Copenhagen Business School

Master Thesis

Cand. Merc. International Business

FANG stocks - *Investigating the factors driving the stock prices*

Supervisor:

Björn Preussen

Authors:

Kasper Thrane Kofoed-Dam (47745)

Giulia Giorgi (107724)

Pages (characters with spaces):

111 pages (219.517)

Hand-in Date:

May 15th, 2018

Abstract

The purpose of this paper has been to find macro-economic and accounting factors that impact the stock prices of the FANGs, which represent today's leading Internet companies. First, a price index and share weights were calculated by applying the share-weighted index method. Subsequently, the macro-economic and accounting factors' impact on share prices was tested by OLS regression. The analysis found that none of the macro-economic factors (oil price, CPI, M1 and M2) were significant, while both of the accounting factors (market-to-book value and earnings-per-share) were found to be significant although the results for earnings-per-share were unstable. The model developed for the FANG stocks was tested on two other samples containing data for respectively today's biggest Chinese Internet companies and the biggest Internet companies of the pre-dotcom bubble. The results for the two other samples supported the findings based on the FANG sample. The main finding, which is supported across all samples, is that market-to-book-value has a high impact on Internet stock prices and thus provides a metric that cannot be ignored by both investors and managers. This paper contribute to the existing literature as the impact of macro-economic and accounting factors on the Internet stock prices has not previously been researched. It is the hope of the authors that this paper can lead to future research, which will examine the impact of a broader set of variables on a broader segment of Internet stocks. This could enable research on longer time series and thereby lead to new interesting findings.

Table of content

Table of figures

List of Abbreviations[1]

Abbreviation	Explanation
CH	Chinese tech stocks
EPS	Earnings-per-share
FANG	Acronym for Facebook, Amazon, Netflix and Google
GDP	Gros Domestic Product
MTVB	Market-to-book value
P/B	Market-to-back ratio
P/E	Price-to-earnings ratio

---

[1] Abbreviations are reported in alphabetic order

1. Introduction

Not a day goes by without checking Facebook, “googeling” for information and watching the latest series on Netflix. Not a week goes by without checking the price and reviews of a product on Amazon or downloading a new book for the Kindle. In the news, the headlines are about Facebook’s Cambridge Analytica scandal, its new dating feature for long-term relationships, Alphabet[2]’s autonomous cars and Amazon’s parcel delivery by drones. When exposed to these companies, we see the technology of today, which impacts all of us, while we also get a sneak peak of what the future will bring.

Collectively, these companies are known as the FANGs: Facebook, Amazon, Netflix and Google. The term was coined by Jim Cramer in 2013, and the stocks that are grouped together represents “stocks [that] have the potential to really take a bite out of the bears (Brodie, 2013).”. Since the term was coined in 2013, the price of Google has nearly doubled, the price of Facebook and Amazon has more than tripled, while Netflix’ price has increased six-fold. With a trailing P/E[3] value for some of the companies exceeding exorbitant 250, the price relative to earnings by far exceed what is considered sustainable in the long run (Risager, 2016).

The FANGs are all a part of a group that can be defined as Internet stocks. A group that flourished during the 1990s and since have fostered many stocks which are highly valuated and where the expectations of the future returns are extremely high, and the expectations for growth opportunities seems limitless (Hand, 2001) (Deloitte, 2018). The FANGs have not only become highly valued but have developed from opportunistic growth companies to market dominators in the western world.

Companies have also risen in the Middle Kingdom where it is instead common to check WeChat, go to BaiDu to search for information or hear about the Apollo project that is a serious contender to becoming the leader within autonomous cars. You watch the latest videos on iQiyi and shop on Alibaba or JD.com. These are all services provided by Chinese companies that seek to compete

---

[2] The parent company of Google

[3] Trailing indicate that the earnings component in the P/E ratio is calculated as the earnings of the most recent fiscal year

with the FANGs, in particular on the Chinese market, and some of the companies list among the biggest companies in the world (PWC, 2017).

1.1 Background

The subsequent section will briefly present each of the companies that are included in the FANGs. As the nature of the companies will not be analyzed in this thesis, the purpose of this section is to give the reader an elementary understanding of each of the companies.

Facebook

Facebook, Inc. is operating in the business of online social media and social networking with a mission to "…give people the power to build community and bring the world closer together" (Facebook, 2018). The main brands of Facebook Inc. are Facebook, WhatsApp, Facebook Messenger, Instagram and Oculus. Facebook has 2.2 billion monthly active users and the net income of Facebook Inc. for 2017 amounted to 10.217 million USD (Facebook, 2018).

Amazon

Amazon.com, Inc. is in the business of electronic commerce and cloud computing and has a vision to be "[the] earth's most customer-centric company; to build a place where people can come to find and discover anything they might want to buy online." (Amazon.com, 2018). In addition to the main service of being an online marketplace for books, video streaming, consumer electronics, groceries etc., another major service area of Amazon.com is Amazon Web Services (AWS) which offers a broad range of services within cloud computing. The net income of Amazon.com, Inc. for 2017 amounted to 3.033 million USD (Amazon.com, 2018)

Netflix

Netflix, Inc. is in the business of internet entertainment service and focuses on delivering and producing movie and TV series entertainment. The service has 125 million members across 190 countries. The net income of Netflix, Inc. for 2017 amounted to 558 million USD

Alphabet

Alphabet, Inc. is a conglomerate with Google LLC as its main subsidiary which has the mission to “Organize the world’s information and make it universally accessible and useful” (Google Inc., 2018). Alphabet’s main business is advertising, sales of digital content, applications, cloud offering and sales of hardware products. However, the company is also investing in projects within infrastructure, data management, analytics and AI. The net income of Alphabet, Inc. for 2017 amounted to 12.662 million USD.

1.2 Problem statement

The main subject of this paper, the FANGs, came into play early on in the process of considering potential topics. We are all exposed to the FANGs on a daily basis whether we hear about them in the news or we use the services that they provide. Each of the companies have been pioneers within their field of business and most people have an opinion about the concept and content on Netflix or the latest changes to the Facebook platform; but despite hearing so much about the companies and using their service, there is not spend much time considering what is actually driving the stock value of these companies.

We hear about a new tech bubble, that the prices of the stocks will suddenly fall. We know from the media that bad news can lead to significant drops in share prices while good news causes the price to spike. This made us wonder, do the conventional stock price drivers still apply to this breed of stocks?

Searching for an answer to this question, we quickly realized that an abundance of literature exists on which macro-economic factors drive the stock prices in general and also on which accounting factors have an impact on the stock prices. However, when searching for literature that apply this knowledge and test if the same factors drive the price development of Internet companies, no literature was found. The fact that no literature existed on whether the traditional stock price drivers apply to internet stocks, led to the central research question of this paper:

To which extend is it possible to find factors that explain the price development of today’s biggest Internet stocks, the FANGs, by multiple linear regression?

To help answer the central research question, the following sub-questions will be answered:

- *Which macro-economic and accounting factors are driving the prices of the FANGs?*
- *To which degree are the findings generalizable to other leading Internet stocks of our time, as well as to the leading pre-dotcom bubble Internet stocks on NASDAQ?*

The aim of the thesis is to build upon the existing literature by testing if factors that have been found to be driving stock prices in general are also significant drivers of the stock prices of leading internet stocks, in particular the FANGs.

1.3 Thesis structure

This research is divided into four distinct sections: *introduction*, *theory*, *methodology* and *findings* and *culminating section*. Each section contains one or two main chapters, which again include a set of subchapters. Figure 1 outlines the structure of the thesis; the grey boxes on the left represent the *sections*, the white rectangles denote the *chapters,* and finally the grey rounded rectangles refer to the *subchapters.*

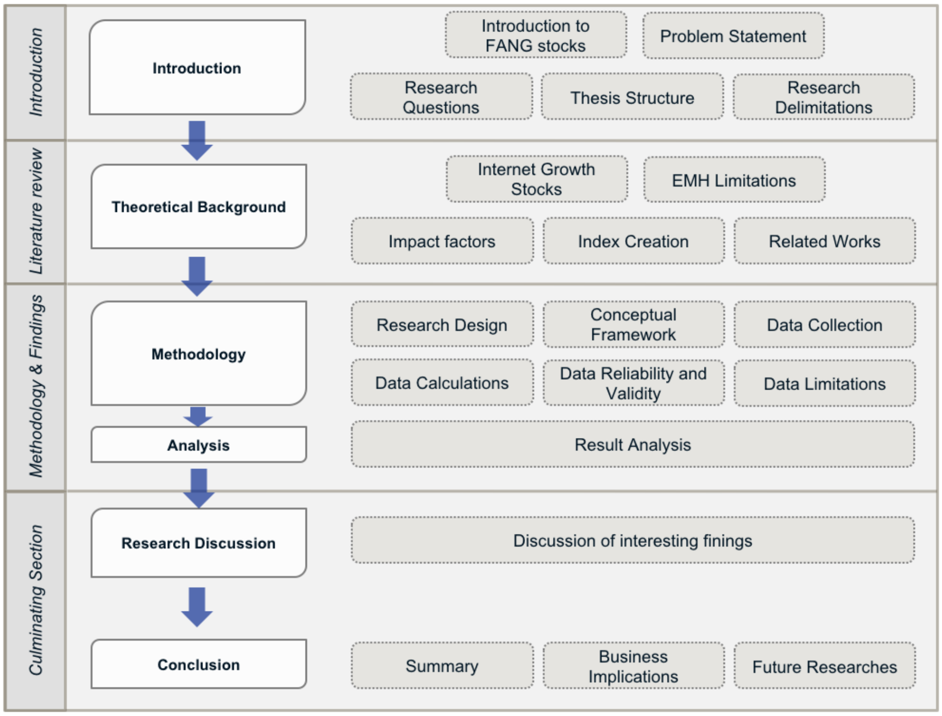


Figure 1 - Thesis Structure. Source: Personal production

The structure of this thesis is meant to provide a transparent structure which makes it easy for the reader to follow. The first section sets the ground for a comprehensive presentation of the theme of the research (*Introduction to FANG stocks*), followed by an introduction to the research investigation and scope (*Problem Statement; Research Questions*). The introduction also includes the thesis structure and the *research delimitations*.

The second section outlines the literature review that supports the research topic. The authors decided first to present the concept of *growth stocks*, and how this type of stocks is associated to modern *Internet technology stocks* i.e. FANG stocks. Afterwards, the theory of efficient market hypothesis is presented. To this extent, the authors' focus is not only on providing the definition but also on emphasizing the limitations that the theory has in real applications. Afterwards, macroeconomic and financial accounting factors will be presented, with the intent to outline what variables affect stock valuation in the past literature. Then, the theory related to the creation of the index is presented. The last subchapter proves a term of comparison by showing a literature of related works.

The third section presents the *methodology*. The methodology is meant to show the readers all the steps that the researches have made in order to achieve the results of their analysis. The purpose of the methodology is two-fold: first, it must clearly show the consecutive steps that the researchers have performed. Second, it must ensure the replicability to a third party who is willing to reproduce the analysis by its own. Subsequently, the *analysis* of results is presented in the fourth section.

The last section also contains two chapters which respectively refer to the *discussion* and the *conclusion*. While the discussion is meant to outline critical considerations related to the topic, the conclusion is intended to summarize and provide additional food for thought for future researches.

1.4 Research Delimitations

This work presents three conceptual delimitations, which refer to what the authors have decided to *exclude* from the scope of the research.

The *first* delimitation concerns the nature of the topic of interest. In this regard, the authors have chosen to limit the scope of the research to the FANG stocks i.e. Facebook (FB), Amazon (AMNZ), Netflix (NFLX) and Alphabet (GOOGL). Recently, some economists also discussed and wrote about FAANGs, incorporating Apple (AAPL) (Mauldin, 2018). Additionally, others refer to FAANMGs, including Microsoft (Oyedele, 2017). In this regard, the authors have chosen to take the original acronym “FANG” coined by Jim Cramer in 2013 (Brodie, 2013).

The main reason behind the choice to limit the research to the *FANG* stocks relies on the fact that Facebook, Amazon, Netflix and Google all are service-provider companies which are relatively young and revolutionary in their respective industry. Their peculiarity is expressed by the implementation and the execution of innovative business models that have disruptively transformed the respective markets.

In this regard, the authors have decided to not include the acronym containing Microsoft and Apple. Microsoft is a well-established company that is highly technological. However, compared to the FANG companies, the company founded by Bill Gates is more traditional as its core business has been established in the early 1990s. On the other hand, Apple Inc. presents a slightly different business classification compared to FANG companies. In fact, the core business of Apple focuses on the manufacturing of hardware devices i.e. iPhone, iPad and Macintosh. The service provision of Apple is exclusively reserved to its specific products.

According to the authors of this paper, Apple and Microsoft share fewer similarities than what the FANG companies share between them. Therefore, the research will exclusively focus on the factors that have an impact on the stocks of Facebook, Amazon, Netflix and Alphabet.

The *second* delimitation is strictly linked to the first one and refers to the choice of the authors to consider the four companies as one entity. In other terms, the authors do not assess the business and financial characteristics of each individual company.

The *third* and last delimitation refers to the dataset and the time period that has been taken into consideration. The authors took the period starting on 01/06/2012 and ending on 31/12/2017 for the FANGs and the top-four-performing Chinese stocks, and the period starting on 01/06/1994 and ending 31/12/1999 for the top-four U.S. technology stocks prior the dotcom bubble. The decision regarding the first time period is based on when the last stock of the FANGs was listed.

2. Literature Review

The general purpose of the literature review is to present a summary of what is acknowledged about a specific topic on the basis of established research evidence (Hammersley, et al., 2011).

The authors have decided to divide this literature in five main focus areas: *growth stocks*, *the efficient market hypothesis (EMH*), *accounting and macroeconomic factors*, *index creation* and the *econometric model*. Each chapter contains one or more sub-chapters. The primary objective of this literature is to provide a theoretical background to the research.

2.1 Internet stocks; a special kind of tech stocks belonging to the growth stock family

2.1.1 Defining growth stocks and growth companies

Financial experts define *growth stocks* as specific securities of companies that are believed to offer strong earnings growth potential in the near future (AXA Equitable Life Insurance Company, 2018). Growth stocks are categorized as having three main characteristics; first, they have high earnings-per-share (EPS) growth. Second, they have a high market price appreciation, which could be measured by both market-to-book ratio (P/B) and price-per-share (P/E). Third, they entail higher risks as compared with value stocks (Bauman, et al., 1998). In this regard, Modigliani and Miller (1961) theoretically classify growth stocks as "riskier than non-growth stocks" i.e. value stocks (Modigliani & Miller, 1961) (pp.425). They argue that growth stocks entail more risk for investors because their worth is highly linked to the uncertainty attached to future length of coming growth opportunities (Modigliani & Miller, 1961). As a matter of fact, it can be determined that the price of growth stocks is primarily estimated on the basis of investors' sentiment and projections (Hillier, et al., 2013).

The primary criteria that determines the nature of growth stocks is given by the price-to-earnings ratio, P/E (AXA Equitable Life Insurance Company, 2018). P/E is quantified as the price of a stock divided by the current year's earnings per share (Hillier, et al., 2013). Smidt (1968) argues that an aggregated sensitive investor reaction to specific high P/E ratios might potentially generate an exaggerated expectation towards the future growth in earnings and dividends of particular securities. Smidt (1968) continues saying that exuberant investor optimism might lead to an average overvaluation of P/E ratios of some stocks (Smidt, 1968). De Bondt and Thaler (1985) argue that the P/E ratio represents a powerful proxy in the valuation of stocks, as a good valuation of P/E ratio would eliminate the incorrect anomalies in security price assessment. On the contrary, an unrealistic assessment of P/E ratio would eventually lead to stock price inconsistencies in the market (De Bondt & Thaler, 1985).

Growth stocks are often highly innovative and highly profitable companies i.e. technology companies (Hillier, et al., 2013). Székely and Richards (2012) specifically define growth companies as organizations "whose revenues are growing significantly faster than overall the economy" (pp.229). From a historical perspective, growth companies commonly manage to achieve outstanding returns despite economic fluctuations (AXA Equitable Life Insurance Company, 2018). That is, *growth* companies are perceived as having the right potential to achieve high and lasting future earnings opportunities (Hillier, et al., 2013).

2.1.2 Growth stocks vs value stocks

Scholars and investment professionals have always debated the antagonism between value and growth strategies. Traditionally, value stocks are associated with a low market price and low earnings-per-share (Bauman, et al., 1998). In particular, Lakonishok et al. (1994) define *value stocks* as those financial securities that present three main characteristics: first, low price-to-earnings ratio; second, negative investor expectations on future growth and finally, lower prices with respect to peer-marketed securities (Lakonishok, et al., 1994). Typically, lower prices and growth expectations might be caused by negative investor reactions to underperformances of the company i.e. unsatisfactory returns, legal issues or negative reputations (Hillier, et al., 2013). Generally speaking, Lakonishok et al. (1994) define value strategies as the transaction of buying stocks that have low P/E ratio and price-to-dividends ratio. As compared with growth stocks, value stocks attract less investor attention because of low returns on their investments (Bauman, et al., 1998).

Investors seeking for growth securities prioritize high-quality, successful companies that have strong past performance, which feeds confident investor expectations (Risager, 2016). The risk of buying growth stocks consists in a sudden sharp decline of their expensive prices (Petkove &

hang, 2005). On the contrary, investors seeking for value stocks are focused on companies that have underperformed but possess sound fundamentals (Hillier, et al., 2013). The concept supporting value investing strategies includes the assumption that good companies are more likely to worsen in the future when the actual intrinsic value is generally understood by a large number of investors (Hillier, et al., 2013).

For many years, experts have sustained that value strategies generally perform better than growth investment strategies (Graham & Dodd, 2009). In this regard, Miller and Bauman (1997) find that value strategies are very likely to perform better than growth strategies in the long-run. In other words, the earnings-per-share growth rate has a reversed mean trend over time. It follows that, low growth associated with value securities tend to accelerate and outperform the high growth rates linked to growth stocks. Miller and Bauman's findings are built upon past works. For instance, De Bondt and Thaler (1985) and Fama and French (1993) have previously proved that value stocks generally earn higher average returns than growth stocks.

However, while there is overall consensus that value strategies generate average superior returns over growth strategies, there is little clarification on the drivers behind this (Lakonishok, et al., 1994).

