Charles University in Prague

Faculty of Social Sciences

Institute of Political Studies

Šimon Vlček LL.B.

Efficient market hypothesis in the modern era

MASTER THESIS

Prague 2016

Author: **Šimon Vlček LL.B.**

Supervisor: **doc. PhDr. Ladislav Krištoufek Ph.D.**

Academic Year: **2015/2016**

Declaration of Authorship

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

The author hereby declares that all the sources and literature used have been properly cited.

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

Prague, May 12, 2016

Signature

Acknowledgments

The author extends his gratitude to his thesis supervisor, doc. PhDr. Ladislav Krištoufek Ph.D., whose advice and guidance have been invaluable. Furthermore the author thanks his family for providing supportive environment necessary to complete this thesis.

Abstract

Efficient Market Hypothesis (EMH) has been the central assumption of financial modelling in the previous decades. At its core, it is a statement about the efficient incorporation of available information in the prices of assets, rendering each price a 'true' representation of the asset's intrinsic value. The notion of informationally efficient financial markets has been, since its formulation, entrenched in the very core of our understanding of how asset pricing works, yet, with ever so increasing frequency, when subjected to empirical scrutiny, it fails to prove its explanatory and predictive prowess.

New academic strands emerged have emerged as a result, attempting to explain those empirical short-comings, with rather mixed results. The new models and theories often either explain a singular anomaly, rather than providing a generalized and consistent theoretical framework, or are exclusive with the general state of financial markets, which tends to be efficient and rational.

This thesis shall explore the relationship of information and financial markets, taking into account developments that have occurred since the inception of the EMH. Subsequently it will present a new theoretical model for asset pricing and *ipso facto* the efficiency of financial markets, based on meta-analysis of information, along with other less conventional measures of market efficiency.

JEL Classification	D50, D53, D59, G12, G14
Keywords	financial assets, market efficiency hypothesis, market inefficiency
Range	≈ 90 000 characters

Author's e-mail	simon.vlcek@email.cz
Supervisor's e-mail	17152303@fsv.cuni.cz

Abstrakt

Teorie efektivních trhů se stala a stále tvoří ústřední předpoklad finančního modelování. Ve své podstatě je tato teorie konstatování o efektivním začlenění dostupných informací do cen aktiv, jež má za následek, že cena je tak shodná s 'vnitřní' hodnotou aktiva. Ačkoliv se pojem informačně efektivní finanční trhy od své formulace zakotvily v samotném jádru našeho chápání oceňování aktiv, když je tato teorie vystavena empirickému zkoumání, stále častěji dochází k prokazování anomálií nevysvětlitelných v rámci této teorie.

Nové akademické prameny se objevily, se záměrem objasnění těchto empirických porušení efektivity trhů, s poněkud smíšenými výsledky. Nové modely a teorie často pouze omezují na specifickou anomálii, bez formulace všeobecného a konzistentního teoretického rámce, nebo je jejich existence neslučitelná s obecným stavem na finančních trzích, která má tendenci být efektivní a racionální.

Tato práce zkoumá vztah mezi informacemi a finančními trhy, s přihlédnutím k vývoji, ke kterému došlo od prvotní formulace teorie efektivních trhů. Následně bude navržen nový teoretický model pro oceňování aktiv a tímto také efektivnost finančních trhů, na základě meta-analýzy informací, spolu s dalšími méně obvyklými metodologickými postupy.

Klasifikace JEL	D50, D53, D59, G12, G14
Klíčová slova	finančních aktiva, teorie efektivních trhů, neefektivnost na trhu

E-mail autora	`simon.vlcek@email.cz`
E-mail vedoucího práce	`17152303@fsv.cuni.cz`

Contents

List of Tables ix

List of Figures x

Acronyms xi

Thesis Proposal xii

1 Introduction 1

2 Theoretical Background 5

2.1 Random Walk Hypothesis 5

2.2 Martingale and Fair Game 8

2.3 Efficient Market Hypothesis 10

2.4 EMH Criticism 14

2.4.1 Proposed Alternatives 17

3 Testing the EMH 26

3.1 Joint Hypothesis Problem 26

3.2 Weak Version 27

3.2.1 Variance Bounds 29

3.3 Semi-strong Version 31

3.3.1 Methodological Challenges 35

3.4 Strong Version 36

4 Bayesian Inference Approach 38

4.1 Informational Landscape 38

4.2 Methodological outline 42

5 Empirical Analysis 46

5.1 Description of the data 46

5.2 Classical Statistical Analysis 48

5.2.1 Model Definition 48

5.2.2 Model Estimation 49

5.2.3 Testing Results 50

5.3 Concluding Remarks 51

6 Conclusion 53

Bibliography 65

A Analysis Outputs I

A.1 Proxy Variable Table II

A.2 Data Plots IV

B Source Codes VI

List of Tables

5.1 Model estimation 49
5.2 Correlations 49

A.1 Information quantity index II
A.2 Information quantity index – cont. III

List of Figures

A.1 Microsoft – Regression IV
A.2 Microsoft – Residuals IV
A.3 Wells Fargo – Regression IV
A.4 Wells Fargo – Residuals IV
A.5 AT&T – Regression V
A.6 AT&T – Residuals V
A.7 Marriott International – Regression V
A.8 Marriott International – Residuals V

Acronyms

AMH Adaptive Market Hypothesis

AR Autoregression / Autoregressive

BSOPM Black-Scholes Option Pricing Model

CAPM Capital Asset Pricing Model

EMH Efficient Market Hypothesis

FMH Fractal Market Hypothesis

GFC Global Financial Crisis

HFT High Frequency Trading

MA Moving Average

REH Rational Expectation Hypothesis

RWH Random Walk Hypothesis

Master Thesis Proposal

Institute of Political Studies
Faculty of Social Sciences
Charles University in Prague

Date: 08.06.2015

Author:	**Simon Vlcek**	Supervisor:	PhDr. Ladislav Kristoufek
E-mail:	Simon.vlcek@email.cz	E-mail:	17152303@fsv.cuni.cz
Phone:	724264448	Phone:	
Specialisation:	IEPS	Defense Planned:	July 2016

Proposed Topic:

Efficient market hypothesis and market inefficiencies in modern era: a sociological approach

Registered in SIS: Yes **Date of registration: 02.06.2015** (in case of No give an expected date)
Remark: the registration must be done by your supervisor but, prior to that, it requires the approval by Dr. Riegl.

Topic Characteristics:

This thesis shall focus on the Efficient Market Hypothesis (EMH) as the cornerstone of asset pricing in economics and finance. EMH has been considered an elegant solution to the problem of transposing the general microeconomic equilibristic assumption of market and supply; in its simplest form, it states that markets are efficient in incorporating all available information about the asset in its price, resulting in correct allocation of those prices. Hence the actual asset prices may be deemed to be fractionally close if not equal to the equilibrium 'intrinsic' prices of the underlying assets.
Since its inception however, markets have repeatedly exhibited so-called moments of 'irrational exuberance' or 'financial bubbles', during which the asset prices became disconnected from their equilibristic counterparts, resulting in fundamentally baseless price fluctuations and volatility. Especially since the global crisis beginning in 2008, the theory has attracted strong criticism, claiming the over-reliance on the theory by the global actors in financial sector at least contributed to the occurrence of the crisis. The focus of this thesis will then further be to examine those critical arguments in relation to EHM, and to develop a sociological theoretical approach, relaxing some of the unrealistic assumptions of mainstream economics strongly relied upon by the EHM, in order to provide an insight into the causal mechanics of financial bubbles and potentially theoretically reconcile the theory with the occurrence of financial bubbles.

Working hypotheses:

1. EHM is not compatible with the market behavior during the occurrence of speculative bubbles
2. Behavior of market agents is not rational during the occurrence of financial bubbles
3. The efficiency and rationality of markets does not decline with increased volumes of information
4. Increasing size of a particular market of securities does influence positively efficiency of such market and *vice versa*.

Methodology:

Both historical and current academic discourse will be surveyed and qualitatively analysed in relation to the occurrence of financial bubbles. Game theoretical approach shall be often applied onto market participants in order to induce behavioral trends in specific cases of market price fluctuations. The analysis conducted as well as the theoretical aspects of the thesis will be inherently microeconomic; focusing on the supply/demand and competition changes, both in cases of rationally and irrationally induced market exchange.

Outline:

1. Introduction
2. Efficient market hypothesis
3. Financial bubbles
    a. Historical outline
    b. Theoretical analysis of occurrences
4. Criticism of EMH
5. Reconciliation of EHM with financial bubbles
    a. Description of the Data
    b. Theoretical model
    c. Behavioral induction and discussion of the results
6. Conclusions
7. References / Bibliography

References / Bibliography:

Fama, Eugene. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." Journal of Finance. 25:2, pp. 383–417

Malkiel, Burton G. 2003. "The Efficient Market Hypothesis and Its Critics" Journal of Economic Perspectives. 17:1, pp. 59-82

Malkiel, Burton G. 1973. A Random Walk Down Wall Street. New York: W. W. Norton & Co.

Constantinides, George, Harris, Milton and Stulz, René, (2003), Handbook of the Economics of Finance, vol. 1, Part 1, 1 ed., Elsevier

Constantinides, George, Harris, Milton and Stulz, René, (2013), Handbook of the Economics of Finance, vol. 2, Part 2, 1 ed., Elsevier

Preda, A. & Knorr Cetina, K. 2012. The Oxford Handbook of the Sociology of Finance. Oxford: Oxford University Press.

Levich. 2001. International Financial Markets: Prices and Policies, 2 ed., McGraw-Hill/Irwin

Fligstein, Neil. 1996. Markets as Politics: A Political-Cultural Approach to Market Institutions. American Sociological Review, Vol. 61, No. 4 (Aug., 1996), pp. 656-673

Stiglitz, Joseph E. 1989. Markets, Market Failures, and Development. The American Economic Review, Vol. 79, No. 2, Papers and Proceedings of the Hundred and First Annual Meeting of the American Economic Association (May, 1989), pp. 197-203

Digitally signed by Ladislav Krištoufek
DN: cn=Ladislav Krištoufek, o=Czech Academy of Sciences, ou=Institute of Information Theory and Automation, email=kristouf@utia.cas.cz, c=CZ
Date: 2015.08.11 13:55:30 +02'00'

. . . .
Author

. . . .
Supervisor

Chapter 1

Introduction

The basic notion that asset prices[1] follow a random walk process, rendering them unpredictable, has been initially formulated by Bachelier (1900). In the 1960s, his work was picked up independently by two scholars, and eventually crystallized in the seminal work '*The Behavior of Stock-Market Prices*' by Fama (1965), along with a proof for the random walk of prices published by Samuelson (1965).

At its heart, EMH is a simple statement about efficiency of financial markets to incorporate existing information into the price of an asset. It should be noted that efficiency in the context of this thesis does not refer to a commonly described feature of markets – resource-allocation efficiency, even though it could arguably still be derived from the micro-economic equilibrium properties of financial markets. Furthermore although market efficiency is often described as a singular hypothesis, it actually consists of three distinct 'forms' of EMH, each of which is a separate statement about different aspects of incorporating information in the prices of assets, thus they must be examined individually in turn, yet still in the context of each other.

[1] Although initially Efficient Market Hypothesis (EMH) has been concerned primarily with stock prices, the theoretical assumptions were extended to essentially all classes and types of assets, including shares, derivatives, and market traded currencies.

Since its formulation, the EMH became incorporated into the corpus of financial and neoclassical economics, with heavy spillovers into policy making, forming the main pillar of support for deregulation of financial markets, and becoming the central assumption of classical financial modelling, including for instance the Capital Asset Pricing Model (CAPM) or Black-Scholes Option Pricing Model (BSOPM). However it has also been subjected to heavy academic scrutiny, due to several theoretical and empirical concerns; it has been argued that the idea of efficient markets is inconsistent with occurrences of market inefficiencies and irrational market behaviour. Even Keynes (1936) famously argued in his '*General theory*' that financial market interactions are much closer to a beauty contests than the ferocious trial of rationality prescribed by the hypothesis[2].

And indeed, as the general governing principle behind asset pricing, the hypothesis has throughout its existence suffered several set-backs and problems. There are instances of consistent empirical violations, various adjustments and reformulations had to be introduced, in order to align the world of *homo economicus*, where the EMH is virtually unimpeachable, with the real world. New academic strands sprung to attention, attempting to improve on the ideas behind the hypothesis or outright replace it. Hence it is accurate to claim that over the course of its history, the hypothesis has substantially evolved.

One particular aspect in the evolution of the hypothesis however has been overlooked. The volume of information published and available about any given asset has increased exponentially, statement especially accurate in the last decade, with the omnipresent increase in the prominence of online communications and specialised technologies implemented for the sole purpose of transferring information, particularly in relation to financial assets. Further

[2] Unless otherwise stated, references to 'the hypothesis' shall refer to the general concept of EMH throughout this thesis.

substantive development on financial markets has been the transformative impact of High Frequency Trading (HFT) combined with the reliance on liquidity provided by these types of market participants. And despite the apparent and direct influence on the both exogenous and endogenous features of financial markets, these changes in the existing state of affairs have been subjected to minimal academic scrutiny or a rigorous theoretical review, especially in relation to the degree of efficiency exhibited by financial markets and their impacts on the theoretical framework of EMH.

This thesis will thus focus primarily on the relationship between information and asset prices on financial markets, with the objective of exploring the topics of EMH testing, and to revisit some of the assumptions underlying the hypothesis, in order to offer a better understanding of how increased complexity and volumes of information could affect the efficiency of markets, and potentially be used to prove consistently that increased volumes of information lead to violations of market efficiency. The thesis will also briefly cover the subjects of bounded rationality of expectation and correlated behaviour through examination of modern alternative theories to the EMH, and a brief theoretical proposal for the utilization of Bayesian methodology to address the presented problem scope.

The thesis shall be structured in the following manner: Chapter 2 shall describe the Efficient Market Hypothesis (EMH), its theoretical formalization, the different versions of EMH, and underlying assumptions. Chapter 3 shall cover the subject of testing in relation to the different forms[3] of EMH, contextualized consideration of the joint hypothesis problem, and particular aspects of the vast body of literature in regards to criticism and proposed alternatives of the hypothesis. It ought to be noted that given the subsequent focus of analysis, a particular methodology shall be considered in detail only for even

[3]In this thesis, references to a form or version of EMH shall be used used interchangeably.

studies, whereas for other types of testing, only brief overview shall be provided. Chapter 4 shall examine a theoretical approach of data analysis *via* Bayesian inference in order to extract information allowing for excess returns. Chapter 5 shall explore empirical analysis and empirical testing of selected aspects of EMH. Chapter 6 shall consider the results, summarize the empirical findings, conclude what may be inferred, and suggest potential areas of future research directions.

Chapter 2

Theoretical Background

2.1 Random Walk Hypothesis

It was arguably Random Walk Hypothesis (RWH) that pre-dated definition of EMH in the work of Bachelier (1900), where he stated "*if one could be sure that the price would rise, it would have risen already*". Without a formal specification, RWH holds that prices[1] of assets follow a random process, whereby each individual value is uncorrelated to any previous value.

The RWH has been long subject to an empirical corrosion, as there is very heavy and statistically highly significant evidence of different types of correlation, as well as various momentum and seasonality effects. Partially, the issue was mitigated by shifting the primary subject of analysis from arguably highly correlated absolute prices to returns, which may, for the purposes of clarity, be formally specified as:

$$R_t = \frac{(P_t - P_{t-h}) + D}{P_{t-h}} \tag{2.1}$$

Where:

R_t = Returns from an asset at the time t

[1] Or returns, as it shall be further discussed in this section.

P_t = Price of an asset at the time t

D = All future cash-flows from the asset[2]

h = Holding period of the asset

A precise formulation of the RWH may be expressed as a process of time series sequence, which fulfils following properties:

$$X_t = \mu + X_{t-1} + \epsilon_t \tag{2.2}$$

Where:

X_t = Logarithm of the price of an asset at the time t

μ = Drift of the random walk

ϵ_t = Random increment variable

And because:

$$X_t = log(P_t) \tag{2.3}$$

Equation 2.2 can be rearranged as $X_t - X_{t-1} = \mu + \epsilon_t$, which in combination with Equation 2.3 may be rewritten as $log(P_t) - log(P_{t-1}) = \mu + \epsilon_t$, which equals to:

$$log(\frac{P_t}{P_{t-1}}) = \mu + \epsilon_t \tag{2.4}$$

Therefore it follows that $r_t = \mu + \epsilon_t$, which by virtue of Central Limit Theorem implies normal distribution in random walk. Thus it may be said that the asset returns consist of the drift plus the random increment, and are normally

[2]It should be noted that the dividend factor is commonly omitted, and only the price difference of the asset is taken into account in modelling.

distributed. [3] (Lawler & Limic 2010)

Based on different statistical burdens, Campbell *et al.* (1997) defined three distinct types of random walk.

Random walk 1 - i.i.d. increments

The first form of the random walk is said to be the simplest, but also the most restrictive, hence merely theoretical in relation to stock prices. The formal requirement of ϵ_t is that it must be independent of its past values and identically distributed (*i.i.d.*) with a mean 0 and variance σ^2. Normal distribution is assumed based on the Central Limit Theorem (Norstad 2005), further reducing its applicability in the real world, hence the final specification of the first random walk is:

$$X_t = \mu + X_{t-1} + \epsilon_t, \epsilon_t \overset{iid}{\sim} N(0, \sigma^2) \tag{2.5}$$

This specification equals to discrete Brownian motion, which is essence cannot be applied to the financial markets. (Campbell *et al.* 1997, pp. 32-33)

Random Walk 2 - independent increments

In order to ease and relax the strict prescriptions of the *i.i.d.* increments, the assumption of identical distribution is dropped, thus in this form, the increments are only independent from each other. This allows unconditional heteroskedasticity of the increments, allowing seasonality of increments, if the independence condition is fulfilled. The independence of increments though is still fairly imposing assumption, as it requires the increments to be uncorrelated in all their

[3] Random walks can also be without a drift, which yields specific statistical properties, such as mean-reversal.

linear and non-linear forms; retaining the property of increments, which cannot be predicted or forecasted, by the virtue of past increment observation.

Random Walk 3 - uncorrelated increments

The assumption of independent increments from the previous form is further relaxed in the weakest form of random walk, whereby the only necessary condition that needs to be satisfied for all k is:

$$Cov[\epsilon_t, \epsilon_{t-k}] = 0 \tag{2.6}$$

And thus non-linear correlation between increments may occur for some k – $Cov[\epsilon_t^2, \epsilon_{t-k}^2] \neq 0$.

2.2 Martingale and Fair Game

The idea of a martingale process governing the movement of stocks had been formulated prior to the RWH, however in a sense, RWH is both more restrictive, and also does not sufficiently cover the entire notion of informationally efficient markets. Where random walk merely restricts the nature of the process by imposing statistical limits on the distribution of returns, martingales specifies a direct and precise relationship between available information and asset prices.

Generally martingales is a process, which satisfies:

$$E[x_t|\Omega_{t-1}] = x_{t-1} \tag{2.7}$$

Where:

Ω_t = Information set of all publicly available information at the time t

x_t = Random integrable variable given the σ-field Ω_t

$E[.|\Omega]$ = The conditional operator for the expected value given information set Ω

This implies that the following condition holds:

$$x_t = E[x_{t+1}|\Omega_t] \tag{2.8}$$

And by defining the difference operator as $\Delta x_t = x_t - x_{t-1}$, the 2.8 may be rewritten as:

$$E[\Delta x_t|\Omega_t] = 0 \tag{2.9}$$

Process satisfying the condition set by Equation 2.9 is said to be a *fair game*, as it dictates that any analysis of available data in order to earn excess returns must be fruitless for any market participant, if the returns follow such process. It is obvious that martingales imposes much weaker limits on the process of asset returns (namely autocorrelation of conditional variance in asset prices) than the random walk, hence is less likely to be violated by empirical anomalies. (LeRoy 1989)

One problem of martingale process in relation to financial market is the requirement of risk-neutrality of market agents. (Samuelson 1965) Subsequently LeRoy (1973) concluded that martingale property would not be violated only under a restrictive model, assuming constant and uniform risk-aversion and AR(1) process in the stock returns, however concurrently efficient markets apparently exhibit at least close approximation of martingales even in the presence of risk-aversion. This conclusion was questioned by Ohlson (1977), who briefly outlined empirically looser restrictions of a risk-averse model of efficiency. Risk-neutrality still however poses problems in relation to market efficiency, due to the empirically demonstrated relationship between risk and returns. (Brown *et al.* 1988)

2.3 Efficient Market Hypothesis

As stated in Chapter 1, EMH refers to the ability of markets to efficiently incorporate all public information in the price of an asset, so that the price is a reflection of the collective rational choice of the market participants, who base their decisions on all available information about the asset. Fama (1965) in his seminal work separated the hypothesis into three forms: the **weak version**, which states that all historical and present information in relation to market data about the asset has already been incorporated in the current price, hence analysis of previous market behaviour in relation to an asset does not provide any indication of future behaviour, which would consequently allow gaining excessive returns. The **semi-strong version** states that beyond the effects of the weak version, the price of an asset at any given moment also already reflects all public information about the asset. Furthermore the semi-strong version also maintains that any new information that becomes public becomes incorporated in the price almost, if not completely, instantly, and any subsequent changes in the price still follow a random walk. Lastly the **strong version** builds upon the semi-strong version and postulates that prices also reflect any non-public information immediately upon a disclosure of such information to a closed group of market participants takes place, rendering any 'private' or 'insider' information unprofitable in regard to predictability of prices or returns. EMH implies existence of the no-arbitrage hypothesis, which refers to the ability of investors to extremely quickly and efficiently identify arbitrage opportunities[4], and subsequently efficiently exploit and exhaust them. Thus

[4]In this context, arbitrage refers to the exploitation of mispricing of a single asset in a single market, rather than the more common definition, referring to the exploitation of mispricing of the same asset in multiple markets.

it is presumed that for any market participant there will not be any arbitrage opportunities available, as identification of an arbitrage opportunity by a given market participant already presupposes it had been exploited and hence would not allow for such individual market participant to exploit it further.

Summers (1986) concisely summarized EMH in the following way:

$$P_t = P_t^* = E\left[\left(\sum_{t=1}^{\infty} \frac{D_t}{(1+r)^t}\right)\middle|\Omega_t\right] \tag{2.10}$$

Where:

P_t = Market price of an asset at the time t

P_t^* = Real or 'intrinsic' price of an asset at the time t

D = Cash flows from the asset from t until infinity

Ω_t = Set of all publicly available information about an asset

r = Expected constant rate of return for a risk-free asset

The first equality in Equation 2.10 signifies the reliance of EMH on the rational market hypothesis, thus the notions of profit and utility maximization are central to the hypothesis. Based on the assumptions of Rational Expectation Hypothesis (REH), the basic model for valuation of an asset is simply the expected value of all cash-flows and *dividends* payable to infinity, discounted to present by the rate of return of a risk-free asset, which is assumed to be constant, given a particular set of information available to the market at any given time.

Rationality, however, is but one of several assumption central to EMH, and although not explicitly defined *ex ante* in his analysis, the consequence of Fama's EMH implies following assumptions about the financial markets:

(i) The acquisition of information also bares no transaction costs, and information is uniformly distributed amongst the market participants.[5]

(ii) All market participants are rational in their behaviour and expectations are uniform as well as rational.

(iii) All market participants are utility maximising.

(iv) Market participants are risk-neutral.

All of those assumptions have been used in order to mount various criticisms, with vast samples of empirical evidence. And due to this empirical erosion that the hypothesis has been subjected to, there has been several changes and amendments introduced along the way. Fama (1991, p. 1575) made revisions pertaining to the uniformity of informational distribution, and introduced the assertion that "*prices reflect information to the point where the marginal benefits of acting on information do not exceed the marginal costs*". The adjustment also allowed to fix the assumption of risk-aversion uniformity, resulting in the relaxation of the prescriptions of the weak form, namely shifting the requirement of RWH to the less restrictive and more realistic fair game process. Lucas Jr (1978) in fact concluded that RWH is not a necessary nor sufficient requirement of market efficiency.

These changes however injected further ambiguity into constituting elements of the hypothesis, and as result there have been references to an apparent divide in Samuelson's and Fama's versions of market efficiency; this was perhaps best highlighted by LeRoy (1989, p. 1613):

> The failure of many financial economists to appreciate the extent of the gulf separating market efficiency interpreted as economic equilibrium and market efficiency interpreted as the martingale model has led them vacillate between

[5] Of course the assumption of uniform informational distribution does not hold for the strong from.

> viewing market efficiency, on one hand, as hard-wired into their intellectual capital and unfalsiable and, on the other hand, as consisting of a specific class of falsifiable models of asset prices.

Thus the argument can be summed as distinguishing two separate elements contained in Fama's version of EMH; first is the "fair game" element, as shown by Samuelson (1965), and secondly it is Fama's equilibristic model of fundamental prices of financial assets. And it is the latter element, which is disputed for the apparently inaccurate predictions and tautological definition. LeRoy (1989)

The equilibrium was expressed by Fama (1976) as an equality of expected returns:

$$E[p_{t+1}|\Phi_t] = [1 + E[r_{t+1}|\Phi_t]]p_t'' \tag{2.11}$$

Which LeRoy (1989) pointed out as tautological, because the conditional expectation is applied to identity of rate of return, which in turn equals p_{t+1}/p_t or r_t. Fama however denied any tautological element, and did not attempt to amend the specification.