Although representing two opposite strategies, growth and value investments should be perceived as complementary, rather than exclusive (Risager, 2016). The discussion between growth and value has been going on for decades, each faction has always been able to present statistics to support their case (Risager, 2016). Instead of choosing one-sided investing strategies, Risager (2016) recommends that astute investors should always find the way to manage investing risks by converging both growth and value strategies in their portfolios.

2.1.3 Internet Growth Stocks

From the establishment of the internet industry in the mid-1990s, Internet companies are commonly associated with two characteristics: high market value and extremely positive growth expectations (Hand, 2001). For these reasons, technology Internet companies can be theoretically associated to growth companies (Deloitte, 2018). Like growth companies, Internet technology companies are, by investors, believed to be a source of limitless growth opportunities, and therefore, the value of their stocks tend to be extremely high and differ from the value that fundamentals justify. (Deloitte, 2018)

An attempt is made by Trueman et al. (2000) in valuating internet stocks by associating their market values with fundamental accounting information e.g. paid-out dividends and the P/E ratio. Their study reports that Internet stocks are traded at prices that are not necessarily aligned to the companies' accounting information (Trueman, et al., 2000). In this perspective, Schonfeld (2000) argues that internet stock valuations lack on the established financial metrics that Wall Street has traditionally used to evaluate companies in the market (Schonfeld, 2000). Aligned to Schonfeld's argumentation, Trueman et al. (2000) debate that the value and stock prices of Internet companies are commonly assessed by weighting more non-accounting factors over quantitative financial data (Trueman, et al., 2000). For instance, in year 2000, Amazon stocks were priced $130 per share rather than $30 per share because the investors' attraction to the company won over traditional financial analysis (Trueman, et al., 2000).

Trueman et al. (2000) argue that Internet stocks are extremely difficult to value for two main reasons. First, the Internet industry is evolving at an extremely high rate. It follows that stock analysis accuracy weakens. Second, the Internet industry is relatively young and dynamic, and therefore, growth history does not represent a pivotal factor in assessing future predictions. Furthermore, the Internet industry generates an intense amount of non-financial data that have a heavy impact on the investors' judgment of the internet stocks (Trueman, et al., 2000).

2.2 The EMH and how the theory breaks when dealing with growths

2.2.1 The Efficient Market Hypothesis (EMH) definition

The claim that "stock markets are efficient" is a fundamental part of financial economics (Summers, 1986). According to the theory of the Efficient Market Hypothesis (EMH), the market is defined as *efficient* when all significant information is fully reflected in current stock prices (Fama, 1970). In this regard, Modigliani and Miller (1961) argue that in an efficient market, the valuation of securities should be ruled by a fundamental principle which states that "the price of each share must be such that the rate of return on every share will be the same throughout the market over any interval time" (pp. 412). Under this principle, three main assumptions hold:

perfect capital markets, rational behavior of investors and perfect certainty (Modigliani & Miller, 1961). The EMH theory is also proved by Jensen (1978), who has challenged, tested and confirmed the Efficient Market Hypothesis for several different market cases, including the New York and American Stock exchanges, the English, the Australian and the German stock markets (Jensen, 1978).

Three are the main conditions that cause market efficiency (Shleifer, 2000). Fist, the *rationality* condition implies that all investors in the market are rational, and consequently, when information is released in the market, all investors are able to absorb the news and adjust their expectations on securities in a rational and logical way (Risager, 2016). Second, the *independent deviations from rationality* rely on the assumption of "offsetting irrationalities" (Hillier, et al., 2013) (pp. 355). In this perspective, the market is populated by as many irrational optimistic investors as irrational pessimistic investors. It follows that although the market is led by irrational investors, the two sides end up offsetting each other; and therefore, the status of the market remains inviolably "efficient" (Shleifer, 2000). Finally, the last condition denotes the concept of *arbitrage* which represents the set of professionals in charge of correcting abnormal stock prices due to the speculation of investors. Assuming that arbitrage is effective, the market would still be efficient (Hillier, et al., 2013).

The Efficient Market Hypothesis has two main implications: one related to investors and one associated with companies (Hillier, et al., 2013). The first implication affirms that information is reflected in the prices immediately and that prices adjust before the investors have time to trade and make profits on it. The second implication suggests that companies should receive a fair (par value) for selling stocks, and thus, firms should not be able to make a profit out of the selling transaction. Both implications relate to the starting point that unconditioned information disclosure is granted and available in an efficient market (Fama, 1970).

2.2.2 The flaws of the Efficient Market Hypothesis

Generally speaking, there is consensus within academia that the principle about the market not always being efficient is true, and those who believe it to be are naïve and imprudent investors (Risager, 2016). In fact, several economists like Keynes (1936), Smidt (1968), Shiller (1980), Buffet (1989) and Samuelson (1998) attribute the lack of market efficiency primarily to investor sentiments. In particular, Smidt (1968) argues that market inefficiency is due to an *inappropriate* market reaction to a new information released (Smidt, 1968). Consistent with Smidt's discovery, Arrow (1982) contributes by adapting the psychological concept of "irrational decision making". The "irrational decision making" was applied to financial economics by Tversky and Kahman (1981) who found that irrationality in the market can describe a large proportion of the speculative markets. Similarly, Buffet (1989) says that "observing correctly that the market was *frequently* efficient, they went on to conclude incorrectly that it was *always* efficient. The difference between these propositions is night and day" (Buffett, 1989) (pp. 18). Samuelson (1998) contributes to the argumentation, reporting that another fundamental reason why markets are sometimes mispriced is because there is a lack of a self-correcting rational mechanism that drive price to the right levels.

Mispricing anomalies can last for a long period of time, as it is proved by Keynesian economists. Keynes and the academic circle that supported Keynesian ideas argue that the market could stay overvalued much longer than what fundamentals allow by theory. As a consequence, if *happy times* last for several years in the market, investors tend to generally become more optimistic and overconfident and perhaps they would also start believing "this time it would be different" (Risager, 2016) (pp.24). In other terms, Keynes (1936) argues that in a stock market where share prices are traded at very high prices for a long time, investors' expectations tend to be driven by an emotional belief that prices are fair, while forgetting the mathematical fundamentals of financial economics (Keynes, 1936). This fact would eventually lead to investor disillusionment, which would rapidly create a radical shift in expectations and actions (Keynes, 1936). In the event that such changes happen in the broader market, the general market outlook could change and take the economy into recession (Risager, 2016).

In a broader context, Lamont and Thaler (2003) sustain that companies, which trade their stock prices beyond economic fundamentals and their intrinsic value, should attract too much capital which would ultimately lead to a bubble outbreak. Although particularly significant, this proposition related to the EMH is challenging and difficult to prove, as intrinsic value is neither quantifiable nor statistically observable. What is measureable and testable is the relative valuation of overvalued stocks compared to other similar stocks in the market (Lamont & Thaler, 2003).

Other literatures find weaknesses of the EMH in the relationship between financial indicators i.e. macroeconomic fundamentals and stock prices (Kumar & Puja, 2012). The Efficient Market Hypothesis assumes that macroeconomic changes are instantaneously incorporated into stock prices and consequently changes of any macroeconomic variables should not influence the performance of stock prices (Kumar & Puja, 2012).

Along with an extensive literature supporting the EMH theory, several scholars affirm that macroeconomic variables might influence the performance of stock returns by affecting stock prices (Kumar & Puja, 2012). Fama and Schwert (1977) estimate the extent to which the price of the securities in the market are influenced by specific financial indicators i.e. inflation. The results suggest that macroeconomic variables do affect returns from stocks by acting through stock prices (Fama & Schwert, 1977). Modigliani and Cohn (1979) attributed undervaluation of the securities market to inflation, which misleads the investors assessments by leading stockholders to commit mistakes when evaluating i.e. wrong capitalization in inflationary periods (Modigliani & Cohn, 1979). Shiller (1980) affirms that the only condition to hold the EMH is that investors must assume that forecasted dividends differ from their historical trend much more than what they have done in the past (Shiller, 1980). Kumar and Puja (2012) empirically investigate the impact of macroeconomic fundamentals on stocks in the specific case of the Indian security market. The study highlights a long-term correlation amongst tested financial indicators and Indian stocks (Kumar & Puja, 2012). Past literature debating the linkages between macroeconomic factors and stock prices are numerous, and they primarily focus on the influence between macroeconomic indicators and the price of securities (Kumar & Puja, 2012).

2.2.3 The Efficient Market Hypothesis and growth-tech stocks

A vast preponderance of the literatures debates the matter of stock evaluation of fast-growing technology firms in relation to financial economic fundamentals. In a context where investors avidly pursue optimistic profitable opportunities i.e. technology industries, there is no doubt that the majority of the market players would rather seek for growth stocks, instead of much less troublesome stocks i.e. value stocks (Durand, 1957). That is, forward-looking investors would always prefer securities that promise lasting growth potential over stocks that do not (Durand, 1957).

Typically, when an aggregated buying strategy aims towards a specific set of securities, the price of these specific stocks would ultimately increase (Holt, 1962). Consequently, when growth portfolios massively outweigh value investments, investors are more likely to simply base their portfolio decisions on whether a stock entails or lacks growth (Durand, 1957). As a result, the market becomes highly speculative as traders neglect to question whether the market is rewarding companies too much for expectations of future growth (Durand, 1957). In this perspective, Durand (1957) sustains that speculative growth strategies are powerfully dominant in sophisticated markets i.e. technological industries.

However, appraising growth stocks involves two main implications which are strictly linked to the underlying causes of market inefficiency (Durand, 1957). First, it is not possible to accurately assess future sales levels, upcoming earnings ratios and forecasting dividends. Second, in such a context it is extremely difficult to transform future projections into reliable present value assessments (Durand, 1957). This extreme ambiguity related to future expectations also impede a fair and sound valuation of growth stocks at the moment in which investors formulate their expectations (Székely & Richards, 2012).

The St. Petersburg Paradox by Bernoulli (1738) perfectly explicate the difficulty of assessing future growth expectations at the present time (Durand, 1957). This theory has been studied and revised by several academics, including Durand (1957), Székely and Richards (2012) and Garcia (2013). The St. Petersburg Paradox finds its *ad hoc* application in the inefficiency attributed to high speculative stock markets i.e. growth technology markets. In this regard, Székely and Richards employ the paradox to the specific case of the Crash of High-Tech Stocks in 2000 (Székely & Richards, 2012). In order to assess the high uncertainty linked to the *glamorous* expected growth of tech-securities, Székely and Richards (2012) argue that the actual value of tech-stocks in 2000 were far lower than what investors perceived at that time. Along with this proposition, Garcia (2013) proposes an empirical solution which provides the resolution of the St.

Petersburg paradox based on *expected utility* of investors. Conferring with Garcia's findings, the resolution of overvaluation of growth securities would be empirically reduced to the concept of *expected utility* that each investor in the market would subjectively attribute to the growth stocks, according to its beliefs and forecasting (Garcia, 2013). From Garcia's perspective, expected utility refers to "how much practical use investors think to gain" (pp.5). This is to say that investors' perception of utility towards growth securities would eventually specify a fair price for growth securities in the market, which in turn would ultimately become *fairly* efficient (Garcia, 2013). To some extent, the introduction of expected utility provides a solution to the St. Petersburg Paradox and provides order to the chaotic and inefficient image of the growth securities markets.

Lewis (1989) refers to Wall Street as the perfect sample of speculative security traders and portrays the investing practices as an "awful manifestation of a minimalist definition of correct" (Langevoort, 1992) (pp 852). Aligned to this view, Székely and Richards (2012) define technology security markets as a place where trading investors show foolish future expectations, which turn to be more optimistic than historical trends. In such a context, investors express a deep excitement and belief in skyrocketing returns from tech stocks. Excitement soon transforms into *irrational exuberance* which rapidly exceed the true asset value of the stocks (Székely & Richards, 2012). The steep increase in returns offered by growth stocks i.e. technology securities have sparked an active discussion about whether investors are wise to purchase growth shares.

Lamont and Thaler (2003) find in the *arbitrage* the only mechanism able to drive inefficient markets towards becoming efficient. They define arbitrage as the activity of jointly buying and selling the same stock which is wrongly traded at two distinct prices (Lamont & Thaler, 2003). Along with Lamont and Thaler's argumentation, Hillier et al. (2013) sustain that *arbitrage* is the additional key factor in granting the validity of the *law of one price*, which states that two similar securities must be priced and exchanged at identical prices (Hillier, et al., 2013). Therefore, in order to defeat divergence in prices of two similar securities, an arbitrageur should jointly short the overpriced security i.e. growth tech stock A, and instantaneously purchase the peer value stock B (Lamont & Thaler, 2003). This transaction would eventually grant efficiency to the market where both security A and B are freely traded. Through this example, Lamont and Thaler

(2003) demonstrate how simply the presence of overvalued growth stocks i.e. technology securities breach the fundamentals of the Efficient Market Hypothesis and of the law of one price (Hillier, et al., 2013).

2.3 Introduction to factors that have an impact on Internet stock prices

Given the relevance of the link between finance and macroeconomic factors, a voluminous literature tackles the interrelationships amongst stock prices, accounting variables and macroeconomic factors (Peirò, 2016). Keynes (1936) is one of the first academics arguing that in the long-run there are some factors, other than accounting factors, which exert their "compensating effects" (pp.103) over price fluctuations. On the other hand, Graham and Dodd (1988) approach stock valuation through a pure accounting perspective, debating that investors transact according to their perceptions related to the *intrinsic* value of stocks. Generally speaking, *value* is defined as the agreed price which reflects the actual worth of the stock (Gottwald, 2011). Specifically, Graham and Dodd (1988) define intrinsic value as "[…] that value which is justified by the facts e.g. the assets, earnings, dividends" (pp. 64). In other terms, if investors perceive that the stock price is below its intrinsic value, they would purchase the security. Otherwise, if investors think that the stock price is above the intrinsic value, they would sell off the security (Graham & Dodd, 1988). In this regard, a significant literature debates the importance of fundamentals in establishing stock prices. Ansotegui and Esteban (2002) determine a long-term relationship between the Spanish stock market and its fundamentals and analyze to which extent the interrelationship can be used in forecasting stock prices. Sing et al. (2002) investigate the relationship between fundamentals and stock prices in Singapore and finds significant relationships.

Along with the Keynesian perspective, Bhargava (2014) also discusses that security prices are affected by both macroeconomic and microeconomic factors i.e. data regarding companies' accounting fundamentals. He argues that accounting data is commonly compiled on a quarterly basis. However, variations in stocks prices might occur in between due to external causes such as a shift in investors' behavior and a change in interest rates (Bhargava, 2014). Finally, Bhargava (2014) concludes that differences in security valuations, based on firm-level accounting variables and based on macroeconomic shocks to the economic systems, explained by macroeconomic factors, do play a significant role in describing stock prices. Along with Bhargava (2014), Periò (2016) affirms that companies listed on the stock market are very dependent on the global economic outlook as well as on domestic economies. A huge amount of past published literature also argues that macroeconomic factors that affect the stock markets primarily relate to factors in the domestic economies (Altinbas & Biskin, 2015).

On the contrary, Wasserfallen (1989) assumes that the effect of macroeconomic events may principally depend on the specific industry and the characteristics of the firm, as some industries' and firms' economic situation is dependent on e.g. international trade. In such a case, the firm's economic situation is deeply interconnected with the amount of international trade (Wasserfallen, 1989). Given the importance of the topic, abundant literature exists providing numerous examples of both microeconomic and macroeconomic factors impact the value of a company and thereby its stock (Peirò, 2016).

2.3.1 The financial accounting factors

Chang et al. (2008) argue that if stock prices do not follow a random walk trend, they would rather follow a *mean-reversion* process. If this condition holds, Chang et al. (2008) sustain that price movements can therefore be predicted by a firm's fundamentals.

Relevant microeconomic factors are discussed in the academic literature. For instance, Preinreich (1932) argues that the strongest fundamental factors associated to growth are earnings-per-share (EPS) and market-to-book (P/B). Also, Fama and French (1992) found significant relationships between stock returns and price-related accounting variables i.e. earnings-per-share, market-to-book and leverage ratios. Similarly, Ball (1978) argues that fundamentals like earnings-per-share and book-to-market are comprehensive proxies for a wide range of factors affecting security returns. Alternatively, Kaplan and Ruback (1995) argue that the discounted-cash-flow (DCF) model provides significant estimates of the market value of highly leveraged transactions from 1983 and 1989. In their research, Kaplan and Ruback (1995) assume that leveraged transactions are more likely to have stable operating cash flows. That is, cash flows from operations are assumed to be less variable than equity flows. Their results conclusively show that the DCF model has general accepted accuracy in determining the security value of a company (Kaplan & Ruback, 1995).

2.3.2.1 Earnings-per-share (EPS)

In academic literature, there is a strong interest in factors calculated based on earnings. The focus has especially been on earnings-per-share (EPS) which indicates the profit after net tax and dividends, divided by the total amount of outstanding shares. Assuming that the company dispenses all of its income, financial analysts calculate EPS in order to see how much money shareholders would receive, related to each share in their possession (The Financial Times, 2018).

Earnings are primarily used to assess and forecast financial performance of companies (Barefield & Comiskey, 1975). Modigliani and Miller (1958) are among the first academics discussing the connection between earnings and securities value. They theorize the link between the two variables by defining that the value of a company is the result of the stream of profits over time. Consistent with Modigliani and Miller's (1958) theory, Lintner (1965) proves that the value of a company is contingent to the probability distribution of the future earnings. After analyzing the market reactions to variations occurring in stock prices, Foster (1973) concludes that both the aggregated market and the individual investors have a point in believing that earnings-per-share entail informational content which affects the value of the company and its securities (Foster, 1973). Beaver et al. (1987) also study the information content of stock prices with reference to accounting earnings. Beaver et al. (1987) report that changes in historic stock prices have a significant explanatory power when trying to explain variations in present earnings. Kormendi and Lipe (1987) explore the magnitude of the positive relation between earnings and stock returns. Freeman (1990) examines the relationship between the earnings reported in the accounting books with stock returns among small and big companies and finds that the cointegration between the variables is high. McAnally et al. (1997) debate that earnings and security prices are jointly subjected to the same underlying information, and therefore, a relationship must exist between the two variables. Chang et al. (2008) affirm that earnings-per-share (EPS) represents one of the main proxies for assessing the firm's fundamental value and stock price. Consistent with findings of Campbell and Shiller (1988), Chang et al. (2008)

examine the earnings against stock returns in order to formulate an explanation of long-term stock price fluctuations.