The divide of the hypothesis is however somewhat arbitrary, under the assumptions of rationality of market participants, it seems that one element would logically imply the other. If we are to accept the assertions of the first element, but not the second, we are also to assume that market participants are essentially more irrational than rational. And by disposing of the rationality assumption, there is no viable equilibrium model, which could be used to explain the market interactions or used as the basis for modelling. Therefore the discussion may be reduced to a simple dilemma, which is whether market participants in financial markets should be considered rational or not. If the answer is no, and only fair game is accepted as the constitutive element of market efficiency, it introduces further problems. If the irrationality is systematic, which is implied by the virtue of the answer, then it must be shown that such behaviour happens by chance as to maintain the martingale property of the process.

These amendments of the otherwise stringent specification of EMH however induced a sustained academic attack on market efficiency theorists, as non-convergent rational beliefs of individual market participants rendered any attempts to test market efficiency essentially meaningless. (Michailidou 2015) Overall it could be asserted that the proponents of market efficiency did not leave the theory to the numerous academic and policy attacks over the years, but attempted to revise the theory in order to provide a more realistic model of the real financial markets, whilst maintaining some of its explanatory and predictive powers. On the other hand, it could be argued that since the publication of his thesis, Fama (1965) engaged throughout the years until 1998, which were signified by his aggressive defence of the hypothesis, in dubious practice of 'moving the goalpost', significantly changing the original specifications of market efficiency, based on the 'latest' empirical issues presented in academia. Such practice could suggest that the hypothesis is being moulded around the data, rather than providing any useful insight or allowing for improved analysis of the underlying issues.

2.4 EMH Criticism

In this section, the focus shall be on generalized and abstract notions of criticism (and a brief rebuttal where appropriate), as specific violations and methodological issues are further explored in Chapter 3.

The hypothesis is often described to be discouraging in active participation in the financial markets, due to martingale property of returns, whereby acting on new information does not confer any extra benefit – abnormal returns. Therefore any market agent would be better of simply passively investing in the market as a whole, which in turn would cause the liquidity to dry up, and the financial markets would hence come to an unequivocal halt. (Grossman 1976) And even very early empirical enquiries in this topic found that buy-and-hold passive strategies are most likely to either match or surpass returns of funds that had been actively managed (see *inter alia* Malkiel 2003, Jensen (1968), and Malkiel (1995)). This claim is however largely a misconstrued argument based on false interpretation of the hypothesis prescriptions. The hypothesis presupposes an equilibristic market behaviour relying on fierce competition between the market agents, and it prescribes that opportunities to exploit the efficiency are scarce and fleeing precisely because of that. Therefore it can be claimed that this line of criticism "*confuses a statement about an equilibrium "after the dust settles" and the actions required to obtain that equilibrium*". (Ball 2009, p. 10)

Additional arguments often go towards the inability of EMH to reconcile a prolonged periods of irrational market behaviour, or significant departures of prices from the fundamental value of the underlying assets.[6] This argument is based on the misconception that market efficiency does not allow any irrationality at all. Conversely, market efficiency presumes there might be individual agents, who will act irrationally, and it is due to the arbitrage seeking, utility maximizing, rational agents, that such irrationality is quickly exploited, and therefore balanced out. Thus EMH in fact does provide a general framework to reconcile irrationality of market participants, whereby rational agents acts as the counter-balance to irrational agents, eventually reaching equilibrium. There is however a reservation to the rebuttal; the described behaviour

[6]The nomenclature is rather wide in this regard, and may include asset bubbles, market crashes, irrational exuberances and many more, which all seem to describe the same phenomenon.

is only feasible if there is sufficient number of rational agents to balance out the irrationality, market efficiency seems to break down if markets are in a state of overwhelming irrational panic.

It has been also claimed that EMH should have prevented, or at least predicted financial bubbles. Here is where most of EMH criticism has been concentrated, especially since the beginning of the Global Financial Crisis (GFC). And the argument is also further fuelled by the ambiguity of the term 'financial bubble'; bubbles are defined as prolonged and significant departures of prices from the intrinsic value of the asset, however identification of such state of affairs is only possible *ex post*, *i.e.* after the bubble had burst. Therefore the efficiency proponents claim, that until the time of the market crash, the bubble is reflection of rationality, until the moment of sudden irrationality, which is followed by rational correction. (Ball 2009)

Often cited is also the issue of excess trading volume, volatility, and consistent violations of martingales/RWH[7]. (Thaler 2010) The argument of excessive trading activity is based in empirical data, which would suggest that there is too much trading and price changes to be justified solely based on changes of rational expectation in dividend/bond yields, or interest/premiums. (Shiller 1981) The response of the market efficiency camp has been largely inconsistent and unconvincing, generally relying on the micro-macro economic dichotomy (see Section 3.3) or exclaiming the impossibility to exploit these inefficiencies in way contradictory to the hypothesis (Malkiel 2003). This field of the empirical research however arguably helped to spawn the behavioural approach to market efficiency (*e.g.* Shiller 2003) explored in the Subsection 2.4.1.

Third objection relates to the inefficiencies in the testing methodology of the hypothesis, which leads to the conclusion it is safer to reject the hypothesis in favour of other testable theories. This problem is further explored in the fol-

[7]Including seasonality and momentum effects, short-term predictability of returns *etc.*

lowing section, however generally the testability problem, although exemplary in the case of EMH, infects the study of economics in its entirety, and hence ought not to prevent academic endeavours aiming at enlightening those respective fields. Berk (2007) extends this argument by claiming that the conclusion based on the unrealistic assumptions of the hypothesis in fact hinder academic progress and enquiry into the inner-workings of financial markets. However the vastness of academic literature pertaining to market efficiency, both with the focus on invalidation and replacement by other more realistic theory, ought to serve as a sufficient ground to dismiss such objection. In this sense, it could be even said that market efficiency is the forefather of a particular strand of behavioural economics or financial chaos theory.

2.4.1 Proposed Alternatives

Having established the fundamentals basis of the EMH criticism, it is necessary to subsequently explore the new academic strands conceived to address these critiques. The intent is to present a general overview of the major directions of research and how they address the perceived problems associated with the hypothesis, rather than presenting a detailed and technical analysis of these topics.

Behavioural economics

Behavioural economics are argued to be the most significant contender and a natural successor of market efficiency, due to the superior approach to bounded rationality of market participants, and the ability to model specific economic behaviour. It attempts to address the underlying issue of unrealistic assumptions about the rational market participant and over-reliance on rational expectation theory, by introducing the notion of human element into the metaphorical equation. Thus in behavioural economics it is entirely possible market participant behaviour is sub-optimal, irrational, or guided by misinterpreted/incomplete information. (Barberis & Huang 2006)

Tversky & Kahneman (1974) identified several cognitive biases, as a result from heurestic decision making process, which 'guide' irrational behaviour. The group of biases initially conceived by Tversky & Kahneman has been since expanded to include various other biases, based mainly in empirical evidence. The most significant, both theoretically and empirically, biases are:

Representativness A tendency to fit new and unrelated events to already observed events, with the increased likelihood of identifying common features, despite the overall divergence of the two events.

Availability A bias whereby importance of data is derived based on its availability.

Overconfidence A tendency to overstate one's own skills, aptitudes and knowledge about the markets, resulting in irrational behaviour. (De Bondt & Thaler 1994)

Herding A tendency of individuals to follow behaviour of others rather than basing their actions on their own information. Arguably this is perhaps the most analysed and known, yet least understood bias. (Banerjee 1992)

Adjustment and Anchoring A bias of initiating the decision making process within an informational realm, or 'anchor', which essentially presupposes the desired conclusion. Similarly, adjustment refers to the amendment of the initial 'starting point' prior to decision making, in order to reach the desired conclusion. (Sewell 2007)

Over/Under-reaction There is strong empirical evidence, which shows that reaction of market agents to new information is not adequate and therefore does not correctly reflect in the prices on an asset. (Bondt & Thaler 1985) As discussed below, the empirical evidence for this type of bias has been often disputed or dismissed.

Gambler's fallacy The false belief that previous occurrence of an event reduces its probability of it occurring in the future. Thus market agent is likely to assume that stock prices are likely to go down, because previously they were rising for a prolonged period of time. (Kudryavtsev *et al.* 2013)

Besides the biases, Kahneman & Tversky (1979) developed new theoretical framework, 'Prospect Theory', which according to its authors violates the basic assumption of expected utility theory contained within the EMH. The basic premise of the model states that a person replaces true probability distributions with adjusted probabilities, where he will treat "*extremely improbably events as impossible and extremely probable as certain*" (Kahneman & Tversky 1979), therefore the subsequent decision is likely to be based on the subjectively perceived gains. Kahneman & Tversky then derived behavioural fallacies from such framework, which govern market agents' decision making:

Loss aversion Through this fallacy, market agents allocate a lower tolerance toward risk of losses than to an equal risk of reward, as there is inner stigma associated with loss. (Kahneman & Tversky 1979)

Regret aversion This fallacy arises as a result of negative emotional response associated with the realization of bad decision, which is not limited to the material matters (financial losses), but includes the responsibility for a bad decision *etc.* This means that people are reluctant to conclude a losing market action – a trade – and keep the asset for irrational periods of time. (Zeelenberg 1999)

Mental accounting The last fallacy refers to the notion of separating our activities into a separate mental compartments by the nature of the activities, and treating each compartment separately and independently of each other. For instance Gultekin & Gultekin (1983) showed that investors tend to treat January as a distinct period independent from the previous months, which leads to the so-called 'January effect'.

Behavioural economics however suffer from a lack of a unified approach to market behaviour in relation to asset pricing, only explaining various individual aspects, such as various market inefficiencies, which on its own cannot fully replace an overarching theory such as the hypothesis. Conversely, behavioural economics could be used to reconcile or bridge the hypothesis with the existing empirical challenges, however if a generalized framework for the interaction and inter-operation of those approaches is feasible, it is yet to be seen. In his review of the empirical evidence and literature, Fama (1998) dismissed most of the evidence as non-systematic and short-term only, hence rejecting the idea of behavioural economics providing better understanding of the above-described market behaviours. Fama further stresses that the anomalies such as overreaction and underreaction have been shown to occur inconsistently, or by chance, and behavioural economics thus do not provide any new insight. Regardless of the criticism, behavioural economics still provide superior approach to issues of bounded rationality, and were the basis for the Adaptive Market Hypothesis (AMH) to be developed.

Adaptive Market Hypothesis

Lo (2004, 2005) proposed an alternative hypothesis, incorporating elements of 'evolutionary psychology' in an attempt to show that behavioural economics can coexist with the market efficiency in a 'intellectually consistent manner'.

Central to the new hypothesis is the idea that market agents are driven by their evolutionary tendencies of competition, adaptation and natural selection, and prices of assets reflect both these individualistic tendencies, as well as environmental characteristics of markets.

Thus the processes of market trading is an adaptive process formulated by tendencies to optimize behaviour of individual agents through trial and error process, incorporating past experience into the current decision making process. Through this process a system of heurestic knowledge is developed, utilized to solve problem associated with different market circumstances. And as those circumstances change continuously, the primary requirement for obtaining optimality is the ability to adapt the heurestics onto the new set of circumstances.(Tseng 2006) The asset price is thus a result of enviromental conditions, as well as the number and nature of "species", which are a specific group of market participants with a common characteristic (*e.g.* HFT funds, mutual funds, individual speculative traders, or various institutional actors). (Lo 2004, p. 18)

The adaptive market hypothesis acknowledges that emotion and irrationality and cognitive biases may get introduced into the adaptation process, however only those that are able to learn from those 'mistakes' and adapt their behaviour will survive – earn returns. (Lo 2005) Thus applying the hypothesis, markets tend to the market efficient equilibrium through this vicious process of adaptation, however it is never reached given the adaptive behaviour of market participants. And although in theory the equilibrium prices dictated by the EMH are impossible to reach, theoretically the efficient market hypothesis is not mutually exclusive with the adaptive market hypothesis. According to Lo (2004) AMH implies a dynamic non-linear relationship between risk and reward, based on the 'current' market conditions, and due to the continuous adaptive process arbitrage opportunities are bound to occur. Lo (2005) also specified that the governing principle of interaction on financial markets is survival, profit/utility being of secondary nature.[8] Lastly Lo identified innovation as the predictor of survival, given it provides the tools necessary for an improvement of adaptability.