Healy et al. (1999) argue that the link between EPS and stock valuation relies on the accounting disclosure. In their work, Healy et al. (1999) explore the advantages of voluntary disclosure of accounting information i.e. earnings-per-share over the capital market. They argue that the results are coherent with the model of disclosure prediction, which states that increased disclosure guides investors to reconsider their expectations and tend to lead them towards more positive valuations of the firms' stocks (Healy, et al., 1999). To conclude, there has been a pivotal surge in empirical works dealing with earnings, as it appears to be a trustworthy proxy for assessing a firm's fundamentals and the value of the stock (Chang, et al., 2008).

2.3.2.2 Price-Earnings (P/E) ratio and market-to-book (P/B) ratio

Like earnings-per-share, P/E and P/B ratios represent two of the most used valuation tools in the stock market (Chishom, 2009). Watsham (1993) formally defines price-earnings (P/E) ratio as the relation between actual stock prices and EPS. Chishom (2009) specifies that price-earnings ratio serves as indicators of whether stock values in a given industry are correctly assessed[1]. The P/B ratio represents the financial indicator that links stock prices of a firm with its accounting value of shareholders' equity per share (The Financial Times, 2018).

Although both ratios are equally reliable from a theoretical point of view, P/B ratio usually receives less academic consideration. In the late 1980s, Lakonishok and Shapiro (1986) first recognized the significance of this ratio by emphasizing the strong role of book-to-market in describing average returns of securities, focusing on the particular case of the Japanese market. Fama and French (1992) analyze how market-to-book provides "a simple but powerful characterization" of average security returns in the commodity industries, for the for the period 1963-1990 (Fama & French, 1992) (pp.429). Fama and French (1992) give a lot of conceptual credit to the market-to-book ratio by attributing it with great explanatory power in relation to predicting stock returns. Also, Rosenberg et al. (1985) and Stattman (1980) contribute by analyzing how average returns of US securities are positively affected by the market-to-book

ratio. They also report that overall book-to-market equity has a steadily greater role in average stock returns over time (Stattman, 1980).
Comparably, price-earnings ratio is also an indicator of stock performance and is often referred to as the principal earnings growth indicator (Malkiel & Cragg, 1970) (Litzenberger & Rao, 1971). In practice, this ratio is very important as it is used for identifying mispriced stocks (Basu, 1977). According to Basu (1977), the bias in stock prices is a direct result of the fact that markets are far from being efficient, and consequently, the price-earnings ratio has an impact on the investors' expectations. The relationship between P/E ratio and stock performance is also documented by Smidt (1968), who argues that overconfident investors tend to overvalue security prices as a consequence of an overreaction to the information content entailed in the P/E ratio. In other terms, Smidt (1968) argues that there is a concatenated chain of events: exuberant investor optimism leads to an average higher P/E ratio and consequently, an increased price for securities. Conversely, if investor expectations are pessimistic, the mechanism is reversed (Smidt, 1968). Based on past empirical studies, it can be concluded that companies having higher P/E ratios than their peers generally report expectations of higher future earnings growth. This fact does not always turn out as a sound investment if market expectations are not met (Gottwald, 2011).

Given the definition of the P/B and P/E ratios, it is no surprise that the roles of these two financial indicators are highly interconnected. Both P/B and P/E ratios predict stock returns and performances in a way that is consistent with the accounting fundamentals (Reggiani & Penman, 2010). From a pure theoretical perspective, Penman (1996) argues that the distinction between what is explained by the P/E ratio and what is left to the P/B to explain is sometimes a bit unclear. However, the distinction results to be perfectly clear if the sole accounting perspective is taken into consideration. In fact, information contained within the P/E ratio is disclosed in the income statement; whereas, the P/B ratio is revealed in the balance sheet (Penman, 1996). Consistent with the considerations provided by Penman (1996), Wilcox (1984) makes an attempt in articulating both ratios under a single equation. By presenting his "P/B-ROE valuation model", Wilcox (1984) formulates that "each ratio is a transformation of the other" granted by the following expression $\frac{P}{B} = \frac{P}{E} \text{ x } \frac{E}{B}$, where E stands for *Equity*, P stands for *stock prices* and B stands for *book value* (Wilcox, 1984). Penman (1996) further contributes to Wilcox' (1984) research by providing directions on how to measure the ratios jointly. The starting assumption of Penman (1996) is that P/E strictly refers to future growth in earnings, while P/B solely reflects expected return on equity. By making this distinction, Penman (1996) succeeds in estimating the interconnection between the two accounting ratios and their impact on stock performance.

2.3.2.3 The Discounted Cash Flow (DFC)

The use of the discounted-cash-flow (DCF) model is well-established in finance theory and investment practice (Hillier, et al., 2013). Typically, the discounted-cash-flow method is used to solve and assess valuations related to compound interest rates measurement. However, modern revalidations of the model employ discounted-cash-flow techniques to capital budgeting and stock valuation (Koller, et al., 2010)((Damodaran, 2012).

The concept of discounted-cash-flow is based on the assumption that expected cash flow is exclusively driven by two distinct factors: forecasted returns on invested capital and revenue growth (Hillier, et al., 2013). This implies that DCF is determined as expected cash flows discounted at the cost of reinvested capital (Damodaran, 2012). DCF model is therefore expressed as follow (Koller, et al., 2010):

$$\frac{Value}{Invested\ Capital} = ROIC * \left(\frac{1-\frac{g}{RONIC}}{WACC-g}\right)$$

Where *ROIC* stands for return on the firm's current capital, *RONIC* equals the return on incremental capital and *WACC* refers to weighted average cost of capital. Therefore, the DCF model is centered around the joint understanding of infinite future cash flows and accrual accounting models (Penman & Sougiannis, 1998).

Like book-to-market and price-per-earnings ratios, DCF denotes a powerful tool in measuring security expected value. However, this method entails a relevant problem. That is, it is not always possible to estimate future cash flows (Hillier, et al., 2013). For this reason, the DCF model is not always indicated for valuating high dynamic industries i.e. technology Internet markets (McKinsey&Company, 2015). Nevertheless, there are cases where it could be meaningful to use the DCF model on technology companies; for instance, when price-to-earnings or other earnings multiples indicate negative earnings (Koller, et al., 2010). When analysts decide to apply the DCF to technology companies, they have to take into account the industry-specific characteristics. Therefore, a greater emphasis should be placed on the examination of expected long-term development of the market, rather than on the company's past performance (McKinsey&Company, 2015). In this specific context, a revision of the traditional DCF model e.g. the Reversed DCF might be more appropriate. The Reversed DCF differs from the traditional model as it first estimates the future growth and then determines the proper discount rate and the final growth rate (Nasdaq, 2013).

2.3.3 The macro-economic factors

A voluminous literature demonstrates that financial fundamentals are often challenged by macroeconomic conditions. One of the first academics who formalized the relationship between securities and macroeconomic growth was Irving Fisher in 1907. Fisher (1907) discusses two main implications: first, the current real interest rates include relevant information about expected economic growth. Secondly, expectations regarding future growth do affect security valuation in t_0 (Fisher, 1907). After Fisher, several economists dedicated their studies to the subject that associated security valuation with macroeconomic factors. Keynes (1936) also discusses the presence of non-financial factors that affect the investors' valuation. Similarly, Campbell (1989) states that the general economic outlook is the key determinant of the discounted value of expected cash flows. Since the expected cash flows determine the security value, the stock prices are therefore indirectly determined by the general economy. Basically, he argues that the link between the security market and the real economy is expressed by the fundamental valuation of equity, which in turn is highly correlated with the general economic outlook (Campbell, 1989).

According to the Efficient Market Hypothesis, security prices should embed all the information available in the market, and the security prices should react to unexpected macroeconomic events immediately (Pearce, 1984). However, the evidence observed in the market does not always support the EMH view. In reality, a distinction must be made between what is *anticipated* and what is *unexpected*[4] in the market (Pearce, 1984). In this regard, Schwert (1981) presents a result

---

[4] Pearce (1983) highlights the substantial conceptual difference between *anticipated* and *unexpected* announcements or events. According to the EMH theory, both anticipated and unexpected events should be immediately incorporated

that contradicts the EMH as he shows that stock prices slowly adjust to unexpected changes of inflation values. On the contrary, Pearce and Roley (1983) prove the validity of EMH in the case of weekly money stock announcement. They analyze the reaction of security prices on weekly basis after the routinely announcement. They prove that when the macroeconomic news is expected, the investors are ready and rationally react to the new information. In addition, they examine that only the unexpected change of money stock has a relevant effect on the stock valuation (Pearce & Roley, 1983).

However, Roll (1984) and Cutler et al. (1988) argue that expected and unexpected macroeconomic variables and events can usually explain approximately *one third of the returns on stocks.* By taking the stock market as a whole, Cutler et al. (1988) analyze both the monthly stock returns from 1926 to 1985, and the annual returns from 1871 to 1976. Their results show that macroeconomic events only explain a fraction of the market reaction and stock prices which is approximately equal to one third of the stock return (Cutler, et al., 1988).

Typically, academic literature link macroeconomic factors with stock returns and study whether a set of parameters affect a specific stock during a given period of time. For instance, Wasserfallen (1989) examines the unexpected changes of specific macroeconomic indicators, i.e. exchange rates and the index of import prices, and their impact on some European stock markets e.g. Switzerland, West Germany and Great Britain. The results of his work suggest that the macroeconomic events have a very small short-term impact on stock returns for these domestic markets. Consistent with Wasserfallen, also Pethe and Karnik (2000) study the impact of some macroeconomic factors, i.e. exchange rates between rupee and US dollar, index of industrial production and prime lending rate and money supply, on the Indian stock market. However, as technology is a key aspect of the Indian market, they conduct their analysis emphasizing also the importance of this non-financial factor. The results of their research suggest that the cointegration between macroeconomic factors and the Indian stock market is not very conclusive for the short-run (Pethe & Karnik, 2000). Along with Pethe and Karnik's findings, also Maio and Philip's (2015) conclude similar considerations after taking 124 macro-variables and applying the VAR

---

into the stock prices. In practice, the likelihood, that an unexpected event is immediately and correctly absorbed by the investors in the market, is extremely low. Therefore, in the case of unexpected events, the EMH theory is generally violated.

method on six common factors drawn by statistical estimations from the sample of the variables. They argue that macroeconomic variables do not have a relevant impact on stock returns in the short run (Maio & Philip, 2015).

However, there might be some instances where macroeconomic indicators may have an impact in the short-medium term. This is the case argued by Peirò (2016), who examines whether interest rates and production indexes affect the stock markets in France, Germany and the UK. Firstly, Peirò (2016) finds that a change in production would start a mechanism of adjustments which would ultimately impact stock returns in the short term. In fact, the paper argues that a change in production would cause proportional and equal changes to stock prices through the assessment of new expected future dividends. Secondly, an increase in interest rates would imply higher discount rate and therefore lower stock prices. Furthermore, it is pointed out that there might also exist a link between interest rates and production. If this is proved true, the increases in interest rates may cause decreases in investments and thus in future production. Peirò (2016) concludes that interest rates may eventually affect stock prices in two ways: directly through changes in discount rate or indirectly through changes in future production.

2.3.3.1 Oil prices and the stock market

Oil prices represent a key macroeconomic indicator because increases in oil prices are often indicative of the general macroeconomic outlook. In fact, oil prices could be determinant in explaining the inflationary pressures in the economy, which in turn could indicate the future of interest rates (Sadorsky, 1999). In this regard, Adelman (1993) affirms that “oil is so significant in the international economy that forecasts of economic growth are routinely qualified with the caveat: ‘provided there is no oil shock’” (pp. 537). Hamilton (1983) argues that the oil prices are involved and partially responsible for all the post-World War II recession in the United States. Consistent with Hamilton’s findings, also Burbidge and Harrison (1984) and Loungani (1986) test and prove that Oil price fluctuations are highly explanatory when associated with the post-World War II U.S. recessions. By contrast, Jones and Kaul (1996) specifically investigate the relationship between oil process and the financial markets. They study quarterly data of the post-World War II period in the United States, Canada, Japan and the United Kingdom, and test whether the reaction of international stock markets to shocks in oil prices can be explained by the present and incoming changes in real cash flows and changes in expected returns (Jones & Kaul, 1996). They conclude that the oil price shocks are comprehensively accountable for the shock that occurred in the real cash flows in the U.S. and Canadian stock markets. However, the results in Japan and the United Kingdom are not as strong as in the previous cases (Jones & Kaul, 1996). Also, Huang et al. (1996) analyze the relationship between daily U.S. shock returns and daily oil future returns. Their analysis suggests that the returns of oil futures have a medium-long term impact on some individual oil company stock returns. However, oil future returns do not have any significant impact on broad-based market indices i.e. S&P 500.

Overall, it can be argued that oil prices are among the major Leading Economy Indicators (LEI) because of their explanatory efficacy in describing the general economic outlook (Yamarone, 2016). Past academic literature agrees with the statement that oil prices play a significant role in affecting the economic activity (Sadorsky, 1999). However, the oil price impacts on security prices are observable in the medium-long run with little significance in the short run (Yamarone, 2016).

2.3.3.2 Money Supply (M) and the stock market

Investors often consider money supply as an indicative factor in influencing changes in stock prices (Yamarone, 2016). The underlying reason is primarily provided by the Friedman and Schwartz (1963), who argue that the turning point of business cycles is associated with the earlier changes in the growth rate in the security market. Consistent with these findings, Sprinkel (1964) observes that changes in growth rates are anticipated by changes in stock prices, which in turn are caused by a shift in consumer goods and services. In addition, Sprinkel (1964) argues that investors in the market can actually predict the changes in stock prices by observing past knowledge related to money supply fluctuations in the market. In other terms, Sprinkel (1964) proves two main causal facts; first, money supply in a long-term leading indicator of stock prices. Secondly, smart investors are given the chance to beat the EMH theory and gaining above-average returns in stock investment if they are attentive enough to use the money supply indicator. Along with Sprinkel, also Palmer (1970), Reilly and Lewis (1971) support the theory that monetary changes predict fluctuations in security prices. In this regard, Rozeff (1975) affirms that the relationship between stock prices and money supply is causal and observable in the long-run. However, Rozeff (1975) supports that money supply can be considered an indicator of stock prices only if one condition holds, that is, investors should be attentive towards the market fluctuations and should have an idea of the present phase of the business cycle. Indeed, he argues that without a consolidated knowledge of the theory of business cycle and without an accurate forecast of the different phases, investors are not able to exploit the potential benefits provided by the money supply indicator.

2.3.3.3 Gross Domestic Product (GDP) and the stock market

Economists and politicians consider Gross Domestic Product (GDP) above all other economic indicators because it is the broadest and most comprehensive measurement of the current economic outlook, besides providing additional insights for future financial growth (Yamarone, 2016). For decades, practitioners have recognized the strict relationship between security markets and GDP. The relationship derives from the association between economic activities tracked and future economic growth which generates expectations across the market and therefore impacts the current values of stocks (Foretti, 2007). In other terms, if the economy is expected to move upward and enter a growth phase, the security market will foresee this shift and consequently it will adjust current prices by bidding up the prices of those stocks which are estimated to increase (Risager, 2016). The real GDP growth is provided by the annualized quarterly growth rate contained in the real GDP report. When it reports positive results and numbers, it is a good symptom for the general economy, corporate profits and finally, security valuation in the stock markets (Yamarone, 2016). The market reactions, either negative or positive, are far more accentuated when the announced numbers differ from the forecasted ones (Hillier, et al., 2013). While the real GDP growth can depict negative values, it is rare that the nominal GDP growth value records negative results. This is because nominal GDP growth entails inflation variables which is excluded in the calculation of real GDP. The only condition that leads the nominal GDP to report value below 0 is when deflation occurs (Yamarone, 2016).

2.3.3.4 Inflation, interest rates and the stock market

By measuring the difference between nominal and real GDP, it is possible to calculate the inflation present in the real economy. In fact, Fisher (1907) defines inflation as the difference between nominal and real GDP. This calculation is called the *implicit price deflator* (Yamarone,

2016) (pp. 22). There exist several distinct deflators, i.e. implicit GDP deflator, personal consumption expenditure deflator etc. Overall, all deflators provide information regarding the broad price activity in the real economy. In light of Fisher's findings in 1907, a voluminous literature developed discussing the causality between inflation and interest rates and its association to stock returns. Fama (1981), Ram and Spencer (1983), Geske and Roll (1983), James, Koreisha and Partch (1985), Stulz (1986) and Kaul (1987) all argue the negative link between real interest rates and inflation. Indeed, they discuss that both bonds and stocks should be somehow hedged against inflation. Specifically, Fama (1981) argues that the negative correlation between stocks and inflation should be exclusively attributed to the negative relation between inflation and real economy. In other words, he proves that there exists a positive correlation between stock returns and real economic activity, and that the negative link between stock returns and inflation is primarily induced by the negative relation between real activity and inflation (Fama, 1981).

In order to contrast the rise of the inflationary levels, the interest rates increase (Mankiw, 2005). It follows that investors are likely to misprice growth stocks by failing to recognize when growth stocks become value stocks (Mankiw, 2005). Therefore, a negative correlation between growth stocks and inflation exists.

Rather than inflation, stock returns are sensitive towards interest rates (Pearce & Roley, 1983). Interest rates represent the tool used by central banks to intervene into the economy (Mankiw, 2005). Interest rates have a direct and speedy impact on the inflationary levels, oil prices, the exchange rates and on the money supply in a domestic economy (Risager, 2016). Therefore, general macroeconomic variables do not really have a direct and short-term impact on the stock valuation. But it can be argued that macroeconomic variables indirectly affect security returns through changes in interest rates, and that the impact is medium-short term (Pearce & Roley, 1983) (Cutler, et al., 1988) (Serven & Solimano, 1989). For instance, two are the main effects of the increased short-term interest rates on stock markets. First, the increase in interest rates leads to a proportional increase in the discount rates which is associated with more expensive costs of borrowing for investors and companies. Second, the higher borrowing costs reduce the positive net present value (NPV) of projects as lower cash flows are expected. The result of these two mechanisms is that stock prices would decrease in value (Degiannakis, et al., 2017).

2.4 Index Creation

Stock market indexes are created for various reasons and can be used for different purposes. Fundamentally an index is a "Statistical composite that measures changes in the economy or in financial markets, often expressed in percentage changes… Indexes measure the ups and downs of stock [markets],…" as defined by (*Index Definition - NASDAQ.com*, 2018). Depending on which stocks that are included in an index, it can mirror everything from the performance of a broad market to highly specific market segments or investment strategies (Speth, 2003). Stock indexes are also used as underlying assets for derivatives, as well as a benchmark for stock portfolios. Furthermore, indexes are used in academic research (Aboura and Chevallier, 2017). The main uses of stock market indexes are; to mirror a part of the stock market or an investment strategy as underlying assets for derivatives and a tool to be used when carrying out academic research.

In the financial industry common methodologies exists for creating indexes. The three most common ways of creating an index are capitalization weighted, equally weighted and price weighted (Blackman, 2003; Hulbert, 2003; Speth, 2003; Aboura and Chevallier, 2017). The common factors of these three methods are that they are easily interpretable and that they are based on a single measure. In a capitalization weighted index, such as Standard and Poor's 500, a stock enters the index with a weight based on its value. This value is calculated as the total dollar value of all publicly traded stocks. In an equally weighted index all stocks in the index enter with the same weight, whereas a price weighted index, such as Dow Jones Industrial Average, uses the price of one stock to calculate the index weight (Blackman, 2003). While the aforementioned methods have a long history in the financial industry, academia has been seeking out new ways of creating indexes.

In the academic literature, alternatives to the traditional index methods exist. (Speth, 2003) develops a share-weighted index where the individual stock price is multiplied by an adjustment factor which is calculated based on a number of factors that are deemed relevant by the creator of the index. More complex methods also exists where statistical methods of Principal Component Analysis and dynamic equicorrelation (DECO) models are applied (Aboura and Chevallier, 2017). The academic literature has developed more complex methods for creating indexes, some which depend on advanced statistics while others are easier understood and applied.

When evaluating the feasibility of the methods for creating indexes there are several things to consider. The simplicity of the capitalization weighted model, the equally weighted model and the price weighted model are the main advantage of these models (Blackman, 2003; Speth, 2003; Aboura and Chevallier, 2017) but each of them have serious drawbacks. While the capitalization weighted index is the index model that is most frequently used in the financial industry, it is generally accepted that it has one major flaw, namely that price changes in larger companies predominate over the price changes in smaller firms (Blackman, 2003; Hulbert, 2003; Speth, 2003). This means that the development of an index can be driven mainly by the stocks of a few giants. While this might be macro consistent, it is not necessarily representative for how the individual stocks perform in general. Both the equally weighted method and the price weighted method lack scientific evidence (Aboura and Chevallier, 2017). While the price weighted index method alleviates the concerns that stocks of giants dominate an index, it creates another problem. The method not only ignores that a giant company, and thereby its stocks, might influence the market to a greater extent than a small stock, it even implies that a stock of a big company might have less impact than that of a small one. This would be the case if the smaller company has a higher price per share than the big company (Speth, 2003; Aboura and Chevallier, 2017). In the academic literature, methods based on scientific evidence are developed. While these models may better fit the available data, they are in many cases hard to implement in the financial sector due to lack of simplicity, transparency and the easiness of interpreting index changes. The two main areas to consider when selecting a method for creating an index are the simplicity in understanding, applying and interpreting the index as well as an index’ ability to mirror its subject, be it a broad market, a narrow market segment or an investment strategy.

2.5 Statistical analysis of stock performance

The behavior of stock-market prices is a pivotal topic in financial economics. In this regard, a voluminous economic literature studies the relationships between the stock performance and specific financial variables i.e. macroeconomic and accounting factors (Mukherjee & Naka, 1995). In this connection, statistics and stochastic processes play a key role in supporting the empirical researches. Of particular importance are the concepts of *ordinary least squares multiple linear regression* and *time series* (Allison, 1999).

Traditionally, security prices are analyzed through the theory of *time series*. Brockwell and Davis (2016) broadly describe time series as a set of observations, x_t, at a specific time t. The formal definition defines times series as "a set of regular time-ordered observations of a quantitative characteristic of an individual or collective phenomenon taken at successive points of time" (OECD, 2005). Usually, time is broken down into equal intervals and is considered as a discrete variable, meaning that $t = 1, \ldots, T$ are discrete points (Kjersti & Dimakos, 2004). By contrast, time series could either be *continuous time series* or *discrete time series* analyses. Continuous time series take into consideration observations that are recorded on an ongoing basis, over a specific period of time. On the other hand, discrete time series are series that study observations that are reported at specific and fixed time intervals (Brockwell & Davis, 2016).

Allison (1999) refers to *least squares multiple linear regression* as a statistical method that examines the link between a specific dependent variable Y and one or more independent variables Xs. The multiple linear regression technique can be used for two distinct purposes: forecast or causal analysis. The purpose of prediction analysis is to build a mathematical formula that explains the future trends of the dependent variable through the observation of independent variables (Allison, 1999). The strength of the multiple linear regression model arises from its logical simplicity which consists of formulating an equation that expresses the direct relationship between the variable of interest and a set of potential explanatory variables (Montgomery, et al., 2012). In multiple regressions the relationship between the dependent and the independent variables is assumed to be linear, meaning that the quantities are assumed to be proportional to each other (King, 1997).

In a predictive analysis, the multiple linear regression combines a set of explanatory variables with the aim to forecast the behavior of the dependent variable (Allison, 1999). Typically, empirical researches are classified into two major schools of thoughts. On one side, a large group of economists affirm that *past stock performances* represent an important pool of information that partially explains current stock prices and affects future stock valuation. By contrast, other academics support the theory of *random walk* which basically assumes that future stock performances do not depend on past trends, but that they are randomly distributed (Fama, 1965). In other terms, some economists support the first theory based on the concept that history repeats itself as past stock behaviors are likely to repeat in the future. Whereas, others support the theory of random walk which argues that the past cannot be used to explain and forecast the future in any relevant perspective. The main difference between the two forecasting theories is whether successive prices are considered as dependent or independent from historical trends (Fama, 1965). In this regard, Fama (1965) and Sydney (1967) argue that in practice the *perfect* independence does not exist. Hence, the random walk theory does not accurately explain what happens in the real markets (Sydney, 1967). In real markets, the hypothesis formulated by the random walk theory must be adapted by stating that the independence assumption holds as long as "the series of successive price changes is not above some minimum acceptable level" (Fama, 1965) (pp. 35).

In causal analysis, the multiple linear regression looks at the independent variable as causes of the dependent variable (Allison, 1999). In other terms, the purpose of the analysis is to study whether the set of independent variables have an effect on the dependent variable and to assess the scale of that specific effect (Cohen, et al., 2003). From a theoretical standpoint, the *causal inference analysis* represents a special application of the prediction analysis in which the aim of the study is to forecast what would happen under different explanatory variables (Gelman & Hill, 2006). Although multiple regressions can be used for both predictions and causality, the *least squares multiple linear regression* is primarily used for forecasting time series, outside of academia (Allisaon, 2014).

The choice regarding which method to use should be based on the results deriving from the statistical tests of cointegration i.e. the Engle-Granger and Johansen tests (Hobdari, 2016).

Granger (1981) introduces the idea of cointegration vectors in the study of non-stationary time series. The term *cointegration* defines whether there exists a linear and stationary combination between a set of non-stationary time series (Granger, 1981). In collaboration with Engle, Granger (1987) suggests that the method of *regression* could be used to estimate the cointegration relations between time series. Alternatively, cointegration can be tested by applying the Johansen Cointegration test (Hobdari, 2016). From a statistical standpoint, the Johansen test is a multivariate generalization of the augmented Dickey-Fuller test. Unlike the Dickey-Fuller test, the Johansen test measures whether linear combinations of variables for unit roots exist (Hjalmarsson & Österholm, 2007).

If the tests of cointegration suggest that correlation exists between the time series used in the analysis, then the *ordinary least squares multiple linear regression* (OLS) model is not the best method for the analysis as the results would suffer from a high degree of bias, which makes the results inconsistent (Hobdari, 2016). However, if both test results report that no cointegration exists, the OLS model can be used, and the non-stationarity condition can be solved by differencing the time series (Hobdari, 2016).

In this regard, when cointegration holds, the past literature employs two major statistical models: the Vector Auto Regressive (VAR) or the Vector Error Correlation (VECM) approaches. Both methods are specifically used to investigate the causal relationships among financial explanatory variables and stock performance. For instance, Fama and French (1988) and Lamoureux and Zhou (1996) apply a VAR model in order to examine to what extent stock prices are ruled by random walks and stationary processes. Similarly, Brandt and Kang (2004) study the relationship between stock return volatility and mean returns trough the specific application of the latent VAR model. Nasseh and Strauss (2000) analyze the long-term relationship between stock prices and economic activity in six European countries. First, they assess to what extent macroeconomic activity influence stock prices in the long run; then, apply both VAR and VECM models and compare the outcomes. By contrast, Mukherjee and Naka (1995) apply the VECM model to the case of the Japanese stock market and show that Japanese securities are long-term cointegrated with some macroeconomic variables e.g. inflation, money supply, exchange rates, industrial production index and government bonds. Likewise, Al-Majali and Al-Assaf (2014) prove first the cointegration of variables and then use VECM model to assess the long-term relationship between Amman Stock Exchange (ASE) and a set of macroeconomic factors i.e. Gross Domestic

Product (GDP), Consumer Price Index (CPI), credit to private sector and the weighted average interest rate on time period.

3. Methodology

3.1. Research Design

As defined by Saunders et al. (2016), *research design* defines the comprehensive methodology that the authors of a paper have followed, with the intent to solve a problem statement and answer the research questions of the paper. In this regard, this chapter will introduce the reader to the authors' decisions regarding *research philosophy* (see Chapter 3.1.1. Research Philosophy), *approach* (see Chapter 3.1.2. Research Approach), *strategy* (see Chapter 3.1.3. Research Strategy) and *method* (see Chapter 3.1.4. Research Method). Afterwards, more details regarding the methodology, the data collection techniques and the analysis will be presented in "3.3 Data collection" paragraph. An adapted version of the well-known "Research Onion" modeled by Saunders et al. (2016) is applied (see Figure 2 - The Research Onion (Saunders, et *al.*, 2016)).

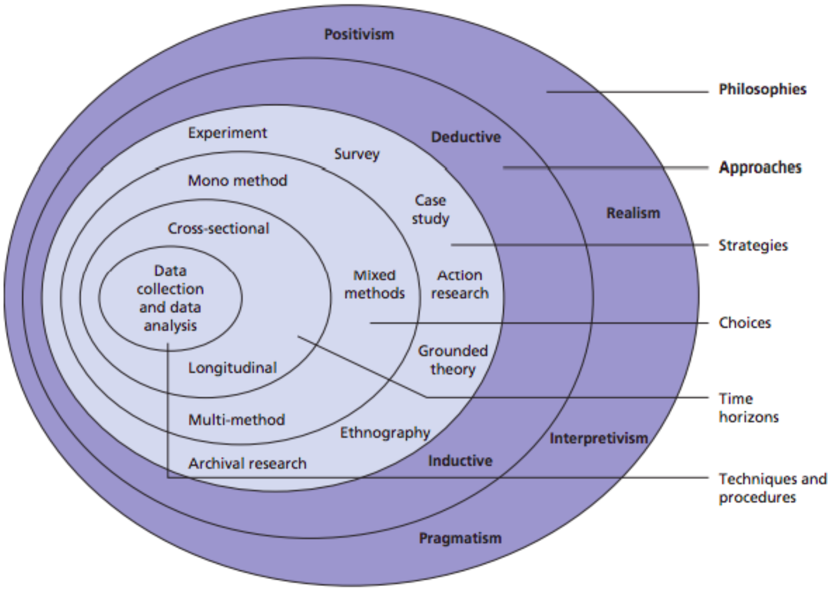


Figure 2 - The Research Onion (Saunders, et al., 2016)

3.1.1. Research Philosophy

According to Saunders et al. (2016), *research philosophy* defines the relationship between the knowledge and the research processes, by developing a comprehensive understanding of a specific topic of interest. In other terms, the research philosophy contains several key assumptions about how the authors of a paper see their study. Consistent with the findings of Johnson and Clark (2006), the primary purpose of this section is to reinforce the philosophical commitment of the research, before specifying the research strategy and approach.

As suggested by the Research Onion framework, the authors of this paper have decided to approach the selection of the research philosophy i.e. *positivism*, *realism*, *interpretivism* and *pragmatism*[5], by considering three philosophical gateways i.e. *ontology*, *epistemology* and *axiology*[6] (Saunders, et al., 2016). Overall, this research can be classified as both *realistic* and *pragmatic* (see *Table 1*)

Table 1 - Research Philosophies matrix. Source: Personal production

Philosophical Gateways	**Research Philosophies**			
	Positivism	**Realism**	**Interpretivism**	**Pragmatism**
Ontology		**X**		**X**
Epistemology		**X**		**X**
Axiology				**X**

From a pure ontological point of view, the authors have chosen a scientific enquiry, which relates to the broad topic of financial stock valuation. Therefore, both the analysis and the findings are

---

[5] The first research philosophy type is the positivism, which approaches the topic with the philosophical stance of the natural science. The realistic philosophy refers to an objective and trustful representation of the data and results, with no subjectivism. The Interpretivism emphasizes the difference between the topic of interest and the context. Therefore, it suggests that the critical interpretation is key factor. Last, the pragmatism argues that the most important part of the research is the formulation of the research questions. The pragmatism allows the application of different methods for the analysis, according to the best practices.

[6] The ontology perspective describes the researchers' view regarding the nature of the topic of interest. Whereas, the epistemology view refers to the authors' approach regarding the existing empirical literature to be taken into consideration. Finally, the axiology refers to the authors' view on how to value the results.

presented with statistical evidence. From an epistemological standpoint, the authors have based their discussion on both observable and objective existing knowledge, related to econometrics and financial economics. Finally, the authors have decided to achieve a conclusive holistic picture which combines the strengths of both subjective value-based discussions and objective data-driven argumentation. Specifically, this paper has drawn conclusions based on statistical evidence, which have then been enriched by interpreting the results in the light of general economic and business considerations.

3.1.2. Research Approach

Saunders et. al (2016) classify the *research approach* into two distinct categories: the *deductive approach* and the *inductive approach*. The main difference between the two approaches consists on the methodical procedure; in the deductive approach, the researchers develop a theory and hypothesis, and consequently execute the analysis in order to prove or reject the hypotheses (Bryman, et al., 2011). By contrast, the inductive approach suggests to first collect the data and then build a model on the basis of the results derived by the analysis (Bryman, et al., 2011).

In this regard, the authors of this paper have decided to adopt a *deductive empirical*[7] *approach*, as the nature of the investigation is primarily quantitative and statistics-based. There are three main reasons for why this paper follows the deductive approach. First of all, the deductive approach is a natural choice when seeking to describes the causal relationships between variables (Saunders, et al., 2016). Second of all, this approach is the best explanatory method when quantitative data is involved (Bryman, et al., 2011), because it offers a very structured and logical methodology. Finally, the structure of the deductive approach allows the readers to easily replicate the methodical steps of the analysis; therefore, replicability is fully ensured (Saunders, et al., 2016).

As the expression *deductive empirical approach* suggests, the nature of this work is investigative, and the results will be outlined based on a methodological observation of quantitative data.

To conclude, the research approach adopted in this paper is the *deductive empirical* methodology. It follows that, the data collection was carried out to *gather* an exhaustive pool of data for the

---

[7] The term *empirical* refers to a research that presents results based on an exploratory approach (Bryman, et al., 2011).This concept will be further discussed in the next section, "*Research strategy*".

subsequent analysis. Likewise, the application of the econometric model was intended to *explore* and *assess* the explanatory power of some economic factors over the FANGs. Finally, the results have been collected and will be *discussed* in the light of both empirical business knowledge and the theory presented in the literature review (Bryman, et al., 2011).

3.1.3. Research Strategy

After the choice of the approach, Saunders et al. (2016) suggest deciding the *research strategy*. Yin (2013) outlines three different types of strategies: *exploratory*, *descriptive* and *explanatory*. *Explanatory* studies analyze the causal relationships between the variables of interests (Saunders, et al., 2016). Whereas, *Exploratory* studies refer to the analysis of phenomena in a new light (Robson & McCartan, 2015). By contrast, *Descriptive* studies report the phenomena with descriptive accuracy.

In this regard, this work is explanatory to the extent that it attempts to assess an *influential relationship* between a set of independent variables e.g. earnings-per-share and book-to-market values, and the variable of interest i.e. the FANG index. Additionally, it is also exploratory to the extent that it proposes a new angle of analysis, when assessing the value of FANG stocks.

Once the research strategy is established, it guides the researchers on which research method to apply. (Saunders, et al., 2016).

3.1.4. Research Method

The next step is defining the *research method*, which is an important part as it dictates the methods of analysis that can be considered.

Saunders et al. (2016) suggest that in order to decide the research method, the authors should decide whether the investigation should be *qualitative* or *quantitative*. As mentioned in the previous chapters, the authors have decided to investigate an economic topic, which is based on the financial and econometric theoretical foundations. Therefore, a quantitative analysis of data is performed.

However, qualitative considerations will be discussed in the chapter related to the discussion of the results. The purpose of qualitative reflections in the discussion is to underpin the quantitative results, derived from the mathematical application of the econometric model. Therefore, qualitative considerations will be primarily related to general economic reasoning and business knowledge.

When choosing the research methods, the authors have decided to adopt the so-called *mono method* (Saunders, et al., 2016). Unlike the *multiple method*, the mono method implies the assortment of data through a single quantitative data collection technique (Bryman, et al., 2011). In fact, the authors have extracted the majority of the required data from a single data provider. The next chapter will discuss the collection of data in detail.

3.2 Theoretical framework

As a fundamental input in the econometric model, we will require a single metric which represents the development of the stocks we are examining. The logical way to create one metric across a number of stocks is to create an index.