Certainly by comparison to other alternatives AMH is a novel theory, yet it has been gaining increasing amounts of academic attention. It is without a dispute that such cross-disciplinary approach to market interaction ought to provide invaluable insight into the inner-workings of market behaviour. The superiority of the AMH is perceived to be in its notion of reconciliation with the existing theoretical framework and its consistency with the assumption of market efficiency centred equilibrium, or rather an approximation of it.

Market reflexivity

Market reflexivity was firstly conceived as a sociological principle, which states "*the situations that men define as true, become true for them*" (Thomas 1923). In general, it describes a circular relationship, where cause and effect is interchangeable and bidirectional.

In context of financial markets, it attempts to inject sociological methodology into financial modelling, with the major proponent of the theory, George Soros, describing it as:

> Instead of a tendency towards some kind of theoretical equilibrium, the participants' view and actual state of affairs enter into a process of dynamic disequilibrium, which may be self-reinforcing at first, moving both thinking and reality in a certain direction, but is bound to become unsustainable in the long run and engender a move in the opposite direction. Soros (1994)

The theory was further revised and expanded by Soros (2008) to include the notions of imperfect information and bounded rationality. The approach

[8]Although Lo often refers to survival and 'survival of the richest', what it is meant by it in the context of financial markets and trading financial assets is left to interpretation.

to these topics is however different than that of behavioural economics. It is asserted that there is a limited intellectual capacity inherent to the human nature, which limits the ability to percept the world and knowledge in its 'true form'. It is also inherently sociological in the sense of not being able to separate the observer from an experiment, which leads to recursive causal relationship between observer theorizing about behaviour of the whole system, of which he is a part of, rendering theory derived from such observation unfalsifiable in the sense of Popper (2002). (Ehnts & Álvarez 2013)

Shiller (2003) subdued the feedback models under the general notion of behavioural economics, however reflexivity theory diverges from the elements of cognitive psychology, attempting to explain the market behaviour through somewhat dubious axiomatic statements about the nature of human comprehension and knowledge.

Although the theory is quite fitting for the cases of macro-economic bubbles, it loses its applicability if we attempt to apply it more generally to the market behaviour, especially in a micro-economic context, as it attempts to completely replace the general theory of market equilibrium, without resolving many empirical and theoretical issues. Particularly the issue of identifying the precise moment when markets 'swing' out of the balance is where the theory lacks any substance, greatly limiting its explanatory power. Furthermore the theory has been conceived by a practitioner, not an academic, without reliance on comprehensive empirical evidence or academic review. This may also be the reason, why the theory has not been well received in the academic literature, although we can observe transposition of the principles of the theory in many market models, including momentum and long-memory based models.[9]

[9] For overview of the academic literature pertaining to the feedback models, see Shiller (2003, p. 93).

Fractal Market Hypothesis and Chaos Theory

Fractal Market Hypothesis was enunciated by Peters (1994), with theoretical foundation based in Chaos Theory. Chaos Theory is concerned with the study of random/chaotic systems, which are still governed by a set of deterministic elements. (Etheridge & Sriram 1993) Commonly the features of such system are described with an analogy of a tree and its branches; although the general derivable rule is that each branch will divide at least once, defining the overall shape of the tree, the precise terms and properties of each divide and subsequent branch are to a degree entirely random. (Anderson & Noss 2013)

The alternative hypothesis is centred around the heterogeneity of market participants expectations due to information and/or preference asymmetry, which is manifested in the markets by virtue of market liquidity. Thus the willingness to trade financial assets between market participants is given by their expectations, which differ based on the individual information and preference, often referred to as 'investment horizon'. (Anderson & Noss 2013)

Kristoufek (2012) tested the predictive prowess of the fractal market hypothesis and concluded it produced accurate model of behaviour during periods of significant market turmoil, and performed superior to the EMH. There are however some theoretical reservations in the eyes of the author, which still need scrutiny. Firstly the hypothesis identifies convergence of investment horizons of market participants as the major cause of market instability, *i.e.* either due to exogenous short or endogenous cause, the fragile balance between the short and long-term investors is broken, eliminating the willingness to trade, hence causing lack of liquidity. This in turn causes prices of financial assets to fluctuate wildly. However the focus of the literature has been almost exclusively aimed toward the transformation of long-term investor to short-term ones. Moreover convergence of investment horizons does not imply information uniformity, which is one of the central assumptions of the alternative hypothesis, therefore even under uniform investment horizons, one would expect arbitrage to occur, due to the information asymmetry between market participants.

All of these theories deal with the challenges of efficient markets differently, however they present very interesting venues of analysis for economics and finance. The success of the theories will of course be dictated by the academic appraisal they receive, and also their applicability to issues in the real world, something that has been arguably absent from the market efficiency discourse for some time. Although it is doubtful a singular theory is to be accepted in a short horizon alike the EMH almost five decades ago, due to the fractured nature of the current academic discourse, to speculate on the success of either of them would be unproductive at this time.

Chapter 3

Testing the EMH

As it has been outlined in the previous sections, when testing the hypothesis, we must consider each version as a separate set of statements which have to be tested. Fama (1991) proposed a general class of tests for each of the forms, categorization which shall be explored in the individual subsections.

3.1 Joint Hypothesis Problem

Fama (1991, p. 1575) concluded that market efficiency is not testable; condition due to the joint hypothesis problem, where even if we reject the null hypothesis of efficient markets due to finding inefficiencies in the pricing of assets, our conclusion is conditional on how we estimate what the 'correct' price should be, and hence any results concluding the rejection of our null hypothesis may be attributed to incorrect estimation of the 'correct' price. It is also useful to mention the conclusion of Summers (1986, p. 596), where he estimated that to have a 50 % chance of rejecting market efficiency, one would need data for approximately 5000 years, which illustrates the vast space between the theory and the data.

Lo (2007) concisely summed the problem in a following manner:

> One of the reasons for this state of affairs is the fact that the EMH, by itself, is not a well-defined and empirically refutable hypothesis. To make it operational, one must specify additional structure, for example, investors' preferences or information structure.

3.2 Weak Version

Providing that weak form implies there can be no inference made about the future behaviour of the price or return based on the current/past information, the focus of testing is on the fair game process, or in stricter context the RWH. Thus the testing can be subdued under the general category of *predictability of returns*. The prediction is commonly asserted based on historically observable variables including past returns, both in short and long time horizons, which has been historically the most densely explored subcategory. Further variables commonly tested often include different performance indicators (dividend yields or profit/earning ratios) or sector specific variables. The general methodology may be summed up as a search for AR and/or MA processes in those variables, which would indicate persistent effects of past variables in the future price of an asset, hence violations of the weak form processes.

Summers (1986) was one of the first to provide a simplified model to approach the testing by specifying the hypotheses as follows:

$$H_0 : P_t = P_t^*$$

$$H_a : P_t = P_t^* + u_t \tag{3.1}$$

Where the actual price of an asset is comprised of the 'intrinsic' price of the asset and a mispricing term, which follows an AR(1) process. Such a simplistic specification of the model proved to be very effective for future research, separating methodologically the subject of analysis as the mispricing term. Given that this type of efficiency is viewed as potentially the most profitable and easiest to exploit, with plenty of empirical evidence in regards to various inefficiencies, such as seasonality of returns, the body of research for this section is extremely rich.

Fama (1991) separated the tests of return predictability into several distinct categories:

1. **Time variation of past returns - long-term returns** – These tests are usually associated with the notion of mean-reversal of asset returns, where returns tend to revert to mean after larger swings or departures from the fundamental values of the asset. Fama (1991) however dismissed any empirical evidence by asserting that the irrational bubbles in asset pricing, which would suggest potential reversal to the mean, are *a priori* indistinguishable from "*rational time-varying expected returns*" (Fama 1991, p. 1581), hence it does not violate the fair game characteristics of asset returns.

2. **Time variation of past returns - short-term returns** – These tests are focused on short-term correlation and momentum effects in returns, as the hypothesis dictates that returns should not be affected by these types of processes, which would render them predictable in short-term horizon. Due to the empirical evidence suggesting otherwise, Fama (1991) argued that these inefficiencies are in fact not exploitable, due to the transaction costs associated with exploiting them, thus the fair game prescriptions are not violated by this. However the increase in the HFT based trading, which rely almost exclusively on this type of testing, would suggest that exploiting of these inefficiencies enables to earn excessive returns.

3. **Seasonality** – These tests search for repeating patterns in returns. Various seasonal effects have been identified, including the weekend effect

(Gibbons & Hess 1981), which is signified by stable negative returns throughout the weekends, or January effect, which is signified by above average returns in January (Harris 1986). There is plenty of empirically focused literature centred around the idea of patterned violations of the predictability of returns, and interestingly Fama's response to the evidence of seasonality has been rather accepting in comparison to the overall dismissive attitude toward empirical evidence of market insufficiency.

3.2.1 Variance Bounds

Variance bounds testing is a specific subset of tests of martingales property, which gained significant traction in academic literature. The intuition behind the variance bounds testing is surprisingly simplistic; if the observed fundamental price of stock is calculated *ex post*, using Equation 2.10, the forecast error between the actual realization of the price P and the rational *ex post* P^* is contained in the *i.i.d.* error term, in the equation Equation 3.1.

Based on the above conclusion, Shiller (1981) then asserted that relationship then implies the following:

$$Var(P_t^*) = Var(P_t) + Var(\epsilon_t) \tag{3.2}$$

And therefore it follows that:

$$Var(P_t^*) \geq Var(P_t) \tag{3.3}$$

Thus if the variance of the actual price was higher than the *ex post* variation estimation, violating the upper bound set by Equation 3.3, this would imply that the variance is not explainable by the change in the rational estimate of the price, and hence price variance is caused by irrational behaviour of market agent. Although the initial definition of the variance bounds is simple, the challenges associated with the testing of the bounds has proven to be challenging. Deriving asymptomatic test distribution for a finite sample is a usually also a problematic feature of the research. And the of course, yet again, the joint hypothesis problem must be mentioned, as a violation of those bounds is defined by the *ex post* naive forecast process, which may be misspecified.

Mankiw *et al.* (1991) attempted to improve on work of Shiller (1981) by altering the *ex post* asset-pricing model in the following manner:

$$P_t^{*h} = \sum_{j=0}^{h-1} \left(\frac{1}{1+r}\right)^{j+1} D_{t+j} + \left(\frac{1}{1+r}\right)^{h} P_{t+h} \tag{3.4}$$

And subsequently derived the new bounds:

$$E\left(\frac{P_t^{*h} - P_t^0}{W_t}\right)^2 \geq E\left(\frac{P_t^{*h} - P_t}{W_t}\right)^2 \tag{3.5}$$

and

$$E\left(\frac{P_t^{*h} - P_t^0}{W_t}\right)^2 \geq E\left(\frac{P_t - P_t^0}{W_t}\right)^2 \tag{3.6}$$

Where:

P_t^0 = *Ex post* naive forecast at the time t

P_t = Optimal forecast at the time t

W_t = Scaling variable at the time t

h = Holding period for an asset

Both of Equation 3.5 and Equation 3.6 imply volatility upper bounds, by imposing minimal expected values on squared errors between the naive and optimal forecasts, as well as between the actual price and the optimal forecast.

Mankiw *et al.* (1985) showed that unlike Shiller (1981), their test did suffer from a bias, and this characteristic of the test would not be affected by the sample size.

Arguably the power of the variance bounds is derived from the fact that the upper bound depends only on the *ex post* rational price forecast, not the information set Ω_t. (LeRoy 1989) Fama (1991) however rejected the whole notion of variance bounds, as they provide no explanation as to whether the variation is rational or not, combined with the joint-hypothesis problem, because the assumed continuous dividend asset-pricing model is jointly tested.

3.3 Semi-strong Version

The testing of the semi-strong version consists of *event studies*, whereby reaction of prices / returns of assets ought not to be predictable by observation of new information announcements.

A general methodology for the analysis of event studies was outlined by Campbell *et al.* (1997):

(i) *Event definition* – Firstly the event — scope of the information that will be subjected to analysis — related to an asset must be specified. Moreover the event window ought to be defined. The event window may be derived based on the frequency of information that is available, *i.e.* analysing daily data would suggest event window spanning one to several days, but it is not, nor it can be, 'set in stone'.

(ii) *Selection criteria* – Subsequently, it is logical to define the particular range of assets, which will be subjected to the analysis. Campbell *et al.* (1997) suggests this may be according to some qualification criteria, such

as public listing on an exchange, or membership in a particular industry, which may be sensitive to the previously defined event.

(iii) *Returns definition* – This step is arguably the most crucial and problematic, as it consists of defining both normal and abnormal return functions, allowing the researcher to compare returns after that even had taken place, and at the same time returns as if the event had not taken place. Abnormal returns may be for instance formalized as:

$$\epsilon^*_{it} = R_{it} - E\left[R_{it}|X_t\right] \tag{3.7}$$

Where X_t refers to the information about the normal returns, and is commonly represented by market model, in which abnormal returns are compared to the market returns.