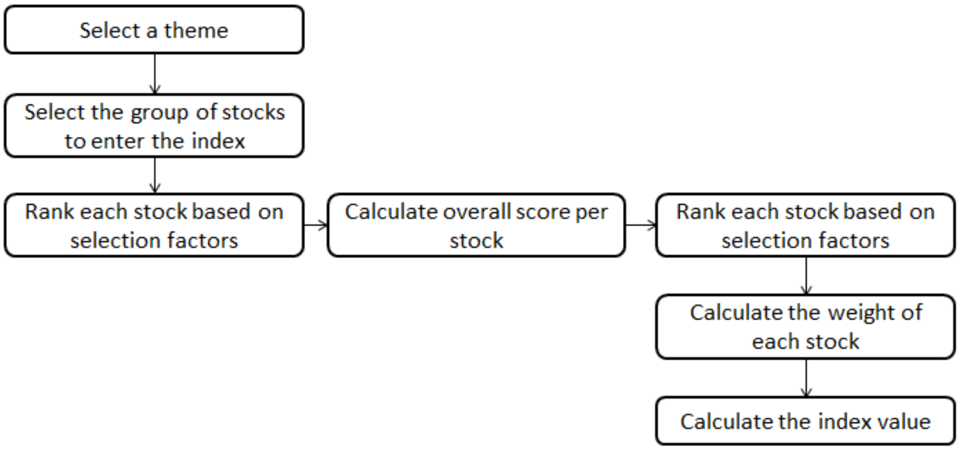
As we saw in section “2.4 Index Creation”, a stock index can be created in a number of different ways. In the financial industry, the capitalization weighted, price weighted and equally weighted indexes are the most common (Blackman, 2003; Hulbert, 2003; Speth, 2003; Aboura and Chevallier, 2017). Often, little emphasis is placed on how an index is developed, and the methods of e.g. S&P500 and DJIA are taken as given despite their methods lack scientific evidence and have a narrow focus on a single factor (market cap and stock price respectively).

Academic literature that uses indexes in their research often do so in an effort to understand a broad section of the market (Kilian and Park, 2009) or to analyze matters concerning futures and derivatives which require that an index is acting as an underlying asset (Stoll and Whaley, 1990; Jagannathan and Wang, 1996; Chan, Chan and Karolyi, 2018). In such research, the common index method is the capitalization weighted index.

As the index value will act as the dependent variable in the econometric model, it is paramount that the index represents the stocks that we are examining best possible. In the evaluation of the different methods for creating indexes, the following has been considered: transparency, how easy the results are to interpret, the ability to capture more than just one aspect of a stock's characteristic and the ability to mirror a narrow market segment such as the FANG stocks.

For the purpose of this paper, the share-weighted index method developed by (Speth, 2003) was found to be the most adequate index model. By incorporating several factors when calculating the index weights, the flaws of the most common, but less scientific index methods, such as the capitalization weighted, price weighted and equally weighted indexes, are avoided. At the same time, the method is easily understood and interpreted, as it does not require complicated statistical calculation to arrive at the index weights, which is the case with e.g. (Aboura and Chevallier, 2017). By striking a good balance between being too complex and being too simplified, the share-weighted index provides a useful method for creating the indexes required in the research that follows in this paper.

The application of the share-weighted index can easily be understood through a series of steps. First a theme must be selected, then a group of stocks to examine are selected, then they are ranked based on appropriate factors and lastly the weight of each stock and the index value are calculated. The approach can be seen in *Figure* 3 - Method for *developing model. Source: Personal production*.



In the following section, the application of the first two steps of the model, displayed in *Figure 3*. First, the index theme must be established. As described in "3.2 Theoretical framework", this paper will carry out a comparative study of three different stock market areas. To describe the development in each market area an index will be created for each area. The theme of each index will be dictated by the market area. Index A will consist of the FANG stocks, Index B will consist of the Chinese equivalents of the FANG stocks and Index C will consist of the most influential American internet stocks from the pre-dotcom era, which can be considered the FANG stocks of that era. Thus, three specific index themes exist as a natural result of the areas that this paper examines.

Second, the stocks to enter each index must be found. To determine which stocks that should enter the index, it must first be determined how many stocks that should be in the index and thereafter which factor that should be used for deciding which stocks that should enter.

The same number of stocks will be chosen to enter each index. The reason for selecting the same number of stocks for each index is that a share-weighted index containing considerably more stocks than another would have a greater impact from “smaller” companies. That is, if the stocks to enter an index are selected based on the market cap of a company; an index consisting of four stocks could represent company giants only, while an index consisting of twenty companies would also consist of large-cap companies. As the characteristics of small and large firms tend to differ (Prakash et al, 2000)(Dean et al, 1998), and it is important to compare apples to apples, all the indexes used in this paper will consist of four stocks. The number of stocks in the index is dictated by Index A, which per definition consists of four stocks (FANG = Facebook, Amazon, Netflix, Google).

As the number of stocks to enter each index has been determined, it is time to elaborate on the method for selecting the four stocks to enter each index. In line with the share-weighted index model (Speth, 2003), a single deciding factor is used for selecting the stocks to enter the index. In the literature review, it was documented that market capitalization has a significant impact on internet share prices, and it is the factor most commonly used for calculating indexes. Thus, the market capitalization is used to find the four stocks to enter index B and C respectively (Index A is already given: FANG).

Now that we have a method to determine which stocks that will enter the index, it is necessary to determine how the weight of each stock is calculated. From *Figure 3,* it can be seen that prior to calculating the weights, it is necessary to choose a set of factors. The stocks are then ranked within each factor, such that a score can be assigned to each stock.

As previously discussed the overall goal of applying the share-weighted index is to use a simple index model while still creating a more nuanced picture than the capitalization weighted model and other nonscientific models offers. Thus, in the selection of factors, focus has been on maintaining the transparency while adding a few factors that could contribute to a more nuanced index, where not only the value of a company is driving the index weight.

The factors that have been chosen are the market capitalization, the yearly return and the yearly volatility of stock returns. Market capitalization is included as this is a well-documented factor in the literature while at the same time being the most commonly used factor in the financial industry. From the literature review, we also learned that a high yearly return is characteristic for internet stocks in the periods that we are examining. While high volatility is a known characteristic of internet stocks, this is not the main reason that it is included as a factor. The reason for including the yearly volatility of the stock return is driven by philosophical considerations rather than by fundamental reasoning, which was the case with the selection of the two previous factors. The overall purpose of the indexes is to depict the development of the underlying “market”, consisting of the four leading internet stocks, rather than showing an isolated snapshot at a given time. For this reason, it is desirable to promote less volatile stocks as this enhances a more stable development. The ranking of the stocks based on yearly volatility of stock returns is inverted, such that the weight of stocks that are more stable is greater than those with high volatility. By ranking the stocks according to these three factors, the indexes that are developed are representative of the development in the leading internet stocks.

In order to have an idea of the homogeneity of the stocks within one index, a correlation matrix is calculated based on the stock prices.

3.2.1 Building a model

Once the indexes have been created, focus can be shifted to building a model that enables us to discuss which factors that impact stock prices, which is what the problem statement requires. In “

3.1.1. Research Philosophy” chapter, this paper was defined as *deductive* type of thesis, where a model is built, so that it can be tested whether a given factor has an impact on the stock price or not. In the previous part, it was presented how an index was created for each of the samples that we wish to examine. Namely the FANGs, the Chinese equivalents and the US pre dotcom bubble companies. As it follows from the problem statement, the model will be developed for the FANG index; and it will then be tested if the same model can be transferred to the two other indexes.

3.2.1.1 Variables in the model

The FANG index will act as the dependent variable in our model. As the dependent variable is in place, it is necessary to select the independent variables that should be included in the model. The independent variables will be formed by macroeconomic and accounting factors. The chapter “2.3 Introduction to factors that have an impact on Internet stock prices” discusses potential factors in depth. The relevance of the individual factors in the context of this paper will be evaluated based on two broad parameters. First, the factor’s relevance for the general stock market, and for Internet stocks in particular, will be taken into consideration. Secondly, the frequency of use in academic literature will be considered as a factor that has been frequently used in the past is more likely to be significant in the model developed in this paper. The remainder of this section will discuss which of the factors that were found in the literature review that should enter the model as independent variables.

In the section “2.3 Introduction to factors that have an impact on Internet stock prices”, we saw that both macroeconomic and accounting factors have been documented to have an impact on the development of share prices. Both Fisher, (1907) and Keynes, (1936) documented the importance of macroeconomic factors. For practical reasons, the number of variables included in the model will have to be limited to those that are expected to be most significant.

Some of the macro-economic factors that are important in general are not included in the model of this paper. The broadest and most used economic indicator is GDP (Yamarone, 2016). While

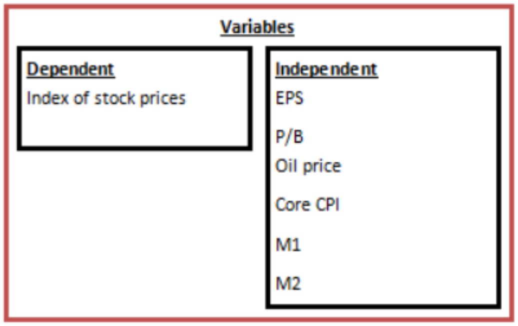
GDP is the most used indicator for the broad economy, it is not included as a variable in our model as it is so broad that it suffers from a lack of focus. In addition, the actual GDP level is not what often drives the market. Instead, it is deviations from the forecasts that usually impact the market. In the academic literature which assesses the impact of macro-economic variables on stock prices, it is common that GDP is not one of the variables included in the models. For these reasons, GDP is not included as a variable in this paper. Another significant macro-economic factor that is not included in the model is the interest rate. The interest rate has a significant impact on the broad economy and on oil prices, CPI, money supply etc., but similarly to GDP one of its shortfalls, when looking for variables to predict stock prices, is that it might be too broad a factor. In the academic literature, it is more common to include CPI and money supply as model variables. Hence the model will not include GDP and interest rate as variables.

Based on the selection criteria and the section "2.3.3 The macro-economic factors", four macroeconomic factors are chosen to be a part of the model. The first factor to enter the model is the oil price. The oil price is an important leading indicator (Yamarone, 2016), and is so significant that economic outlooks often make reservations depending on the oil price development (Adelman, 1993). Another advantage of using the oil price in a model which examines international stocks is that the oil price is not linked to a particular country or geographical region unlike many other indicators. In the academic literature, the oil price is a common factor in models for stock price development. The second factor to enter the model is the core consumer price index (core CPI). Core CPI is a measure of the core inflation. The 'core' refers to the fact that energy and food have been deducted as these components are very volatile. Furthermore, the deduction of energy is a desirable feature when included in the model as a great part of the development in energy prices is already captured by the oil price which is included in the model. Fama (1981) finds that there is a negative correlation between inflation and the stock prices. When building models to forecast stock prices, CPI is a common factor to include as we saw in the literature review "2.3.3 The macro-economic factors". The third factor to enter the model is the money supply. Different measures of money supply exist where M0 is the narrowest definition and M3 the widest. The narrowest money supply includes only the most liquid money stocks, such as cash, whereas the broader money supply includes all that of the narrow money supply plus money stock that cannot be converted into cash as quickly. There is an abundance of literature that highlights the relationship between money supply and changes in stock prices (Sprinkel, 1964) (Palmer, 1970) (Reilly & Lewis, 1971). Money supply is commonly included as an explanatory variable when examining stock price changes e.g. in (Pethe & Karnik, 2000). In the model of this paper, both M1 and M2 are included as it enables interpretation of whether narrow or broad money supply is more significant. The four macro-economic factors that are chosen to be a part of the model are the oil price, the core CPI and the money supply (M1 and M2).

As examined in the paragraph “2.3.1 The financial accounting factors”, a wide range of accounting factors exist. Peinreich (1932) points out that earnings per share, EPS, is one of the most fundamental accounting factors when evaluating share prices. This is supported in both recent and older academic literature (Modigliani & Miller, 1958) (Chang, et al., 2008). In line with the common business understanding of today, Lintner (1965), found that stock prices are highly dependent on future profits and thereby earnings per share. EPS is a commonly used measure when examining factors that are driving stock prices. Another factor that is considered one of the most fundamental accounting factors is the P/B ratio (Preinreich, 1932). Fama, (1992) found that P/B has great explanatory power of stock price development and the measure is commonly included as an explanatory variable, in academic literature, when models to explain stock prices are constructed. Due to the reasons mentioned above, EPS and P/B are included as explanatory variables in the model developed in this paper.

Similar to what was the case for the macro-economic factors, this paper has chosen not to include a few of the common accounting factors. Discounted cash flow (DCF) is often used for assessing stock values. However, as this paper is examining internet stocks, DCF is not an appropriate measure as it is not uncommon that internet stocks, especially in the early phases, are ‘bleeding cash’ while the stock price is increasing, as the spending is seen as an investment in the future. Furthermore, (McKinsey&Company, 2015) points out that future cash flows of tech-stocks are more difficult to predict than for other stock categories. P/E is a common accounting factor used in stock valuation. However, the distinction between what P/E and P/B explains is not entirely clear (Penman, 1996). This in turn means that a considerable part of P/E’s explanatory power can rightfully be assumed to be explained by P/B, which is already included in the model. Furthermore, the earnings per share component of the ratio is already a part of the accounting factors which enters the model. Based on the above, DCF and P/E will not be included as variables in the model of this paper.

Based on the selection criteria, a number of variables have been chosen to enter the initial model. *Figure 4* provides an overview of these variables.



Variables

Dependent	**Independent**
Index of stock prices	EPS P/B Oil price Core CPI M1 M2

Figure 4 - List of chosen variables
Source: Personal production

3.2.1.2 Method of the model

The following section will briefly explain the methodology of developing the model that will be used for testing whether the independent variables found in the previous section have significant explanatory power of the index price. First, data is gathered for the relevant time periods and descriptive statistics is used to get an overview of the data. Once having an overview of the data, it is necessary to consider which statistical method that will be appropriate for testing the impact that macro-economic factors and accounting factors have on the stock price of leading internet companies. From the literature review in section “2.5 Statistical analysis of stock performance”, it is clear that the multiple linear regression is a common method in research when seeking for causal effect. Therefore, this paper will apply multiple linear regression using OLS estimation. The reason for applying OLS estimation is that if the assumptions of OLS are met, OLS is the best linear unbiased estimator. This follows from the Gauss-Markov theorem (Newbold, 2017).

All the data used in our model are time series, which are often non-stationary. Multiple linear regression requires stationarity. Therefore, after getting an overview of the data, the stationarity of the data is tested. In case the data is found to be non-stationary, which is expected, tests for cointegration and lags of the time series are carried out in order to determine how stationarity can be achieved. Once stationarity is achieved, an initial model including all variables can be estimated. When the model is specified, the data is examined for unusual and influential observations. Furthermore, the assumptions of *normality of residuals, linearity, homoscedasticity, multicolliniarity* and *independence of the error term* are tested[8]. Once the initial model has been tested, the overall fit of the model as well as the significance of the individual variables can be evaluated.

Based on the findings in the initial model, a new model can be specified with the variables that are expected to be significant. In practice, this is an iterative process where only one variable is eliminated at a time. However, due to the limited extend of this paper, as well as the limited contribution of reporting each iteration, the models that will be reported and described are the initial model containing all variables and the final model containing only the variables that are found to be significant. Once the final model has been specified, the data is again examined for unusual and influential observations, and, based on analysis, some observations might be eliminated. Subsequently, the final model is tested for the assumptions of normality of residuals, linearity, homoscedasticity, multicolliniarity and independence of the error term. Once these tests are carried out, the model is tested for omitted variable bias, and the model is then ready for discussion and interpretation.

Once the model for the FANG index has been developed, the same model is applied to the sample of the Chinese equivalent, as well as to the leading pre-dotcom bubble internet companies. By applying the same model to the two other samples, the transferability of the model can be tested. Furthermore, it provides a solid background for discussing whether the same factors are significant and whether their magnitude are similar across geographical regions as well as during different time periods.

---

[8] In order to assess the OLS model, five assumptions must be tested beforehand. (Newbold, 2017)

Figure 5 contains a graphical overview of the steps as described above. For each of the steps, the Stata[9] functions that are used can be seen. The individual functions will be briefly explained during the analysis.

---

[9] Stata is the data analysis and statistical tool that is used when developing the model of this paper.

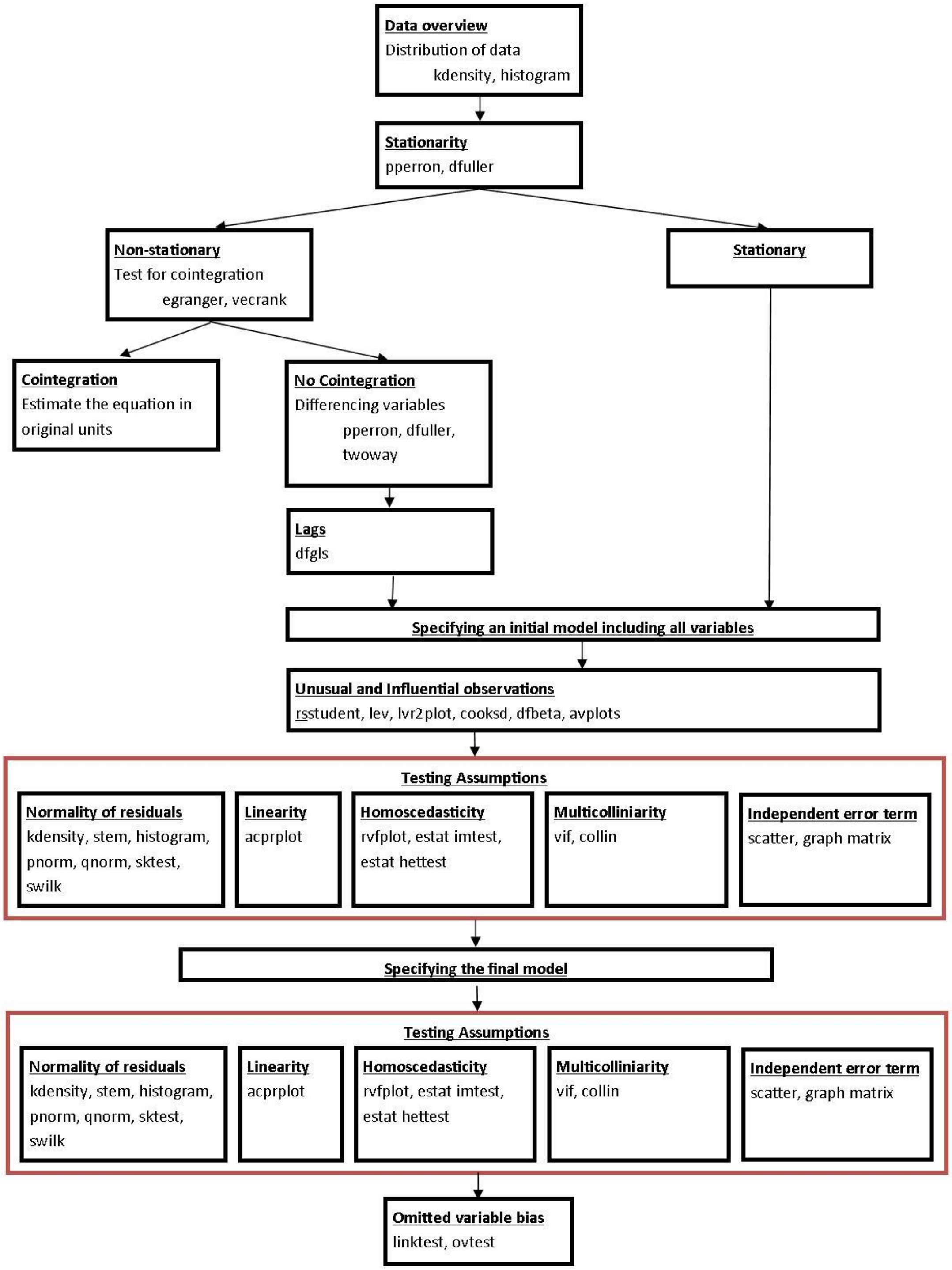


Figure 5- Method for developing model. Source: Personal production

3.3 Data collection

3.3.1 Index

The data has been retrieved from the Datastream platform which is a financial time series database provided by Thomson Reuters (Thomson Reuters Datastream, 2018). A monthly data frequency has been chosen. The choice of monthly data over daily and yearly data frequency is based on the amount of data points required for adequately calculating the factors used for ranking the stocks in the indexes. As the factors and the formation of the indexes are all based on simple calculations with no advanced statistical methods nor any regressions, the required number of data points is limited. While yearly data frequency clearly provides too few data points, daily data frequency is found to be unnecessarily extensive as the value of the additional data points will be very limited. Thus, monthly data frequency has been chosen.