(iv) *Estimation procedure* – Next step is to estimate subset of data with the normal returns model, in order to confirm that the model can be utilized to successfully and significantly, from a statistical point of view, quantify normal returns, excluding data related to the event that is being analysed.

(v) *Testing procedure*[1] – Lastly null hypothesis is defined and tested against the data. Importance of the null hypothesis specification must be stipulated here, as it will influence the subsequent evaluation of results and what, if any, conclusions may be derived in relation to the efficiency of markets.

It must be noted in relation to the methodology that it does not consider data with higher frequency than in terms of days, and similarly Fama (1991) also ponders analysis of events with 'daily' data, which is contrary to the current

[1]There are further steps described by the authors, however those are related to general econometric analysis relating to interpretation of results, hence listing them is unnecessary.

state of affairs, in which information about assets are published in frequencies of seconds and minutes, converging closer to a continuous time series data stream. Moreover, the methodology does not allow analysis outside of excessive returns, such as volatility. Brown *et al.* (1988) examined the post-announcement volatility and concluded that although there is some predictability in the volatility immediately after an announcement, the volatility still follows a white noise processes. Hence post-event volatility does not provide any meaningful way to earn excess returns. In fact, Fama (1991, p. 1587) dismisses volatility as an appropriate test measure, given that it heavily runs into the joint hypothesis problem especially with a market response which could have for instance a higher order lag response to an announcement.

However it is the contention of the author that volatility may in fact be used in event studies as the formative element of the 'Returns definition' element, and that volatility may consequently be utilized in effiency market testing, specifically of micro-economic efficiency.

In regards to this issue, the micro/macro-economic efficiency dichotomy must be briefly examined. Samuelson (1998, p. 36) states:

> we've come a long way ... toward *micro* efficiency of markets: Black-Scholes option pricing, indexing of portfolio diversification, and so forth. But there is no persuasive evidence, either from economic history or avant garde theorizing, that macro market inefficiency is trending toward extinction

From this statement, it may be derived that asset prices are correctly priced relatively to each other, and on a micro-economic level, the no-arbitrage rule still applies. Conversely however, in the case of macro-economic level, the absolute prices of assets may over a period of time diverge from their intrinsic value greatly.

It is easy to accept the assertion of market efficiency limited to micro-level, given the amount of evidence in relation to the boom-bust macroeconomic cycle, most prominently signified by the GFC. Thus focusing exclusively on micro-economic level when dealing with semi-strong form of EMH, following null hypothesis can be stated:

$$P_{t+1} = P_t | \Omega_{t+1} \tag{3.8}$$

Rejection of such hypothesis would imply that semi-strong version of EMH stands. Such hypothesis however will never be rejected because we cannot quantify the true price of an asset P_t^* or the effect of $\Delta\Omega$ on ${P_t}^2$, where $\Delta\Omega$ refers to the change in the body of available information about an asset in the time-frame of t and $t+1$. In order to improve the hypothesis, we have to take into account variation in P_{t+1} unaccounted for in the $\Delta\Omega$, which can be done by adopting the assumption of 'noise trader' proposed by De Long *et al.* (1990), where the error term specified in Equation 3.1 is expressed as a function of variance of misconception in the price of an asset of noise traders:

$$u_t = \sigma^2_{p_t} = \frac{\mu(p_t - p*)^2}{(1+r)^2} \tag{3.9}$$

To evaluate such specification would of course involve the criticized utilization of past volatility variance, to work out confidence intervals, allowing us to estimate excessive volatility explainable by $\Delta\Omega$ in Equation 3.8. Retrospectively utilizing such model would be quite simple, identifying the $\Delta\Omega$ and then looking at the variance of volatility. However the predictive power of the model is *a priori* rather questionable; this is because *ex post* identification of the the change in information leading to the change in price is simple, but assuming *ex ante* position, there is no indication except of the excessive volatility. Hence such model would need to rely primarily on identification of market inefficien-

[2]Except for the case of P being constant, which however would itself violate efficiency market hypothesis.

cies through the lens of volatility variance boundaries, which would assist to distill the precise $\Delta\Omega$ almost immediately.

In relation to semi-strong form, it must also be notes, that there exists some anecdotal evidence in relation to High Frequency Trading (HFT), which would suggest that although generally it would be expected that HFT greatly improves the semi-strong efficiency on micro-economic level, as new information is 'worked into' the price of an asset in an exponentially faster manner in comparison to human cognitive abilities, because prices adjust to news before any person can react, it can cause clear and arguably exploitable market inefficiencies. Unfortunately such argument cannot be really extended beyond an assertion, as there is not sufficient empirical data to properly analyse this phenomenon.[3]

3.3.1 Methodological Challenges

Besides the universal issue of testing the hypothesis, described in Section 3.1, there are other methodological problems associated with the first two steps of event studies. Event studies are almost exclusively performed on observed events[4], which initially seems logical, because collecting data on events, which are yet to be observed, in order to make predictions[5] in a statistically significant manner would be difficult, if not impossible. This however leads to a greatly biased sample of events subjected to analysis, as only those impacting the markets will be analysed. Thus the new information is assigned *a priori* significance, which is then tested against the effect it supposedly had on the market. And although the bias may manifest in the results in contradictory ways, due to the fact that the over-reaction to information may effectively be interpreted

[3] It is worth noting however, that the instances of the described 'flash-crash' scenarios have recently greatly increased.

[4] That is to say, information contained in Ω_{t-1}

[5] And evaluate the performance of our models based on the observation.

both as a sign of near-perfect efficiency and micro-economic inefficiency, the results tend to be skewed in favour of overreaction to new information. (Bondt & Thaler 1985)

Secondly the identification of the even window poses a problem for fairly obvious reasons; the longer the window is extended, the less reliable the results are, as the potential effect of the initial event is diluted by new events taking place and new information becoming publicly available. Therefore by the virtue of increasing the lag of the variables, the results are in turn losing the necessary robustness.

3.4 Strong Version

Lastly, tests of strong version of EMH are *tests for private information*, which inspect informational advantage of individual market agents *vis-a-vis* their ability to obtain excessive returns. Fama (1991) concludes there is sufficient evidence that insiders clearly have information, which allows them earning excessive returns. This resulted in a general shift of the direction to test for private non-insider information, which would allow earning excessive returns after costs are subtracted. The general empirical approach has been to examine the trading results of 'smart money' professional investors and analysts, with the overwhelming results that investors do not have private non-insider information allowing them to earn excess returns. Based on such evidence, Fama (1991) concludes this as another confirmation of market efficiency. This is somewhat controversial, given that this argument also runs head on into the joint hypothesis problem.[6]

With the adjustments of the EMH to allow for information not uniformly available to all market agents, along with the empirically conclusive results,

[6] In this case, one would be testing both market efficiency, but also the 'smart money' investors ability to infer information from market inefficiencies and exploit them.

relating to the absence of true strong market efficiency, the distinction between the strong and semi-strong form essentially became obsolete; and because of the secretive nature of 'smart money' investors, it is unlikely there will be any significant development, as the academic focus shifts away from the strong form of EMH.

However there is a growing body of evidence (see summary in Leuchtkafer (2015)) related to HFT, which shows that high frequency traders might be obtaining publicly available information with a decreased latency compared to other agents, resulting in a consistent ability to exploit the new information. Arguably though there violations could better be subdued under the semi-strong version, providing the as it is difficult to classify information 'public', given that it is only to become public in the future, where the difference in the status of the information in relation to whether it is 'private' or 'public' can be defined as a time range of milliseconds. This is a very prospective field of potential research, which could provide invaluable data relating to efficiency of markets on micro-scale, compared to the analysis of efficiency conducted so far.

Chapter 4

Bayesian Inference Approach

4.1 Informational Landscape

The great impact and influence of information market efficiency has been arguably based in its general overarching application. Once Samuelson (1965) published his proof of the law of iterative expectation, the hypothesis gained mathematically rigorous and sound foundation, which allowed its subsequent prolific settlement into the very core of financial modelling. As such, under the assumption of informationally efficient markets, it is not necessary to inspect quantitative or qualitative nature of information, because it is elegantly reflected in the asset prices.

As both the volumes and diversity of the information about market grows however, it is reasonable to assume the following properties; the distribution of information is fractional between market participants, where the reflection of the information onto the markets takes place asymmetrically, based on the qualitative essence of the information. This is due to the simple fact that there is in fact abundance of data and information about any given asset, which a market participant will receive both intentionally and unintentionally. Fama (1991) claimed:

> In brief, because generating information has costs, informed investors are compensated for the costs they incur to ensure that prices adjust to information.

This, as well as other assumptions about the information however start to break down once generating information is not only free, but so is access to it. It is not thus an informed investor, but an investor which is able to qualitatively assess the abundant information in an efficient manner, which may potentially yield excessive returns.

For the purposes of subsequent analysis it is useful to distinguish between the types of information that affect, or are expected to affect, the financial markets by their qualitative essence. The author proposes a following classification, based on the degree of relevance to the financial asset: the first degree of relevance is information related directly to a particular asset, current and previous prices, asset volatility, market depth. This information essentially signifies the very essence of the financial asset and hence it has been subjected to heavy scrutiny, both in the worlds of practical investment – trading – and academia. And it has been mainly this first degree information, which has been used to mount attacks against the weak form of EMH, with somewhat consistent empirical violations, including various correlations of returns or excessive volatility and liquidity. Second degree of relevance then refers to information indirectly related to the asset. This group is arguably the largest, as its scope includes any 'second-hand' data, including various analysis and reports about a particular asset. Third and the final degree of relevance would be information, which is not 'formally' related to the financial asset, yet is capable of affecting market participants. Such information would include for instance significant legislative changes, resulting in industry-wide turmoil, or political instability causing economic uncertainty.[1]

[1] Obviously in certain cases it may be impossible to assign a piece of information into a single class, however for this thesis, important is the distinction between the first and other degrees of relevance.

The newly and abundantly free information may be subdued mainly into the second degree of relevance, and is closely intertwined with the idea of 'smart money', whereby performance (or rather the lack thereof) of 'smart money' investors has been an important focal point for the validity of semi-strong and strong forms of EMH. Malkiel (2005, pp.8-9) summarizes the historical performance of 'smart money' investors and concludes:

> Thus, even if markets are less than fully efficient, indexing is likely to produce higher rates of return than active portfolio management. Both individual and institutional investors will be well served to employ indexing for, at the very least, the core of their equity portfolio.

The conclusion is in line with the conclusions drawn by Merton (1985) that 'smart money' investors do not have access to any kind of special private excess return earning models, which would logically support the conclusion that the informational efficiency suggested by semi-strong and strong forms of EMH does take place on financial markets.

Although the above-mentioned analyses may be used in support of the EMH forms, it is a secondary statement about the markets, as historical performance of 'smart money' may very well simply mean the inability of market participants to extract the complex relationships contained within the vast informational landscape, rather than confirmation of market efficiency. And there can be no contention that the complexity of such relationships increases exponentially even on day to day basis, which is of course both due to the increased volume of traders, volume of information published about the trading, higher availability and also higher frequency of such data. It is perhaps with regard to this complexity, which resulted in strong focus on the analysis of weak form and the underlying data by both academia and practitioners, as it presents more quantifiable and statistically approachable challenge.

Historically the 'smart money' investors would consider their **private** trading models extremely valuable, given the models would be marketed as earning excess returns in order to gain new clients and subsequently more profits by the virtue of management fees. This of course means that even for *ex post* analysis, there is only limited amount of publicly available information for such assessment, namely the overall performance of the investors, along with usually vague information about the trading strategy[2]. However lately there are increased instances of where 'smart money' investors share their investment strategies 'as they go', *i.e.* their trading activity is made available immediately as an instruction has been executed. This opens completely new field of study, as it is still unclear how the market may or may not react to such information. There is a substantial body of work in regards to the access to private / specialized information, however it is highly questionable, whether the conclusions of those works (see, *inter alia*, Fama 1970; Summers 1986) could be transposed to information that is in high volumes available to the public. Example of such service can be for instance eToro[3] or Tradingfloor[4] platforms, both of which allow to share insights, opinions and analysis in relation to the financial markets, but more importantly allow to share publicly all investment decisions.

Furthermore there is information available about the statistical performance of 'smart money', based on the individual trades they execute. Hence their performance can be evaluated easily, and tracked, in order to determine the 'sentiment' of the market. Arguably both semi-strong and strong versions of EMH would dictate that such information would not provide any edge over the other market participants. However in aggregate, if processed sufficiently quickly, it might be possible to gain informational advantage by examining large

[2]Normally this will include specification of proportion of actively/passively managed funds, sectoral focus and other generalized information, which is not of much importance.

[3]Run by eToro (Europe) Ltd., available at http://www.etoro.com/.