The time period for the analysis has been chosen to be 67 months. As described in the section "3.3 Data collection", the starting point of this thesis was the FANG stocks. Thus, when selecting the time period, we wanted to examine the existing data up until year-end 2017. Among the FANG stocks, the stock that was listed the latest was Facebook, which held its initial public offering (IPO) on 18th of May 2012 (*FACEBOOK INC (FB) IPO - NASDAQ.com*, 2018). As a result, data for Facebook is available from May 2012 and onwards. Thus, the selected time period of analysis for all FANG stocks is May 2012 – December 2017. As a comparative study across the indexes will take place, the duration of the time period for the FANG stocks is used to determine the time period of the two remaining indexes. For Index B, the leading Chinese internet companies, the time period will be the same as for Index A, the FANG stocks. As the purpose of Index C is to determine the leading American internet stocks in the pre-dotcom era, this period's ending point must be prior to the burst of the bubble. As the bubble burst in year 2000 (Risager, 2016), the last period for which data is included is December 1999. As previously explained, the desired duration of the time period is the same as for Index A and Index B. Thus, the time period of Index C must be May 1994 – December 1999.

When the theme of the indexes, as prescribed in the "3.2 Theoretical framework", have been found, it is necessary to find a pool of stocks to test in order to determine which stocks that should enter the index. For Index C, the top 10 NASDAQ companies based on market capitalization prior to the dotcom-bubble is examined. The NASDAQ index peaked on 10th of March which is the date used for finding the top 10 companies. The data for this analysis is readily available in Datastream (Thomson Reuters Datastream, 2018). The Chinese internet companies, which are the subjects of Index B, are not listed on one single index thus the approach used for Index C cannot be applied. For Index B, a list of the top 100 internet companies in China, based on revenue, is examined, and data is collected for the top 10 publicly traded companies (iResearch, 2017). While it is often discussed that data from China can be difficult to both obtain and rely on, the companies we are examining are publicly traded and many are listed outside China e.g. NASDAQ or NYMEX. This means that the data provided complies with international standards and can as such be considered reliable. iResearch, which provides the list, is a Chinese provider of data products, analytics and consulting services.

3.3.2 Econometric model

The dataset for the econometric model consists of the data for each of the variables that enter the model. The data for the dependent variable, the index price, was described in the previous section. Thus, this section will only concern the data collection for the independent variables.

Accounting variables will be collected for the individual stocks, and then a weighted average is calculated based on the weights that are found when calculating the index. Based on this method, the weights are adjusted on a yearly basis. For a full overview of the method, see section “4.1 Index creation”. After creating the weighted average, two accounting factor variables exist for each stock sample: FANG_MTBV, FANG_EPS, CH_MTBV, CH_EPS, US_MTBV and US_EPS. As the macro variables are not firm specific, these will be used without any further adjustments once retrieved from Datastream (Thomson Reuters Datastream, 2018).

The data, for each of the independent variables, was extracted from the Datastream (Thomson Reuters Datastream, 2018) with the following definitions:

Table 2 - Variable description, Source: Personal production

VARIABLE	DATASTREAM CODE	DESCRIPTION
MTBV	MTBV	MTBV is defined as the market value of the ordinary (common) equity divided by the balance sheet value of the ordinary (common) equity in the company. The MTBV is the ratio as represented by the following Datastream codes MVC/WC03501. Datastream's source for MTBV data is Worldscope[10].
EPS	EPS	EPS is derived from interim period earnings and the value change at irregular intervals subject to when updated earings data is available. This means that data frequency is not fixed to e.g. quarterly updates. Datastream's source for MTBV data is Worldscope.
OIL	OILWTXI	The oil price is measured as 'Crude Oil West Texas Intermediate' (WTI) which trades on New York Mercantile Exchange (NYMEX). The price is the close of near month in USD/bbl. Datastream's source for the WTI price data is NYMEX.
CPI_US	USCPCOREE	The CPI is measured as the core CPI (CPI excluding food and energy) as provided by the Bureau of Labor Statistics (BLS)[11]. Index 100 = 1982-1984. The Index is seasonally adjusted and contains urban consumers only. Datastream's source for the core CPI data is BLS.

---

[10] The Worldscope database is provided by Thomson Reuters and offers fundamental data on the world's leading public and private companies

[11] Latest release: https://www.bls.gov/news.release/cpi.htm

M1_US	USM1….B	M1 consists of (1) currency outside the U.S. Treasury, Federal Reserve Banks, and the vaults of depository institutions; (2) traveler's checks of nonbank issuers; (3) demand deposits at commercial banks (excluding those amounts held by depository institutions, the U.S. government, and foreign banks and official institutions) less cash items in the process of collection and Federal Reserve float; and (4) other checkable deposits (OCDs), consisting of negotiable order of withdrawal (NOW) and automatic transfer service (ATS) accounts at depository institutions, credit union share draft accounts, and demand deposits at thrift institutions. The value is adjusted to current prices and is also seasonally adjusted. Datastream's source for the core CPI data is US Federal Reserve[12].
M2_US	USM2….B	M2 consists of M1 plus (1) savings deposits (including money market deposit accounts); (2) small-denomination time deposits (time deposits in amounts of less than $100,000), less individual retirement account (IRA) and Keogh balances at depository institutions; and (3) balances in retail money market mutual funds, less IRA and Keogh balances at money market mutual funds. The value is adjusted to current prices and is also seasonally adjusted. Datastream's source for the core CPI data is US Federal Reserve**Error! Bookmark not defined.**.

Source: Personal production

---

[12] Latest release: https://www.federalreserve.gov/releases/h6/current/default.htm

Based on the previous section, *Table 3*provides an overview of the variables and time periods that have been used for the data collection.

Table 3 - Overview of time period. Source: Personal Production

STOCKS	**TIME PERIOD**	**VARIABLES**
FACEBOOK, AMAZON, NETFLIX, GOOGLE	1/6/2012 – 31/12/2017	*Accounting variable: MTBV,* EPS *Macro variables:* OIL, CPI_US, M1_US, M2_US
TENCENT, BAIDU	1/6/2012 – 31/12/2017	*Accounting variable: MTBV,* EPS
ALIBABA, JD.COM	1/1/2015 – 31/12/2017	*Accounting variable: MTBV,* EPS
NETEASE, CTRIP	1/6/2012 – 31/12/2014	*Accounting variable: MTBV,* EPS
MICROSOFT, CISCO, INTEL, DELL	1/6/1994 – 31/12/1999	*Accounting variable: MTBV,* EPS

During the calculation of the Chinese index values, it was found that values did not exist for the full period for Alibaba and JD.com. As when computing the index value, these stocks will be replaced by Netease and Ctrip until the first year where data is available for all months for Alibaba and JD.com. The reason for only including the macro-economic variable in the FANG sample will become evident when the analysis results are reported in section “4. Analysis”.

When examining the data retrieved from Datastream, the entire data series for EPS of JD.com is found to be missing. This value is instead retrieved from Bloomberg that reports the data on a quarterly basis.

3.4 Data calculations

3.4.1 Index

The first step of creating the index, once the pool of stocks has been determined, is to rank the stocks to select the four stocks that must enter the index. This is done by using the market capitalization of the stock as discussed in "3.2 Theoretical framework". The market capitalization is available in Datastream. In the selection, the average of the market capitalization for the full time period is used.

The three factors used for calculating the index weights, as presented in the "3.2 Theoretical framework", are calculated as follows:

1. The *market capitalization* is calculated as the average of the full time period.
2. The *stock returns* are calculated as monthly discrete compounded returns and then converted to yearly values.

   $r_t = \frac{V_t - V_{t-1}}{V_{t-1}} \quad t = 1 \ldots T$ where V_t is the stock price at time t and r_t is the return at time t

   Conversion to yearly values $r_{yearly} = (1 + r_{monthly})^{12} - 1$
3. The *volatility* is calculated as the monthly standard deviation of the returns and then converted to yearly values.

   Excel function stdev.s is used for calculating the monthly value

   Conversion to yearly values $\sigma_{yearly} = \sqrt{12} * \sigma_{monthly}$

The ranking of each stock within the three selection factors is as follows

1. Market Capitalization: 1-4. Highest market capitalization = 4
2. Stock return: 1-4. Highest market capitalization = 4
3. Volatility: 4-1. Highest volatility = 1 (inverted scale)

Based on the ranking, a score for each factor is calculated by summing the ranking i.e. a stock with the highest market capitalization, the highest stock return and the lowest volatility would receive the score 12 (4+4+4=12).

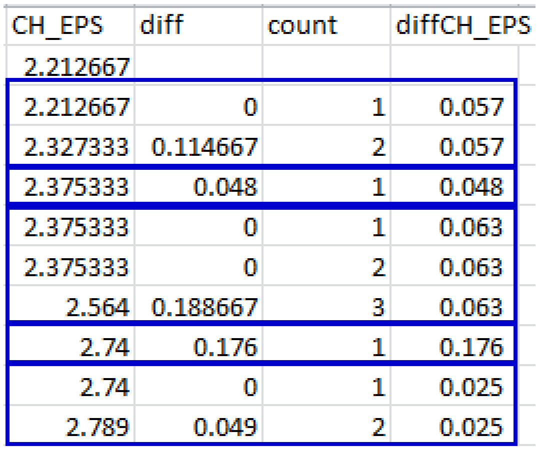
The score that is calculated is then used for calculating the index weight. The index weight is calculated as the score divided by a common divisor. The common devisor that occurs when ranking four stocks based on three factors is 30 ((1+2+3+4)*3 = 30). Based on the score 12, which is the maximum achievable score, one stock can never enter the index with a weight exceeding 40% (12/30 = 0.4).

Consideration must be given to ensure that the indexes, throughout the full time periods, reflect the development of leading internet companies rather than just one or two companies. If any of the stocks that are included in one of the indexes does not contain data for the full index period, as defined in the "3.3 Data collection", the stock will enter the index on the first year where data is available for all months. In the period until the stock enters the index, a replacement stock will be included in the index. The replacement stock is selected in a similar manner as the original four stocks in the index. This means that if one of the four stocks does not contain data for the full period, the stock ranked as number five will enter the index as the replacement stock. This approach ensures that the index will always consist of four stocks. This is a desired feature as an index might otherwise only consist of one or two stocks in the early period of the index which would mean that the index would simply reflect one or two companies rather than the "leading internet companies", which is what we seek to mirror by the index.

The correlation matrix of the stock in each index is calculated by using Excel's 'correl' function.

3.4.2 Econometric model

As will be described in the section "4. Analysis", it has been necessary to carry out a manual differencing of the EPS variable in each of the three datasets: FANG, US and CH. This section will elaborate how the manual differencing is carried out. First, differencing is carried out as it would be done by any statistical program where observation one is deducted from observation two and observation two is deducted from observation three etc. However, as similar values exist for consecutive periods for EPS, this will create a number of zero values. These values do not reflect reality as there was in fact gradual earnings per share during the period. In order to make the differenced variable reflect reality, the data is smoothened such that the difference is divided over the period since the previous change in EPS (see *Figure 6)* where the column "diff" represents the traditional differencing, and the column “diffCH_EPS” represents the smoothened data that are used in the analysis.



CH_EPS	diff	count	diffCH_EPS
2.212667			
2.212667	0	1	0.057
2.327333	0.114667	2	0.057
2.375333	0.048	1	0.048
2.375333	0	1	0.063
2.375333	0	2	0.063
2.564	0.188667	3	0.063
2.74	0.176	1	0.176
2.74	0	1	0.025
2.789	0.049	2	0.025

Figure 6 - Differencing Process. Source: Personal production

3.5 Validity and reliability

Quantitative researches should provide results based on well-conducted research. Therefore, it is extremely important to report considerations on both findings and the consistency of the analysis and the data sources that are used (Heale & Twycross, 2015). In quantitative researches, the analysis and data sources are evaluated based on the *validity*[13], *reliability*[14] (Golafshani, 2003).

3.5.1 Reliability

The following section will discuss the reliability of the data sources used for gathering the data for the analysis as well as reliability of the measures that are used.

Three main sources have been used to gather the required data. The first source is the Datastream platform, provided by the media and intelligence company; Thomson Reuters. Datastream represents a widely consolidated database of both macroeconomic and company specific time series data (Thomson Reuters Datastream, 2018). Datastream delivers a wide range of contents, which are sorted in eleven distinct categories, including for instance “equities”, “bonds” and

---

[13] Validity refers to the extent to which a concept is correctly analyzed and measured in a quantitative study (Heale & Twycross, 2015)

[14] Reliability defines the consistency of the measure (Heale & Twycross, 2015)

“Commodities” (Copenhagen Business School, 2018). Thomson Reuters is a well reputed intelligence service wherefore it can be assumed that they have procedures in place to ensure that the data quality is of the best possible quality. It is however important to note that Thomson Reuters are dependent on other sources for much of their data and the data that is available in Datastream can as such not be better than what is made available to Thomson Reuters in the first place.

The second source is the Bloomberg L.P. database. The Bloomberg platform has been used in order to integrate the missing data that were not available Datastream. Specifically, it has been used to extract the earnings-per-share (EPS) information for the Chinese company JD.com. As Bloomberg L.P. is one of the most reputable financial news service providers, the reliability is expected to be as good as practically possible.

The third source that is used is the Chinese platform iResearch. The iResearch service is used to identify the biggest Internet companies in China. iResearch is a well-established analytics service provider based in Beijing (iResearch Global, 2017). iResearch is among the best providers of financial information related to the Chinese stock markets. This is supported by the fact that some of the world’s most well reputed financial journals e.g. Bloomberg Markets, use iResearch as their data provider for the Chinese market (Bloomberg L.P. , 2018). Since China is a very special case of low transparency market, the authors have decided to trust the reliability of iResearch, as it represents the “best-in-class” financial provider for China.

While it is often discussed that data from China can be difficult to both obtain and rely on, the companies we are examining are publicly traded and many are listed outside China on e.g. NASDAQ or NYMEX. This means that the data provided complies with international standards and can as such be considered reliable.

The reliability of the measures included in our analysis is high. As all the measures that are used in the analysis, both the macro-economic and the accounting variables, are naturally measured and reported on a continuous scale, the reliability of the measures mainly rely on the reliability of the data source which have already been discussed. Had the analysis contained more abstract measures such as happiness and satisfaction, it could be more difficult to ensure the consistency of the research measures. As all measures that are included in the model are tangible and naturally measured on a continuous scale, it is found that the reliability of this paper is good, as best in class data providers have been used.

3.5.2 Validity

Two aspects of validity will be discussed in the following section: *content validity* and *external validity* (National Business Research Institute (NBRI), 2018). Content validity concern whether measures are accurately assessing what we want to know, while external validity discusses whether the results can be generalized.

The content validity can be evaluated for the underlying variables in the analysis, the indexed variables and the model itself. The validity of the variables: price, market-to-book value, earnings-per-share, oil price, CPI, M1 and M2 are all good as they measure exactly what we wish to include in our analysis. However, as earnings-per-share only are available at irregular intervals, it might have negative effects on the parameter estimations and thereby on the validity.

As described in the section “3.3.1 Index”, the price is converted to a price index by the share-weighted method and the weights of each stock are then used for calculating the weighted average of each of the other variables. In this process two choices that impact the validity are made. First, the selection of which stocks that should enter the US and the CH index is made. The objective of the samples is to represent the biggest Internet companies, thus the use of market capitalization to select the stocks can be considered to provide a method with high validity. Second, the weights that are used for both the price index and to calculate the weighted average of all the independent variables are calculated by the share-weighted method. The detailed discussion of this choice can be seen in “3.2 Theoretical framework” but the main driver for selecting this method, rather than using a more conventional index methods such as e.g. the market capitalization weighted method, was to achieve high validity. The advantage of the share-weighted method is that the calculated value represents the nature of all the stocks, rather than one dominating stock, while still attributing some extra weight to the most dominant stocks. The general content validity of our model is found to be high, as the data that is retrieved from the databases are measuring exactly what we wish to measure and because the method used for calculating index weights is chosen to support the validity.

The external validity of the model is to a great extend tested due to the analysis design. As the model is applied on three different samples during the analysis the generalizability of the model has been tested. The US sample contains data which covers a time period that is different from the one for which the model is originally developed. As the analysis reveals, the results are comparable. The CH sample contains data for companies with a different origin than that of the FANG stocks and the results are also found to be comparable. Based on above, external validity has been demonstrated across two time periods and across geographical regions.

3.6 Limitations

While the analysis of the paper has been designed to reduce the limitations, some inevitable limitations exist as a result of the choices made regarding the analysis design. The following section will highlight the limitations of the analysis and the findings of this paper.

A serious limitation of the paper is that no prior research has had a clear focus on establishing the factors that drives the prices of a specific group of high performing Internet stocks by a quantitative method. This implies that the choice of method and the decision on which factors to test has been based on research from other related areas such as the performance of broader stock indexes over longer time periods. The lack of existing research within the area has meant that the paper has been applying an approach that has support in past research but has not previously been proven to work on the specific topic of this paper. The impact on the result is that while some findings are considered both reliable and valid others are found to be less reliable. Furthermore, the work has led to several suggestions for future research.

As discussed in “3.3 Data collection” the analysis covers a limited time period which is a natural consequence based on when the stocks were first listed on the stock exchange. Furthermore, the analysis is only carried out during ‘boom’ periods. As the analysis is carried out for a limited time period and only during periods where Internet stocks develop very positively, the findings cannot be claimed to also hold for longer time periods including e.g. periods containing more depressed market conditions for Internet stocks. The generalizability is further impacted by the choice of focusing on the four largest stocks in each sample. This means that the results cannot be claimed to be valid for a broader range of Internet stocks but only for the very largest, global companies.

A natural result of the limited time period and the selection of a monthly frequency of data is that our samples only consist of 67 observations prior removing outliers, carrying out differencing and lagging variables. While the sample size required for OLS is widely debated, it is clear that 67 observations are in the lower range of the accepted number of observations. An issue of a small sample size tends to be difficulties in finding significant variables. However, as we have seen from the results our analysis does not seem to suffer from this. The limited number of observations are however a restriction when evaluating how many observations that can be removed as outliers or influential observations.

The impact of outliers is an important limiting factor when interpreting the analysis results. When working with the data and examining the outputs, it is clear that the inclusion or deletion of the outliers have a huge impact on both the magnitude of the coefficient and the significance of EPS. The impact is of such an extent that the findings for EPS cannot be relied upon despite removing the most significant outliers. It is noteworthy that many of the outliers are the first month of the year. An explanation for this could be that the index weights are adjusted on a yearly basis.

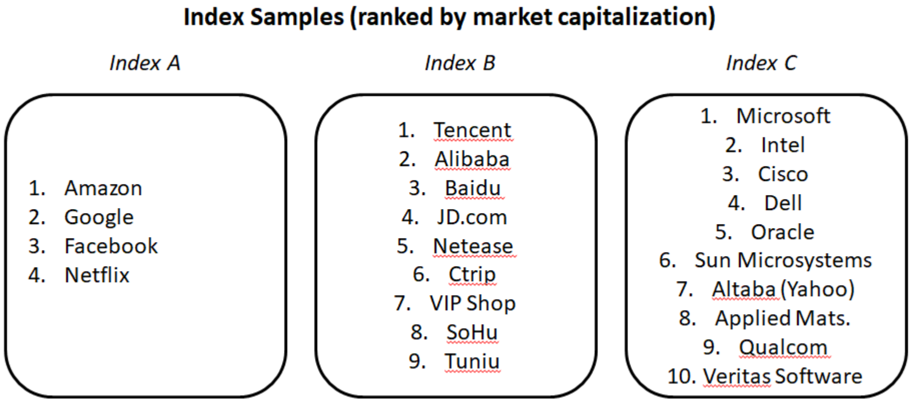
The analysis of omitted variable bias for the FANG model and the US model indicates that omitted variable bias exists, while the analysis of the US model does not indicate presence of omitted variable bias. If a model is specified without a significant predictor variable, the least squares estimates of coefficients are often biased (Newbold, 2017). Leaving out a significant predictor variable will typically also have a negative impact on the R^2 of the model. From our analysis result we notice that all models have a reasonably high R^2. However, it cannot be ruled out that our model does not suffer from omitted variable bias in line with what the formal tests indicate.

Overall some considerable limitations of this paper exist, in particular concerning generalizability for longer time periods and a broader stock category. Furthermore, the results of EPS are unstable and cannot be considered reliable.

4. Analysis

4.1 Index creation

In the case of Index A, the stocks to enter the index are described by the FANG acronym: Facebook (FB), Amazon (AMZN), Netflix (NFLX) and Google (GOOG). The purpose of Index B is to represent the leading Chinese internet stocks. This is done by considering the ten Chinese internet companies with the highest revenues in 2016 that are publicly traded. The purpose of index C is to find the leading internet companies in the time leading up to the dotcom-bubble. This is done by considering the ten biggest stocks, as measured by market cap, on NASDAQ on 10th of March 2000 i.e. shortly before the dotcom-bubble burst. Based on the above, three samples of stocks have been found. In *Figure 7*, the stock samples for each index can be seen.



Index Samples (ranked by market capitalization)

Index A

1. Amazon
2. Google
3. Facebook
4. Netflix

Index B

1. Tencent
2. Alibaba
3. Baidu
4. JD.com
5. Netease
6. Ctrip
7. VIP Shop
8. SoHu
9. Tuniu

Index C

1. Microsoft
2. Intel
3. Cisco
4. Dell
5. Oracle
6. Sun Microsystems
7. Altaba (Yahoo)
8. Applied Mats.
9. Qualcom
10. Veritas Software

Figure 7- Index Samples

Source: Personal production

The four stocks that will form Index B and Index C are selected based on the market capitalization.

Data for two of the candidates to enter Index B is only available at a later date than the start of the time period of the index. The two stocks are JD.COM and Alibaba. The stocks are evaluated based on the period where data is available. The stock "E Commerce Data" is deleted due to inconsistent data. The company can be left out of the sample, as it based on other data, is a far smaller company than the companies that are expected to enter Index B (iResearch, 2017).

As discussed in “3.3 Data collection”, the index must consist of four stocks throughout the period we examine. For both JD.com and Alibaba, the first year where data is available for all months is 2015. Thus, they will only enter the index for 2015 onwards. In the period 2012-2014, the stocks ranked 5th and 6th in the index sample, Netease and Ctrip, will enter the index. Based on the above, the final three indexes will look as depicted by *Figure 8*.

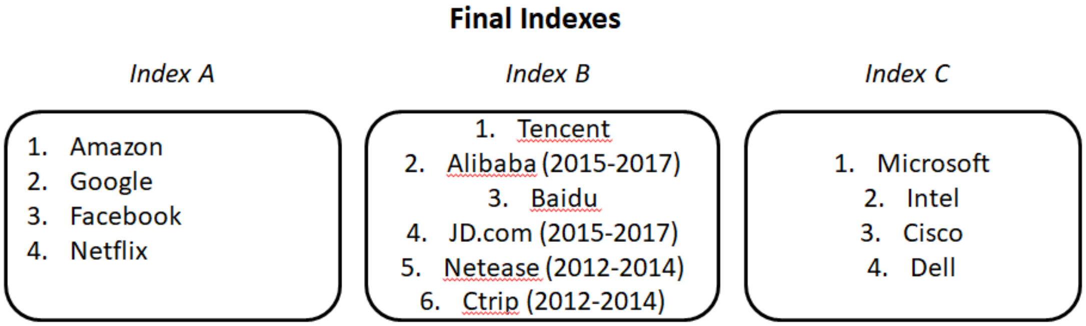


Figure 8- Final Indexes

Source: Personal production

4.1.2 Weights and correlations

As it has been determined which stocks that should enter each index, the yearly weight of each stock can be calculated, according to the method discussed in “3.4 Data calculations”. In the appendixes, the impact factor, the ranking within each factor and the subsequent weight of each stock can be found:

- Appendix X: Index A
- Appendix X: Index B
- Appendix X: Index C

The price correlation matrixes for each index are found to be as follows:

Index A

2012-2017	F	A	N	G
F		0.956883	0.969168	0.977891
A			0.951500	0.958293
N				0.969944
G				

Index B

2012-2014	NT	CT	BA	TE
NT		0.811762	0.917938	0.860732
CT			0.849888	0.912084
BA				0.871970
TE				

2015-2017	JD	AL	BA	TE
JD		0.748474	0.576445	0.721757
AL			0.701152	0.948254
BA				0.515034
TE				

Index C

	MSFT	CSCO	INTC	DELL
MSFT		0.971203	0.951760	0.977056
CSCO			0.920617	0.945401
INTC				0.910077
DELL				

4.1.2 Index prices

As described in “3.3 Data collection”, the index prices are normalized such that all indexes are starting at index 100. *Figure 9* (A) shows Index A and Index B for the period May 2012 – December 2017. On *Figure 9* (B) all three indexes are included in the same graph. The x-axis displays the time period i.e. period 1 = 1st month of data, period 2 = 2nd month of data etc.

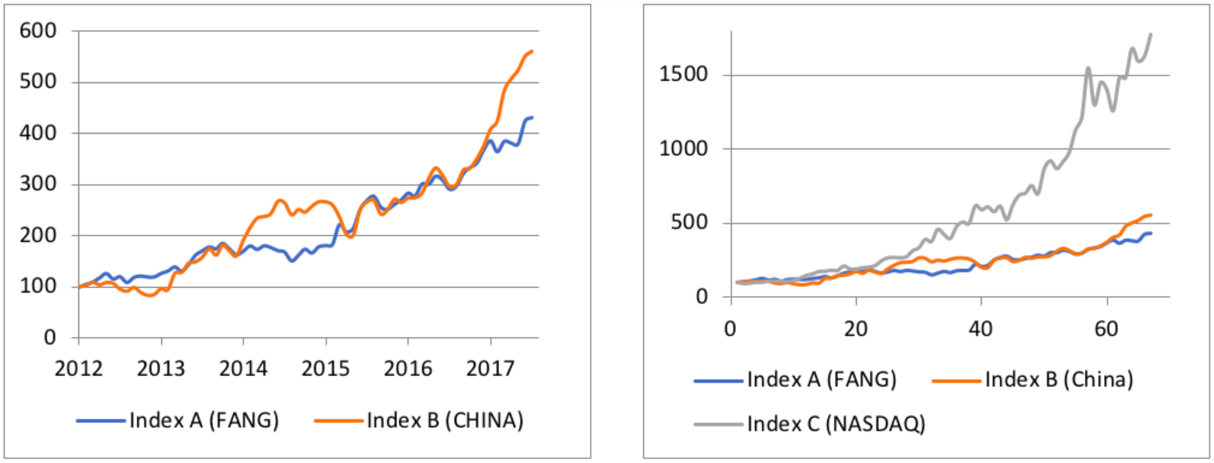


Figure 9 - (A) Plot of Index A and B; (B) Plot of Index A, B and C combined

Source: Personal production, data: Datastream, Bloomberg

4.2 Building the initial model

As the indexes have now been clearly defined, and the relevant data collection has been carried out, in line with section “3.3 Data collection”, the relevant data can now be imported to our data analysis tool, *Stata*. The dataset consists of three data samples that can be seen from *Figure 10*.

As previously discussed, the model that is to be developed will be for the FANG stocks. Thus, the model development, in the subsequent sections, is carried out on the FANG data sample.

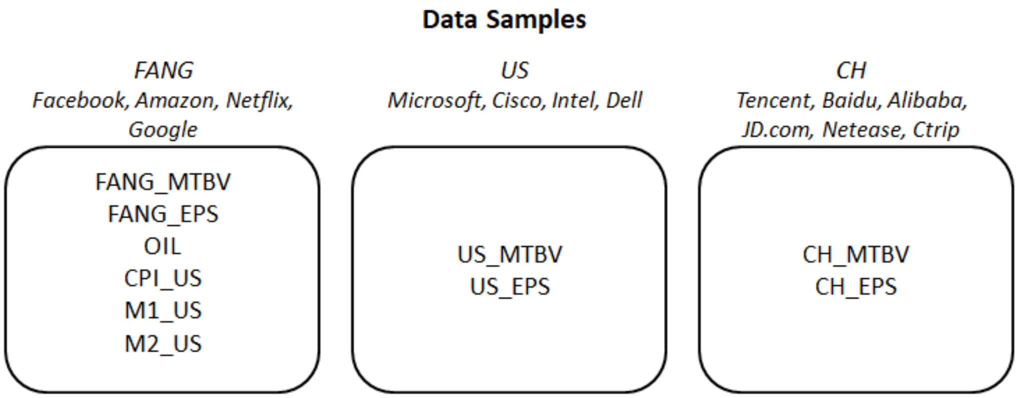


Figure 10 - Data samples. Sources: Personal Production

The structure of the analysis and the tests that are carried out will follow the structure as presented in “3.2.1.2 Method of the model” on *Figure 3*.

4.2.1 Data overview

The purpose of the data overview is to, by simple means, get an overview of the data.

Appendix “8.1 FANG model – all variables” shows that all variables contain 67 observations, and it can thereby be concluded that there are no missing observations. From the graphical overview, it is noticed that CPI_US, M1_US and M2_US seem to have a clear linear relationship with INDEX. The relation between FANG_MTBV, FANG_EPS and INDEX is less clear but also seems to be linear, while the relationship between OIL and INDEX shows a non-linear relationship. The relationship between OIL and INDEX does not seem to follow another common, but non-linear, trend. Thus, the variable OIL is not transformed.

After this quick overview, the distribution of the data for each individual variable is examined. This is done for each variable by examining a histogram as well as using a univariate kernel density estimation plot with a normal density graph fitted to the plot. The Kernel density estimation is a technique that is used to indicate the density from a set of observations (Salgado-Ugarte, et al., 1993). Histograms also visualize the density for different intervals. However, due to underlying differences, the visualization differs a bit which make the two methods complementary. The results can be found in the appendix “8.1 FANG model – all variables”. When examining the kernel density plot and the histogram, they show clear signs of INDEX being skewed to the right which is in line with expectations as the INDEX is known to be an increasing time series. The kernel density plot for FANG_EPS also seems to be right skewed. However, based on the histogram, the skewedness of the observations is not as distinctive as for INDEX. FANG_MTBV, CPI_US, M1_US and M2_US all have a ‘flat’ top and thicker ‘tails’ in the kernel density plot compared to the normal distribution density. Thus, based on visual inspection of the kernel density, the kurtosis seems to differ between the normal distribution and the distribution of FANG_MTBV, CPI_US, M1_US and M2_US. When examining the aforementioned variables by the histograms, one could consider if they are evenly distributed rather than normally distributed. OIL has two distinctive peaks and hence differs from the normal distribution.

Based on the above, it is found that several of the variables do not seem to be normally distributed. However, this does not necessarily cause any issues as only the residuals are required to be normally distributed according to the assumptions of OLS. As the examination of INDEX and FANG_EPS shows signs of being right skewed, we carry out a log transformation of these variables and examine the newly generated variables. As can be seen from the kernel density plot and the histogram in appendix “8.1 FANG model – all variables”, it is clear that the logged values of the INDEX are better distributed after the transformation. The kernel density plot of FANG_EPS also looks better, however, the histogram does not show normally distributed observations.

When deciding whether to log transform variables or not, it is also of relevance how the variable will be used in the model. INDEX will be the dependent variable in our model and log transforming it will enhance the interpretability as the coefficients of the independent variable will then represent the percentage change that one unit increase in the independent variable has on INDEX. For FANG_EPS, it is a little different. Interpreting the effect of a log transformed variable on a log transformed dependent variable (log-log relationship) can easily be done mathematically. However, logically it is not as easily interpreted and understood as a log-linear relationship. As the log transformation of FANG_EPS does not considerably improve the data distribution, as noted by inspection of the histogram, and that a log transformation of FANG_EPS would complicate the interpretation of the model, it is decided not to log transform FANG_EPS[15].

By carrying out the steps in the data overview, we found that it is necessary to carry out a log transformation of INDEX while all other variables remain in their original form.

4.2.2 Stationarity, cointegration and lags

As discussed in “2.5 Statistical analysis of stock performance”, there is a high risk of our data being non-stationary as this is common for time series data. The stationarity is tested by using the Dickey Fuller test for unit-roots as well as Phillips-Perron’s unit root test.

The Dickey-Fuller and the Philips- Perron tests are two statistical techniques that investigate whether the variable of interest follows a unit-root process (Dickey & Fuller, 1979) ((Phillips & Perron, 1988). Both tests assume that the variable contains a unit root, as null hypothesis (H0); while the alternative hypothesis (H1) states that the variable is generated by a stationary process (Fuller, 1996). In statistics, it is very common to run both test simultaneously in order to provide stronger validity to the results.

The detailed test results of the Dickey Fuller test and the Phillips-Perron test can be found in the Stata log file in appendix “8.6 Log files”. At a 5% significance level, both tests fail to reject H_0, that the series is a random walk with a possible drift, for all variables, except for M1_US. In other words, our tests indicate that all variables except M1_US are non-stationary. As M1_US is of similar nature as the other variables and is theoretically closely linked to CPI_US and M2_US, it would be expected that the results of the tests are similar. Based on this knowledge and the fact that time series often lack stationarity, it is decided to treat all variables in a similar way. Thus, all variables are considered to indicate lack of stationarity.

As seen on our methodology overview *Figure 3*, it is necessary to carry out tests for cointegration to determine the next steps on how to develop a model when the variables suffer from lack of

---

[15] All tests for the model and the model itself have also been carried out with FANG_EPS log transformed. The results for tests and the final model does not deviate significantly.

stationarity. The Engle-Granger test of cointegration and Johansen's test of cointegration will be applied to test for cointegration.

The Engle-granger (EG) and Johansen's tests are the most commonly used techniques for assessing the presence of cointegration in time series (Bilgili, 1998). The Engle-granger (EG) is a two-step residual-based test that is used to investigate whether cointegration is present (MacKinnon, 1990). If the null hypothesis is rejected, the time series has cointegration (Engle & Granger, 1987). The Johansen test of cointegration estimates and tests the presence of multiple cointegrating vectors (Becketti, 2013). Because it does not limit the analysis uniquely to residuals, the Johansen test for cointegration is a more exhaustive method when carrying out cointegration analysis (Bilgili, 1998). Nevertheless, best practice suggests using both tests in order to strengthen the validity of the outcomes.

The detailed test results of the Engle-Granger test and the Johansen's test of cointegration can be found in the Stata log file in appendix "8.6 Log files". As the value of the test statistics for the Engle-Granger test is above both the 5% and 10% critical value, we fail to reject H_0. Thus, we find that the variables are not cointegrated. Johansen's test of cointegration finds that one cointegration equation exists in the vector error-correction model[16].

From the analysis above, it is not entirely clear whether cointegration exists or not. For now, it is assumed that cointegration does not exist. To ensure that this is indeed the case, we will run Johansen's test of cointegration again once the variables for the final model have been chosen.