[4]Run by The Saxo Bank Group, available at http://www.tradingfloor.com/

amount of 'smart money' investors, with statistically proven past performance; and if the quantities are sufficient to dismiss any doubt of earning excess returns by chance, and subsequently analysis of this data would allow earning excess returns, it would be quite significant violation of the efficiency hypothesis.

The other type of exponentially growing resources potentially invaluable for the analysis of financial markets is data shared on social networks, namely Facebook[5], LinkedIn[6], and Twitter[7] in particular. These services allow to share information instantaneously with quite literally the whole market, and are used by millions of users on daily basis.

These venues represent a new class of information, which, whilst developing the theoretical framework of the EMH and its subsequent revisions, had not been in existence. The effect of this fundamental transformation of the 'informational' environment remains elusive in academia, due to extreme difficulty in obtaining sufficient amounts of data, as well as imposing requirements for computing resources. However examination of the new landscape is inevitable, providing insight into the deeper relationships guiding the financial markets. Due to the novel nature of the information, new approaches to process the data must also be developed; following section includes a brief and theoretical outline of such approach.

4.2 Methodological outline

The proposed methodological approach in this section is applied upon classes of information, which arguable had not existed throughout the formation of EMH, however the general alternative methodology approach may be easily derived from these examples.

[5]Run by Facebook Inc., available at http://www.facebook.com/.
[6]Run by Linekdin Corp., available at http://www.linkedin.com/.
[7]Run by Twitter Inc., available at http://www.twitter.com/.

Thus in order to test these assumptions, a model needs to be construed, based on the behaviour of smart money. There is however an inherent problem with the data in question, there is bound to be extensive variance in the trading, and hence the individual information may be discarded as trading noise, rather than meaningful information about the potential change in asset returns. It is due to this issue, that the Bayes inference is implemented, rather than a classical deterministic model. It allows us to treat the behaviour of an asset essentially as an unfair coin, which has an unknown probability of getting heads or tails, in this case increase or decrease of the price in time t; and the unknown probability is adjusted every time we observe data, which may suggest bias of the coin, in this case individual investors trading the asset.

In order to distinguish between noise traders, and those investors, whose trade actually have certain market correlation, their performance would be utilized, with normalization through Wilson score interval[8], calculating the adjusted probability interval of 'successful trade' for the individual investor and updating our believed probability.

As a prior, we would assume binomial distribution. Such assumption will be useful for the purposes of the analysis, because behaviour of the asset price may be equated to an unfair coin, where the 'fairness' of the coin is unknown. Alternatively one could use uniform distribution rather than binomial for the prior, in order to avoid possible error in the *ex ante* assumptions, and also avoid the potential consideration of conjugate distributions.

In practical terms, with assumption that in time $t+1$, the probability of a price going up would be equal for the cases of $P(P_t{+}1 > P_t)$ and $P(P_t{+}1 < P_t)$, which would be the prior distribution, as we assume no information about the potential movement of the price of an asset, subsequently the probability

[8] $\pm w = \left(\hat{p} + \frac{z^2_{\alpha/2}}{2n} \pm z_{\alpha/2}\sqrt{\left[\hat{p}(1-\hat{p}) + \frac{z^2_{\alpha/2}}{2n}/4n\right]/n}\right) / \left(1 + \frac{z^2_{\alpha/2}}{2n}\right)$

would be adjusted with each observed data point, in accordance with the Bayes theorem:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \tag{4.1}$$

And update our belief accordingly whether the price is more likely to increase or decrease. The data point in such a case would be an observation of another investor trading the asset in question, where due to the Wilson score adjustment, past performance of market participants is also taken into account.

In formulation of the prior, any other information can be incorporate in the model, including various inefficiencies relating to weak version of EMH, momentum, but also announcement information about the asset *etc.* Thus the Bayesian model can be used both as a complementary efficiency model to crystallize the conclusions, or as the formative element of the model.

Furthermore, if the data contains ordinary non-quantifiable language, there are other steps that have to be taken, in order to derive any qualitative value of the data. In order to do so, it is possible to analyse the sentiment of the ordinary language to assign it a quantifiable rating and possibly derive the impact it might have on the asset price. For such purpose, it would be necessary to employ a machine learning algorithm/s, for instance Naïve Bayes. (Baylis 2015)

This approach also runs into the join-hypothesis problem; even if the model is successful in predicting moments of inefficiency in violation of martingales, where perhaps *ex post* identification of such inefficiencies may seem obvious, but there is no way to quantify the efficiency violations, especially given the distribution of probabilities offered by Bayes. Thus probably the market-based return benchmark would have to be utilized, which in turn invites to questioning of the results by the efficiency proponents. And there is another problem inherent to the Bayesian inference, providing that hypothesis testing is treated simply as any other Bayesian inference, the result from the analysis is a probability distribution over parameters, and there is no clear way to reject null hypothesis, introducing ambiguity into the formal process of testing the EMH.

Bayesian estimation has been recently gaining academic traction in other fields, for its arguably superior explanatory powers to the frequentist approach, including the t test and null hypothesis significance testing. This is due to the ability to derive results even from otherwise statistically insignificant results, as well as improved results due to reflection of the distributional consideration into the probability modelling in a more direct manner. (Kristoufek 2012)

This sections serves to illustrate the theoretical possibilities of alternative or secondary semi-strong EMH tests, which would suggest specific analysis of Ω_t in order to extract market inefficiencies observable only by the virtue of inference from complex data structures. Although this idea is not novel *per se*, the application in relation to EMH is yet to be explored.[9]

[9]The approach described in this chapter is yet to be subjected to a full empirical analysis, which will however require significant computational resources devoted to the task, which at the time of writing this thesis was not possible.

Chapter 5

Empirical Analysis

In previous sections of this thesis, it has been asserted that the increased complexity and volumes of information may cause predictive patterns of increased volatility, as market agents struggle to efficiently reflect the such volumes of information into the prices. Volatility is a market neutral measure[1] and agnostic to the market movement direction, hence it is a useful measure of the net effect of the **volume** of information on the prices of assets, regardless of whether the information causes positive or negative market reaction.

5.1 Description of the data

The subject of the empirical analysis is the test for a linear relationship between price volatility and the amount of information that is being made available. In order to quantify the latter, we shall utilize a proxy variable, which ought to approximate the changes in the quantities of information published about a particular stock. The proxy variable shall be defined as follows:

$$I_t = N_t * (1 + \Delta T_t) \tag{5.1}$$

[1]Usually deemed to be a proxy variable for the riskiness of an asset.

Where:

I_t = Proxy variable – information quantity index for a period t

N_t = Number of indexed search engine results for a period t

T_t = Search engine trend index for a period t

To further elaborate, the first term of the equation (5.1) represents a proportion of internet news items published about a company in a given period, indexed by the search engine Google[2]. The quantity is adjusted to the monthly 'trend' index, which signifies the relative change in number of searches for a particular term via the search engine. Thus the proxy variable reflects both how much information was published, and the 'reach' of the information.

The choice of companies to be included was essentially random, thus cross-industrial panel has been chosen, with various capitalization cap, which proxies both the size of the company and the market depth at any given moment, in order to encompass the different characteristics of the companies, which will also help to identify the possible effects of those characteristics in regards to market efficiency. Due to the resource gathering costs, the overall number of stocks analysed is limited to 4, which however ought to be sufficient in terms of statistical significance. Furthermore a specific consideration had to be given to the uniqueness of company's name and its interchangeability; when assessing the number of news items in the proxy variable, the full name of the company, the stock quote, as well as the common name of the company (if available and widely utilized) have been used, and hence if those names are highly interchangeable with items note related to the company, the proxy variable would become highly unreliable (an exemplary case would be Apple

[2]Google Search is an internet search engine, which indexes publicly available information published on the internet. Other competitor search engine results are not included, primarily due to duplicate results problem (*i.e.* certain published results included repeatedly, whilst other are not), and secondly due to the overwhelmingly dominant position of Goodle Search on the market.

Inc., which is too general to be included in the analysis). Lastly in regards to the period of the analysis, the years 2010 - 2016 have been chosen, due to the relative stability of the stock market, as opposed to the years prior to 2010, which were still tainted by the .

The period frequency in question was chosen to be calculated based on monthly data, which ought to be sufficiently short to reflect on the fact that information is being processed much more rapidly, and at the same time long enough to empirically show a **general** relationship between the data quantity and volatility in returns.

5.2 Classical Statistical Analysis

5.2.1 Model Definition

The basic model for estimation is for each of the selected stocks:

$$V_t = \beta_1 X_{t-1} + \epsilon \qquad (5.2)$$

Where:

V_t = Monthly return variance of a stock at a time t [3]

X_t = Independent variable – the information quantity index at a time t

Because of the length of the data frequency, which is monthly, it is not necessary to include larger lag of the variable, especially with regard to the increasing chances of the omitted variable problem.

[3]There are references to 'volatility' of stock, which refers to the standard deviation of returns. Because we are not concerned with the original units of volatility, we utilize the squared term.

The null hypothesis is that β_1 is zero, for which the EMH holds, as no significant linear relationship is present between the quantity of information and the volatility of stock returns, which would imply that markets react efficiently even to higher volumes of information and incorporate it without a delay into the market.

The values of the proxy variable for all the stocks are listed in Section A.1.

5.2.2 Model Estimation

The results of the estimation, including the p-values of the coefficients are presented in the following table:

Table 5.1: Model estimation

Stock	Coefficient	Standard Error	Adjusted R^2	p-value
Microsoft	-160306	177338.	0.00261915	0.369163
Wells Fargo	-44138	21243.8	0.045239	0.0414616
At&T	-23131.8	32824.8	0.00724337	0.483365
Marriott International	-1537.43	1068.9	0.0150387	0.154862

And also table of the correlation factor between the two estimated variables:

Table 5.2: Correlations

Stock	Correlation
Microsoft	-0.0464633
Wells Fargo	-0.238785
At&T	-0.0766807
Marriott International	-0.225975

We have obtained statistically significant results at the 95 % confidence levels only in one case, Wells Fargo stock. The negative coefficient in all of the estimations indicated the inverse relationship from the initial assumption, hence it would seem that increased volumes of information in fact lead to lower variance in returns. However even for the statistically significant result, when the standard error and adjusted R^2, it becomes fairly obvious that the model is misspecified, as it holds extremely low explanatory element. This conclusion is further enhanced by the information in Table 5.2, which features also very low levels of correlation between the two variables, which would otherwise indicate a more direct and linear relationship. Justifiably, the null hypothesis cannot be rejected based on the results presented above, and in consideration of the coefficient values, even the statistically significant result could not be effectively used to reject the null hypothesis.

5.2.3 Testing Results

Given the initial values of our estimation, additional testing and robustness checks of the model would be superfluous. It is clear that author's proposed model is a clear null finding, with negligible statistical possibility of rejecting null hypothesis. Various transformations of the independent and dependant variables were tested in order to check if improvements could be made to the model, however improvements in the model's adjusted R^2 were also negligible.[4] The best results were yielded with a square root transformation of the independent variable, due to the quadratically distribution, however as previously stated, the overall impact on the model was negligible. It is also evident from the plots presented in Section A.2, that residuals of the model are serially correlated, approximating exponential curve. Transformations specifically introduced in order to address this issue however did not sufficiently alter the residuals of the model, and given the overall results of the estimation, this points to the omitted variable problem. Generally there is a definitive relationship between the data not being captured by the model, which is evident from the identical shape of the residuals for all of the regressions. Restricted samples

[4]Given the low significance of the results, the author did not include all of the model's mutations, as they would add very little value to the analysis.

were also used in estimation, in order to test for correlated values in clusters, however there were no significant results or improvements in the estimation or correlation values using even very specific samples. Although developing a more complex non-linear model for the estimation could yield better results, the focus of this thesis is to show whether **direct and general** relationship exists between volume of information and volatility of asset returns; for which the outlined results shall suffice.

5.3 Concluding Remarks

From the results of the analysis, it would seem that markets actually tend to be more efficient with the progress of time, regardless of the informational volumes, which have to be correctly reflected in the prices. This would suggest that the technological progress enabled market participants to efficiently process vast amount of information, which in turn resulted in more informationally efficient markets. Thus in violation of the our original assumption, there seems to be a negative relationship between the volumes of information and volatility of returns, which would suggest that with more information, the financial markets actually are more efficient. Of course there are also statistical considerations, which could have caused the null finding, including:

1. The chosen data frequency was not fine-grained enough, which caused the regression to be polluted by the aggregate noise data. This could be addressed by estimating weekly or even daily data.

2. The chosen proxy variable poorly reflects the actual state of affairs. It is indeed possible that the scope of the variable is simply too limited, and thus it is unable to capture the aggregate information accurately. This

however is less probable, given the distribution of variables, which were in accordance with the expectation.

3. There is no relationship between the information without direct relevance to the financial asset.

Chapter 6

Conclusion

Ultimately the aim and focus of this thesis was to explore the relationship between information and financial markets. Efficient Market Hypothesis has been highly influential notion and became entrenched in the core of economic literature and mainstream economic discourse. With review of the literature pertaining to the hypothesis however, it becomes clear that the overall theoretical landscape is highly fractured and inconsistent. Moreover the hypothesis had been formulated in environment highly divergent from the current state of affairs.

Prior to the analysis, it has been hypothesised that increased volumes of information may negatively impact efficiency of financial markets, which would result in higher volatility of returns, as market participants attempt to catch-up to the new information. The results of our analysis however were unequivocal in this regard, invalidating the *a prior* assumption in its entirety; we have utilized a proxy variable to assess an overall volume of information published about a particular asset, and then tested whether the increases in such volumes would affect future returns volatility.

The results could be concluded as a null finding, only one out of the five stocks analysed provided statistically significant results, and given the estimate values, it had to be concluded that either there is no linear relationship between the volumes of information and the volatility of asset returns, or in fact there might be negative correlation, thus increased volumes of information might indicated less volatility. Although the interpretation of such finding could be bent to fit interpretation both for and against the hypothesis, the ambiguity does not allow to reject the hypothesis of market efficiency.

It should be noted that the results of our analysis do not indicate absolutely that markets have not been affected by the exponentially increasing amounts of information being published about financial assets. More likely the proxy variable did not portray the overall informational state of affairs in relation to. There is also increased probability that the relationship between data is non-linear and less general than originally assumed, thus a more complex model could extract improved results in relation to the inefficiency of markets. Both these issues are potentially insightful venues for further research.

The author also briefly outlined a new classification mechanism for information, which ought to be reflected in the analysis, given the transformed 'informational landscape' pertaining to financial assets. And based on such classification the author also proposed a generalized framework for processing new types of information that had not existed during the formation of the hypothesis, which could allow to extract more complex relationships of such new data in relation to asset pricing, utilizing Bayes inference.

Thus future focus of research ought to explore in more detail the possibilities of utilizing the theoretical approaches and techniques outlined in Chapter 4, by employing techniques of machine learning, combined with natural language processing and data mining, in order to analyse and uncover complexities of the underlying data and financial asset prices. Secondly the hypothesis needs to undergo a reconciliation process, by which the definitional aspects would be based on solid theoretical foundation, most notably resolving the duality of the hypothesis, *i.e.* the equilibrium versus fair game model, and resolving the issue of market participant rationality and irrational behaviour. Whether such approach will be possible without introduction of new concepts alike notions of behavioural economics is questionable.

Bibliography

ANDERSON, N. & J. NOSS (2013): "The fractal market hypothesis and its implications for the stability of financial markets." *Bank of England Financial Stability Paper* **(23)**.

BACHELIER, L. (1900): "Théorie de la spéculation." *Annales scientifiques de l'École Normale Supérieure* **17**: pp. 21–86.

BALL, R. (2009): "The global financial crisis and the efficient market hypothesis: What have we learned?" *Journal of Applied Corporate Finance* **21(4)**: pp. 8–16.

BANERJEE, A. V. (1992): "A simple model of herd behavior." *The Quarterly Journal of Economics* pp. 797–817.

BARBERIS, N. & M. HUANG (2006): "The loss aversion/narrow framing approach to the equity premium puzzle." *Technical report*, National Bureau of Economic Research.

BAYLIS, P. (2015): "Temperature and temperament: Evidence from a billion tweets." .

BERK, J. B. (2007): "A critique of the efficient market hypothesis." *Working papers*, University of California.

BONDT, W. F. & R. THALER (1985): "Does the stock market overreact?" *The Journal of finance* **40(3)**: pp. 793–805.

BROWN, K. C., W. V. HARLOW, & S. M. TINIC (1988): "Risk aversion, uncertain information and market efficiency." *Journal of Financial Economics* **22(2)**: pp. 355–385.

CAMPBELL, J., A. LO, & A. MACKINLAY (1997): *The Econometrics of financial markets*. Princeton, NJ: Princeton Univ. Press, 2. print. with corrections edition.

COMMITTEE, N. P. (2013): "Understanding asset prices." *Nobel Prize in Economics documents 2013-1*, Nobel Prize Committee.

CORRADO, C. & B. JORDAN (1997): "Risk aversion, uncertain information, and market efficiency." *Review of Quantitative Finance and Accounting* **8(1)**: pp. 51–68.

DE BONDT, W. F. & R. H. THALER (1994): "Financial decision-making in markets and firms: A behavioral perspective." *Technical report*, National Bureau of Economic Research.

DE LONG, J. B., A. SHLEIFER, L. H. SUMMERS, & R. J. WALDMANN (1990): "Noise Trader Risk in Financial Markets." *Journal of Political Economy* **98(4)**: pp. 703–38.

DE MOOR, L., F. VAN DEN BOSSCHE *et al.* (2013): "A tale of market efficiency." *Review of Business and Economic Literature* **58(2)**: pp. 140–158.

DEGUTIS, A. & L. NOVICKYTE (2014): "The efficient market hypothesis: A critical review of literature and methodology." *Ekonomika* **93(2)**: p. 7.

DENG, M. (2007): "Death of the efficient market hypothesis." In "20th Australasian Finance & Banking Conference," .

Dhaoui, A. (2015): "What does matter in economy today: When human psychology drives financial markets." *Arab Economic and Business Journal* **10(1)**: pp. 39–47.

Ehnts, D. & M. C. Álvarez (2013): "The theory of reflexivity – a non-stochastic randomness theory for business schools only?" *Technical Report 28/2013*, Institute for International Political Economy Berlin.

Etheridge, H. L. & R. S. Sriram (1993): "Chaos theory and nonlinear dynamics: An emerging theory with implications for accounting research." *Journal of Accounting Literature* **12**: p. 67.

Fama, E. F. (1965): "The behavior of Stock-Market prices." *The Journal of Business* **38(1)**: pp. 34–105.

Fama, E. F. (1970): "Efficient Capital Markets: A Review of Theory and Empirical Work." *The Journal of Finance* **25(2)**: pp. 383–417.

Fama, E. F. (1976): "Efficient capital markets: Reply." *The Journal of Finance* **31(1)**: pp. 143–145.

Fama, E. F. (1991): "Efficient capital markets: I.i." *The Journal of Finance* **46(5)**: pp. 1575–1617.

Fama, E. F. (1998): "Market efficiency, long-term returns, and behavioral finance." *Journal of financial economics* **49(3)**: pp. 283–306.

Fama, E. F., L. Fisher, M. C. Jensen, & R. Roll (1969): "The adjustment of stock prices to new information." *International economic review* **10(1)**: pp. 1–21.

Flavin, M. A. (1983): "Excess volatility in the financial markets: A reassessment of the empirical evidence." *The Journal of Political Economy* pp. 929–956.

Flood, R. P. & R. J. Hodrick (1986): "Asset price volatility, bubbles, and process switching." *The Journal of Finance* **41(4)**: pp. 831–842.

French, K. (1988): "Crash testing the efficient market hypothesis." In "NBER Macroeconomics Annual 1988, Volume 3," pp. 277–286. MIT Press.

Gibbons, M. R. & P. Hess (1981): "Day of the week effects and asset returns." *Journal of business* pp. 579–596.

Go, A., R. Bhayani, & L. Huang (2009): "Twitter sentiment classification using distant supervision." *CS224N Project Report, Stanford* **1**: p. 12.

Grossman, S. (1976): "On the efficiency of competitive stock markets where trades have diverse information." *The Journal of finance* **31(2)**: pp. 573–585.

Grossman, S. J. & J. E. Stiglitz (1980): "On the impossibility of informationally efficient markets." *The American economic review* **70(3)**: pp. 393–408.

Gultekin, M. N. & N. B. Gultekin (1983): "Stock market seasonality: International evidence." *Journal of Financial Economics* **12(4)**: pp. 469–481.

Harris, L. (1986): "A transaction data study of weekly and intradaily patterns in stock returns." *Journal of financial economics* **16(1)**: pp. 99–117.

Jensen, M. C. (1968): "The Performance Of Mutual Funds In The Period 1945–1964." *Journal of Finance* **23(2)**: pp. 389–416.

Jensen, M. C. (1978): "Some anomalous evidence regarding market efficiency." *Journal of financial economics* **6(2/3)**: pp. 95–101.

Kahneman, D. & A. Tversky (1979): "Prospect theory: An analysis of decision under risk." *Econometrica: Journal of the Econometric Society* pp. 263–291.

Keynes, J. M. (1936): *The General Theory of Employment, Interest and Money*. Macmillan.

Konstantinidis, A., A. Kataracha, G. Borovas, M. E. Voutsa *et al.* (2012): "From efficient market hypothesis to behavioural finance: Can behavioural finance be the new dominant model for investing." *Scientific Bulletin–Economic Sciences* **11(2)**: pp. 16–26.

Kristoufek, L. (2012): "Fractal markets hypothesis and the global financial crisis: Scaling, investment horizons and liquidity." *Advances in Complex Systems* **15(06)**: p. 1250065.

Kruschke, J. K. (2013): "Bayesian estimation supersedes the t test." *Journal of Experimental Psychology: General* **142(2)**: p. 573.

Kruschke, J. K., H. Aguinis, & H. Joo (2012): "The time has come bayesian methods for data analysis in the organizational sciences." *Organizational Research Methods* **15(4)**: pp. 722–752.

Kudryavtsev, A., G. Cohen, & S. Hon-Snir (2013): "'rational'or'intuitive': Are behavioral biases correlated across stock market investors?" *Contemporary economics* **7(2)**: pp. 31–53.

Lawler, G. F. & V. Limic (2010): *Random walk : a modern introduction*. Cambridge studies in advanced mathematics. Cambridge, New York: Cambridge University Press. Table des matières http://www.loc.gov/catdir/enhancements/fy1009/2010017817-t.html.

LeRoy, S. F. (1973): "Risk aversion and the martingale property of stock prices." *International Economic Review* pp. 436–446.

LeRoy, S. F. (1989): "Efficient capital markets and martingales." *Journal of Economic Literature* **27(4)**: pp. 1583–1621.

Leuchtkafer, R. T. (2015): "High Frequency Trading, a bibliography of evidence-based research." *Working Papers 1*, Finance Watch.

Lewellen, J. & J. Shanken (2000): "Estimation risk, market efficiency, and the predictability of returns." *Technical report*, National Bureau of Economic Research.

Lo, A. W. (2004): "The adaptive market hypothesis: Market efficiency from an evolutionary perspective." *Journal of Portfolio Management* **30 (2004)(15-29)**.

Lo, A. W. (2005): "Reconciling efficient markets with behavioural finance: The adaptive markets hypothesis." *Journal of Investment Consulting* **7(2)**: pp. 21–44.

Lo, A. W. (2007): "Efficient markets hypothesis." .

Lo, A. W. & A. C. MacKinlay (1987): "Stock market prices do not follow random walks." .

Lucas Jr, R. E. (1978): "Asset prices in an exchange economy." *Econometrica: Journal of the Econometric Society* pp. 1429–1445.

Malkiel, B., S. Mullainathan, & B. Stangle (2005): "Market efficiency versus behavioral finance." *Journal of applied corporate finance* **17(3)**: pp. 124–136.

MALKIEL, B. G. (1995): "Returns from investing in equity mutual funds 1971 to 1991." *The Journal of Finance* **50(2)**: pp. 549–572.

MALKIEL, B. G. (2003): "The Efficient Market Hypothesis and Its Critics." *Working Papers 111*, Princeton University, Department of Economics, Center for Economic Policy Studies.

MALKIEL, B. G. (2005): "Reflections on the efficient market hypothesis: 30 years later." *Financial Review* **40(1)**: pp. 1–9.

MANKIW, N. G., D. ROMER, & M. D. SHAPIRO (1985): "An unbiased reexamination of stock market volatility." *The Journal of Finance* **40(3)**: pp. 677–687.

MANKIW, N. G., D. ROMER, & M. D. SHAPIRO (1991): "Stock market forecastability and volatility: a statistical appraisal." *The Review of Economic Studies* **58(3)**: pp. 455–477.

MANSHARAMANI, V. (2011): *Boombustology: spotting financial bubbles before they burst.* John Wiley & Sons.

MERTON, R. (1985): "On the current state of the stock market rationality hypothesis." *Working papers 1717-85.*, Massachusetts Institute of Technology (MIT), Sloan School of Management.

MICHAILIDOU, D. M. (2015): *The Inexorable Evolution of Financialisation: Financial Crises in Emerging Markets.* Palgrave Macmillan, 1 edition.

NORSTAD, J. (2005): "Random walks."

OHLSON, J. A. (1977): "Risk-aversion and the martingale property of stock prices: Comments." *International Economic Review* pp. 229–234.

PAK, A. & P. PAROUBEK (2010): "Twitter as a corpus for sentiment analysis and opinion mining." In "LREc," volume 10, pp. 1320–1326.

PESARAN, M. H. (2010): "Predictability of asset returns and the efficient market hypothesis." .

PETERS, E. E. (1994): *Fractal market analysis: applying chaos theory to investment and economics*, volume 24. John Wiley & Sons.

PHILLIPS, P. C. & S. JIN (2014): "Testing the martingale hypothesis." *Journal of Business & Economic Statistics* **32(4)**: pp. 537–554.

POPPER, K. (2002): *The Logic of Scientific Discovery.* Routledge Classics. Taylor & Francis.

RITTER, J. R. (2003): "Behavioral finance." *Pacific-Basin Finance Journal* **11(4)**: pp. 429–437.

SAMUELSON, P. A. (1965): "Proof that Properly Anticipated Prices Fluctuate Randomly." *Industrial Management Review* **6**: pp. 41–49.

SAMUELSON, P. A. (1998): "Summing up on business cycles: opening address." *Conference Series ; [Proceedings]* **42(Jun)**: pp. 33–36.

SCHOOT, R., D. KAPLAN, J. DENISSEN, J. B. ASENDORPF, F. J. NEYER, & M. A. AKEN (2014): "A gentle introduction to bayesian analysis: applications to developmental research." *Child development* **85(3)**: pp. 842–860.

SEWELL, M. (2007): "Behavioural finance." *University of Cambridge. Available from internet: http://www. behaviou ralfinance. net/behavioural-finance. pdf* .

SEWELL, M. (2011): "History of the efficient market hypothesis." *RN* **11(04)**: p. 04.

Shiller, R. J. (1980): "Do stock prices move too much to be justified by subsequent changes in dividends?"

Shiller, R. J. (1981): "The Use of Volatility Measures in Assessing Market Efficiency." *Journal of Finance* **36(2)**: pp. 291–304.

Shiller, R. J. (2003): "From efficient markets theory to behavioral finance." *The Journal of Economic Perspectives* **17(1)**: pp. 83–104.

Shiller, R. J. (2015): *Irrational exuberance*. Princeton university press.

Shleifer, A. & L. H. Summers (1990): "The noise trader approach to finance." *The Journal of Economic Perspectives* **4(2)**: pp. 19–33.

Soros, G. (1994): "Speech to the mit department of economics world economy laboratory conference." [Accessed: 2015 11 20].

Soros, G. (2008): *The New Paradigm for Financial Markets: The Credit Crisis of 2008 and What It Means*. PublicAffairs.

Summers, L. H. (1982): "Do We Really Know That Financial Markets Are Efficient?" *NBER Working Papers 0994*, National Bureau of Economic Research, Inc.

Summers, L. H. (1985): "On economics and finance." *The Journal of Finance* **40(3)**: pp. 633–635.

Summers, L. H. (1986): " Does the Stock Market Rationally Reflect Fundamental Values?" *Journal of Finance* **41(3)**: pp. 591–601.

Thaler, R. H. (2010): "The end of behavioral finance." .

Thomas, W. I. (1923): "The unadjusted girl: With cases and standpoint for behavior analysis." *Criminal science monographs* **(4)**: pp. 1–257.

Timmermann, A. & C. W. Granger (2004): "Efficient market hypothesis and forecasting." *International Journal of Forecasting* **20(1)**: pp. 15–27.

Tseng, K. (2006): "Behavioral finance, bounded rationality, neuro-finance, and traditional finance." *Investment Management and Financial Innovations* **3(4)**: pp. 7–18.

Tversky, A. & D. Kahneman (1974): "Judgment under uncertainty: Heuristics and biases." *science* **185(4157)**: pp. 1124–1131.

Vuolteenaho, T. (2002): "What drives firm-level stock returns?" *The Journal of Finance* **57(1)**: pp. 233–264.

Walia, M., D. M. R. Kiran *et al.* (2010): "Efficient market hypothesis, price volatility, and performance of mutual funds." *Global Journal of Management And Business Research* **10(2)**.

Zeelenberg, M. (1999): "Anticipated regret, expected feedback and behavioral decision making." *Journal of behavioral decision making* **12(2)**: p. 93.

Appendix A

Analysis Outputs

A.1 Proxy Variable Table

Table A.1: Information quantity index

Period	Microsoft	Wells Fargo	AT & T	Marriott International
01.01.2010	13900	1110	3160	61
01.02.2010	13500	1160	3080	47
01.03.2010	16400	1410	3800	61
01.04.2010	15500	1270	3760	80
01.05.2010	15800	1310	3830	61
01.06.2010	16300	1400	4110	82
01.07.2010	16700	1730	4210	81
01.08.2010	16200	1700	4120	55
01.09.2010	18500	1850	4340	77
01.10.2010	23000	2610	5720	101
01.11.2010	17800	2090	4330	88
01.12.2010	19400	2140	4390	68
01.01.2011	20500	2230	4920	84
01.02.2011	19500	2320	4980	87
01.03.2011	21100	2830	5400	122
01.04.2011	21100	2570	5420	102
01.05.2011	27500	3510	6450	129
01.06.2011	23800	2810	5120	140
01.07.2011	23200	2880	5300	95
01.08.2011	24600	2840	5160	100
01.09.2011	25400	3150	5010	111
01.10.2011	23700	3160	5570	116
01.11.2011	26500	3090	4960	130
01.12.2011	24100	3020	4700	141
01.01.2012	23800	3490	5140	116
01.02.2012	25700	3520	5370	153
01.03.2012	29100	3810	5840	173
01.04.2012	28800	4000	6000	193
01.05.2012	36300	4770	8490	241
01.06.2012	29700	4050	6650	199
01.07.2012	29600	4160	6000	219
01.08.2012	30900	4540	6710	244
01.09.2012	32300	4360	6600	358
01.10.2012	35300	4600	6060	351
01.11.2012	33700	5020	6320	354
01.12.2012	31300	4490	6920	313

Table A.2: Information quantity index – cont.

Period	Microsoft	Wells Fargo	AT & T	Marriott International
01.01.2013	34600	5730	7360	368
01.02.2013	32300	5450	7860	387
01.03.2013	34900	6100	7540	422
01.04.2013	35800	5940	7930	417
01.05.2013	49300	7500	10300	598
01.06.2013	40600	6190	7120	479
01.07.2013	40800	6490	8040	448
01.08.2013	43400	7690	9000	449
01.09.2013	48700	7500	8650	497
01.10.2013	53600	8550	10700	602
01.11.2013	50600	8590	9770	535
01.12.2013	43500	7300	9750	443
01.01.2014	56300	9080	9840	550
01.02.2014	51900	9040	9990	557
01.03.2014	55200	9440	11200	650
01.04.2014	57700	9980	11200	652
01.05.2014	81500	12700	15300	787
01.06.2014	69300	10800	12800	599
01.07.2014	67900	11600	11800	645
01.08.2014	68800	11400	13700	582
01.09.2014	75400	12100	14400	749
01.10.2014	77700	13400	16100	782
01.11.2014	74400	13200	13600	727
01.12.2014	90400	16500	18800	890
01.01.2015	101000	16300	18500	839
01.02.2015	92100	17200	18500	787
01.03.2015	104000	19500	19800	960
01.04.2015	104000	19500	19800	900
01.05.2015	125000	19700	20800	920
01.06.2015	121000	18500	19800	948
01.07.2015	208000	20000	21600	975
01.08.2015	126000	21900	22500	822
01.09.2015	247000	23800	28200	987
01.10.2015	376000	24500	25200	1200
01.11.2015	275000	27200	27200	1630
01.12.2015	1620000	25500	24900	1140

A.2 Data Plots

Figure A.1: Microsoft – Regression

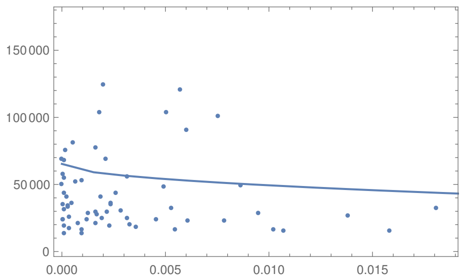


Figure A.2: Microsoft – Residuals

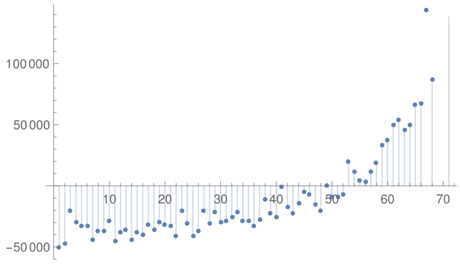


Figure A.3: Wells Fargo – Regression

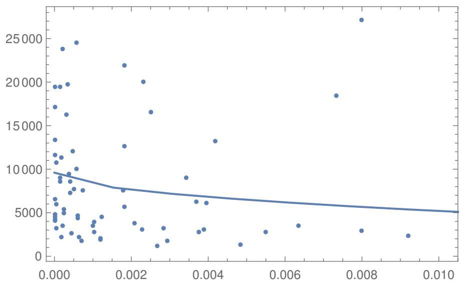


Figure A.4: Wells Fargo – Residuals

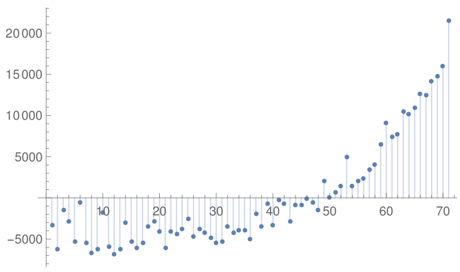


Figure A.5: AT&T – Regression

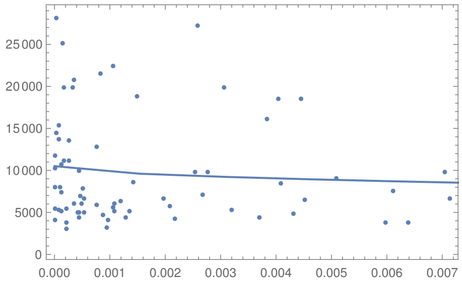


Figure A.6: AT&T – Residuals

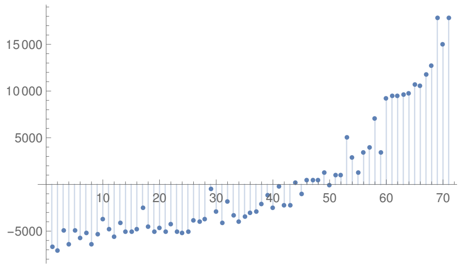


Figure A.7: Marriott International – Regression

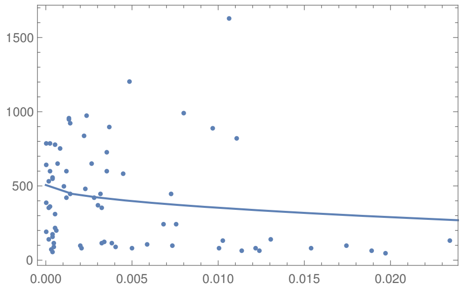
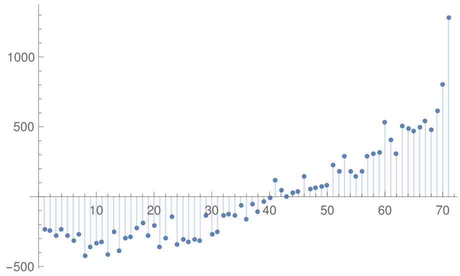


Figure A.8: Marriott International – Residuals



Source: Author's computations.

Appendix B

Source Codes

Below is the basic source code in Mathematica, used to estimate models. The sources and data files have been also enclosed digitally.

General placeholders were introduced in place of an function parameters, where appropriate.

```
data = Import["PATH", "CSV"];
clean = Delete[data[[All, 1]], 1];
clean = Delete[clean, 72];
return = Differences[
   Log[FinancialData["STOCK", {{DATE1}, {DATE2}, "Month"},
     "Value"]]];
returnavg = Mean[return];
return = (# - returnavg)^2 & /@ return;
descriptives = {Mean[return], Variance[return], Kurtosis[return],
   Max[return], Min[return], StandardDeviation[return]};
regress = Transpose[{return, clean}];
ListPlot[return]
model = LinearModelFit[regress, {Sqrt[x] }, x];
model["BestFit"]
model["AdjustedRSquared"]
Show[ListPlot[regress], Plot[model[x], {x, 0, 5}], Frame -> True]
model["ParameterTable"]
Correlation[return, clean]
ListPlot[clean]
ListPlot[model["FitResiduals"], Filling -> Axis]
```