Based on the above, the variables are determined to be non-stationary and not cointegrated. From *Figure 3,* it can be seen that stationarity can be achieved by differencing. After carrying out the differencing, the first observation in the series will be dropped as it no longer contains data as a result of the differencing. After differencing, the Dickey Fuller and the Phillips-Perron tests for unit roots are carried out again to test if stationarity has been achieved. As can be seen from "8.6 Log files"., both tests indicate that stationarity has been achieved for all variables at a 5% significance level.

---

[16] The Vector Error Correction Model (VECM) represents a statistical technique for assessing cointegration among two or more times series (Newbold, 2017)

While Stata has been used for generating the differenced variables; diffLOGINDEX, diffFANG_MTBV, diffOIL, diffCPI_US, diffM1_US and diffM2_US the differencing was carried out manually for diffFANG_EPS. The reason for carrying out manual differencing lies with the nature of the diffFANG_EPS. As described in the section "3.2.1.1 Variables in the model", the values do not change on a monthly basis but are instead changing at irregular intervals depending on when new earnings information becomes available. The issue with such data is that standard differencing by Stata will create '0' values in months where a change in earnings did in fact occur. This happens when the same value exists for e.g. two months. Once the differencing is carried out, it will indicate that there were '0' changes in earnings in the first month, whereas the reality would normally be that the change happened gradually over the two months. The method used for calculating the manual differencing can be found in "3.4 Data calculations".

As stationarity is defined as the variable being mean reverting and having constant variance, we can graphically plot this for each of the variables over time. The results can be found at "8.1 FANG model – all variables". For all of the variables, the graph clearly indicates that they are mean reverting. The constant variance is more difficult to interpret visually. However, there are no signs of e.g. seasonality, and the series seems to be displaying close to constant variance over time.

As stationarity of all series have now been achieved, we need to check if any of the series need to be lagged according to *Figure 3*. To test the number of lags required, we use a modified Dickey-Fuller t test for unit roots on lag 1-5.

The modified Dickey-Fuller t test represents an adaptation of the Dickey-Fuller t test, and it is used when the time series has been reformulated by a generalized least-square regression (Elliott, et al., 1996). The only difference between the standard Dickey-Fuller (dfuller) test and the modified version (dfgls t) is that the time series is processed and regressed in the second case. Elliott et al. (1996) have proved that the modified Dickey-Fuller t test has a greater significance than the normal test (Stata.com, 2013).

The detailed results are available in the log file. By using the ng-perron t, we find that the following variables should be lacked diffFANG_EPS (2 lags), diffOIL (2 lags), diffM1_US (4 lags). The number of suggested lags differs for diffOIL and diffM1_US dependent on which measure is used for evaluation (ng-perron seq t, min SC or min MAIC. For diffFANG_EPS the measures are identical, and all suggest two lags.

Based on theory and logical reasoning, it is not unlikely to observe that a change in a macro-economic factor only has an impact on the company after a certain period of time as it might take time before a macro-economic event impacts the immediate environment that surround the company. For the accounting factors such as FANG_EPS, it would however not be expected to observe a lag as one would expect that the company's earnings per share has a direct and immediate impact on the stock price. As lagging diffFANG_EPS by two periods does not have any theoretical or logical support, it is decided to not lag the variable in the model that is created.

4.2.3 Formulating the initial model

As data transformation has been carried out, stationary variables now exist and the relevant lags are known, we can specify our initial model where all variables that are expected to be relevant, based on the chapter "2.3 Introduction to factors that have an impact on Internet stock prices", are included.

Our hypothesis is that the following model will be significant:

$$H_1: difflogINDEX = \beta_0 + \beta_1 diffFANG_{MTBV} + \beta_2 diffFANG_{EPS} + \beta_3 L2.diffOIL + \beta_4 diffCPI_{US} + \beta_5 L4.diffM1_{US} + \beta_6 diffM2_{US} + \varepsilon$$

4.2.3.1 Unusual and influential observations

In order to identify unusual and influential observations, we look at the studentized residuals, leverage, leverage-versus-squared-residual plot, dfbeta influence statistics and added-variable plot.

In order to evaluate *unusual and influential observations,* two important statistical concepts are considered: *Outliers* and *high leverage points* (Stata.com, 2013). Tests for outliers (residuals tests) and tests for leverage are used for evaluating the influence of an outlier. Outliers with high residuals indicate that they are far away from other observations in the y-direction when considering a simple regression plot. High leverage points are points that are far away from the other observations in the x-direction when considering a simple regression plot. If outliers have high leverage, it implies that they are very influential. Influential observations are observations that dramatically change the regression line, i.e. the parameter estimates, if they are deleted.

The Studentized residuals test represents a statistical approach where residuals are adjusted with respect to their standard errors (Bollen & Jackman, 1990). In order to assess the influence, the combined effects of residuals and leverage can be examined in a leverage-versus-squared-residuals plot. In fact, the leverage-versus-squared-residual graph represents one of the most useful diagnostic plots for influential observations (Stata.com, 2013). The dfbeta value and cooks d are computed to display the *influence* on a given parameter if the observation is deleted.

When examining the output of all of the tests, see “8.6 Log files” and appendix “8.1 FANG model – all variables” for the full set of results, the following point stands out as the most unusual and influential observations: 2016-01.

While the individual diagnostic tools all found more unusual and influential observations than the one listed above, the number of deleted observations is limited to 2016-01. In the selection, consideration has been given to the fact that outliers can be real observations and thus capture information about the population we seek to understand. The observation that is deleted is an example of highly unusual observation with a high impact on the regression.

4.2.3.2 Testing assumptions

As the most unusual and influential observation has been removed, it can now be tested whether the assumptions for the OLS regression are met for our initial model H_1.

Normality of residuals

In order to test the normality of residuals, we look at the kernel density estimation plot of the residuals, the histogram of the residuals, quantiles of residuals against quantiles of normal distribution, standardized normal probability plot, Shapiro-Wilk and Shapiro-Francia tests for normality and carry out tests for skewness and kurtosis. The H_0 of both Shapiro-Wilk's and Shapiro-Francia's tests are that the data is normally distributed.

The detailed results of the tests that can be found in "8.6 Log files" and in the "8.1 FANG model – all variables". Both the kernel density estimation plot and the histogram show clear signs of a normal distribution but with more residual observations centered around the mean. Quantiles of residuals against quantiles of normal distribution show clear indications of normality (though with a slight s-curve trend), while the standardized normal probability plot has a small deviation for large positive values of inverse normal. Both the test for skewness and kurtosis indicate normality. The Shapiro-Wilk and Shapiro-Francia tests for normality find that the data is normally distributed.

Based on both formal tests and visual inspections of the residual data, we conclude that the residuals are normally distributed with only minor deviations from normality.

Linearity

In order to evaluate the assumption of linearity, we use a graph matrix of all the variables against INDEX, and we produce an augmented component-plus-residual plot for each variable. The augmented component-plus-residual plot outlines the nonlinearities in the data (Mallows, 1986).

The outcome can be seen in appendix "8.1 FANG model – all variables". When analyzing linearity, the main concern would be data showing a clear non-linear pattern. From the graph matrix and the augmented component-plus-residual plot, there is no indication of any non-linear relationships. Consequently, it can be concluded that the assumption of linearity is met.

Homoscedasticity

To test one of the main assumptions of OLS, if the data is homoskedastic, a visual inspection is carried out by examining the residual-versus-predictror plot (rvf plot), where after both Breusch-Pagan / Cook-Weisberg test and White's test for heteroskedasticity are carried out.

Heteroskedasticity indicates that the variance of the error terms varies across observations. Conversely, homoskedasticity refers to the situation where the variance of the error terms stays constant across all observations (Newbold, 2017). In order to apply the OLS regression model, homoskedasticity must be assumed and tested.

The results of the formal tests can be found in "8.6 Log files" whereas the rvf plot can be found in appendix "8.1 FANG model – all variables". Both the Breusch-Pagan test and White's test show that we have to reject H_0, that the variance of the residuals is homogeneous. This indicates that our data might suffer from heteroskedasticity. However, the tests for homoskedasticity are known to be highly sensitive and must be combined with visual inspections such as of the rvf plot.

As the rvf plot shows no pattern when the residuals are plotted against the fitted values, we conclude that the data meets the assumption of homoscedasticity.

Multicolliniarity

In order to test for multicolliniarity, the variance inflation factors (VIF) for the independent variables are calculated.

In the linear regression model, the variance inflation factors for the independent variables (estat vif) outlines the centered and uncentered variance inflation factors (VIFs) for each independent variable (Stata.com, 2013).

As can be seen in "8.6 Log files", the VIF for all independent variables are well below the conventional cut off point of 10. Thus, we conclude that the data does not suffer from multicolliniarity.

Independent error term

In order to assess whether the error term of the model is independent or not, we carry out a Durbin Watson test for serial correlation of the error term, which is a common issue when working with time series. A visual inspection is carried out by examining a scatter plot of the residuals over time

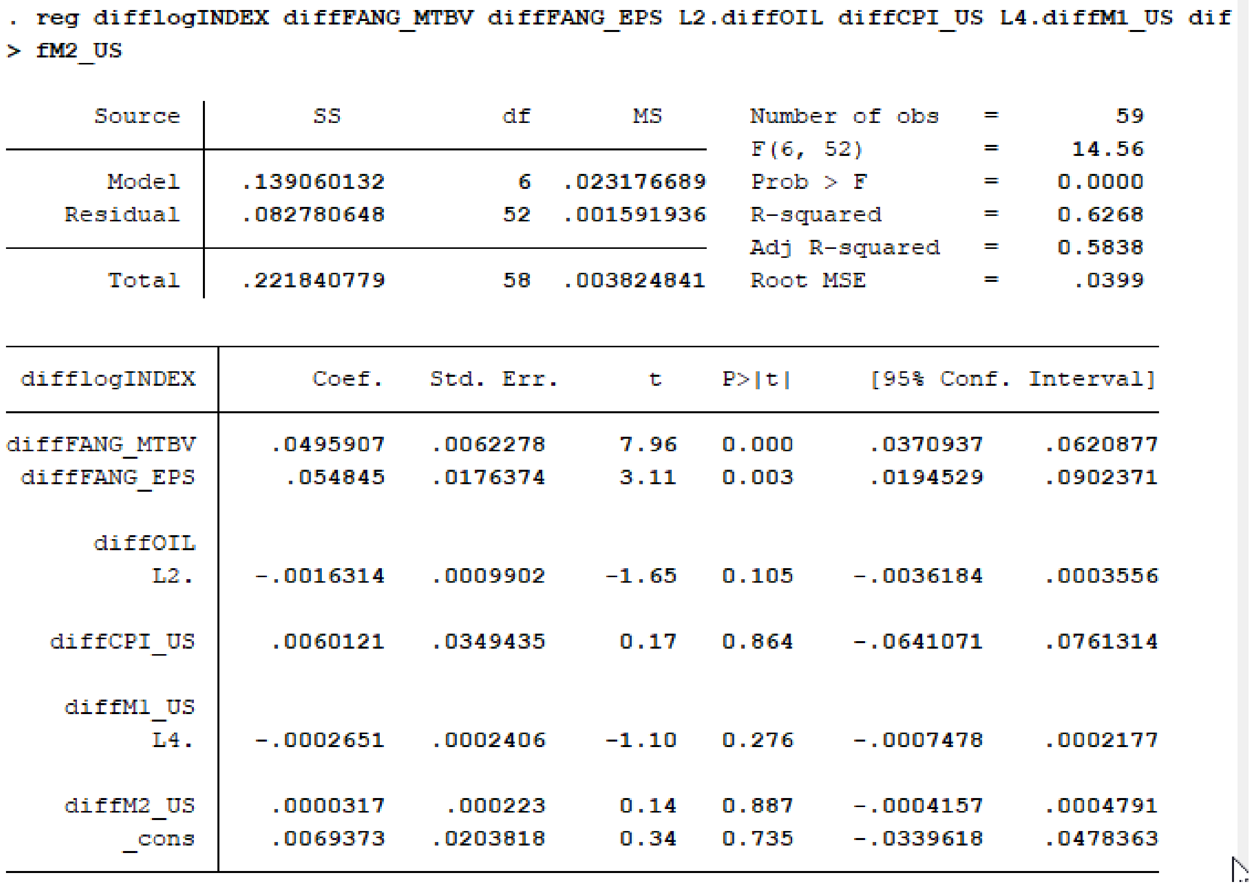
The Durbin-Watson d statistic represents a statistical technique that assesses whether the first-order serial correlation exists among the independent variables (Durbin & Watson, 1950). In other words, it tests if there is disturbance when the regressors are exogenous (Stata.com, 2013).

The detailed results can be found in the “8.6 Log files” and in appendix “8.1 FANG model – all variables”.

The Durbin Watson test value which is relatively close to 2 is a clear indication of the fact that no or very limited serial correlation exists. Furthermore, the scatter plot does not reveal any indication of the error term being correlated over time.

4.2.4 Analyzing model results

Based on the previous sections, it has been documented that the assumptions for OLS are met, which means that OLS is the best linear unbiased estimator for our model. Thus, it is now possible to run the model, as specified in H_1, and interpret the results.



```
. reg difflogINDEX diffFANG_MTBV diffFANG_EPS L2.diffOIL diffCPI_US L4.diffM1_US dif
> fM2_US

      Source |       SS           df       MS      Number of obs   =        59
-------------+----------------------------------   F(6, 52)        =     14.56
       Model |  .139060132         6  .023176689   Prob > F        =    0.0000
    Residual |  .082780648        52  .001591936   R-squared       =    0.6268
-------------+----------------------------------   Adj R-squared   =    0.5838
       Total |  .221840779        58  .003824841   Root MSE        =     .0399

------------------------------------------------------------------------------
 difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
diffFANG_MTBV |   .0495907   .0062278     7.96   0.000     .0370937    .0620877
 diffFANG_EPS |    .054845   .0176374     3.11   0.003     .0194529    .0902371
              |
      diffOIL |
          L2. |  -.0016314   .0009902    -1.65   0.105    -.0036184    .0003556
              |
   diffCPI_US |   .0060121   .0349435     0.17   0.864    -.0641071    .0761314
              |
    diffM1_US |
          L4. |  -.0002651   .0002406    -1.10   0.276    -.0007478    .0002177
              |
    diffM2_US |   .0000317    .000223     0.14   0.887    -.0004157    .0004791
        _cons |   .0069373   .0203818     0.34   0.735    -.0339618    .0478363
------------------------------------------------------------------------------
```

With an adjusted R-squared of 0.5838 and a p-value of 0.0000 for the F-test, the model has a good explanatory power, and it is clear that the independent variables are jointly different from zero. When evaluating each independent variable, it is clear that the two variables, diffFANG_MTBV and diffFANG_EPS, are significant. The two-period lagged variable of diffOIL is close to being significant at a 10% level. In order to specify the final model, the independent variable with the highest p-value is removed and the regression is run again. This is repeated until the variables that are left are all significant at either a 10% or 5% level.

4.3 Building the final model

In the following section, the final model will be specified. As for the initial model, unusual and influential observations will be examined and the assumptions of OLS will be tested for the new model. As this paper has previously applied and explained all the tests that will be used in the subsequent section, the tests will be applied, and the results interpreted without further elaboration.

4.3.1 Model formulation and tests

Based on the findings in the initial model, the hypothesis is that the following model will be significant:

H_2: $difflogINDEX = \beta_0 + \beta_1 diffFANG_{MTBV} + \beta_2 diffFANG_{EPS} + \varepsilon$

4.3.1.1 Unusual and influential observations

The results of the test can be found in “8.6 Log files” and “8.2 FANG model - final”.

While the test finds multiple unusual observations, it is clear that the observations with the following dates are remarkable: 2014-01, 2016-01 and 2017-01. Despite going through the data of the observation, no irregularities have been found. The three observations that have been removed are selected based on the deviation from normal data, as well as the impact they have on the estimation of the regression coefficients. More observations could have been deleted if only considering the tests and graphical displays. However, it is important to keep in mind that if the observations do not contain errors, the information they provide is relevant for the population. Furthermore, the limited size of the dataset must be kept in mind, and it is decided that a maximum of 5% of the data will be deleted under normal conditions.

4.3.1.2 Cointegration

From the work with the initial model, it is known that stationarity was achieved. The test of cointegration was however not entirely clear as Johansen’s test of cointegration indicated that one cointegrated equation existed. To get a clearer picture of whether cointegration exists, the same test is now carried out on the final model. The detailed result can be seen in “8.6 Log files”. The Johansen’s test of cointegration finds that zero equations of cointegration exists. Hence, we can conclude that cointegration does not exist for the variables in the final model.

4.3.1.3 Testing assumptions

Normality of residuals

As can be seen in “8.6 Log files”, the sktest of normality fails to reject that the residuals are normally distributed, while Shapiro-Wilk and Shapiro-Francia both reject that the residuals are normally distributed. When consulting the kernel density estimation plot, the histogram and the standardized normal probability plot and quantiles of residuals against quantiles of normal distribution plot, it can be seen that the residuals resembles a normal distribution. However, with minor violations.

As the visual displays indicate that normality exists with only minor violations, and one of the formal tests indicates that the residuals are normally distributed, it is concluded that the assumption of normality is met.

Linearity

Based on the visual interpretation of the graphs in "8.2 FANG model - final", no non-linear relationship can be seen. Thus, we conclude that the assumption of linearity is met.

Homoscedasticity

Based on both Breusch-Pagan's test and White's test, the residuals are found to fulfill the requirement of constant or near constant variance as can be seen in "8.6 Log files". When consulting the rvf plot, the error term does not display signs of heteroskedasticity. Thus, it is concluded that the assumptions of homoscedasticity is met.

Multicolliniarity

As the VIF is well below the conventional threshold level of 10, there is no cause for concern in regard of multicolliniarity. See ("8.6 Log files").

Independent error term

As can be seen from appendix "8.2 FANG model - final", the visual inspection of the graph matrix indicates independence between the error terms and the other variables. Furthermore, when examining the residuals over time in a scatter plot, and when carrying out the Durbin Watson test, "8.6 Log files", there is support of independent error terms, and the condition is met.

4.3.1.4 Omitted variable bias

As our final model is limited to two independent variables, we proceed and test for omitted variable bias. Omitting relevant variables will lead to a misspecification error. When a model specification error occurs, it can have a substantial impact on the estimates of the regression coefficients (Newbold, 2017).

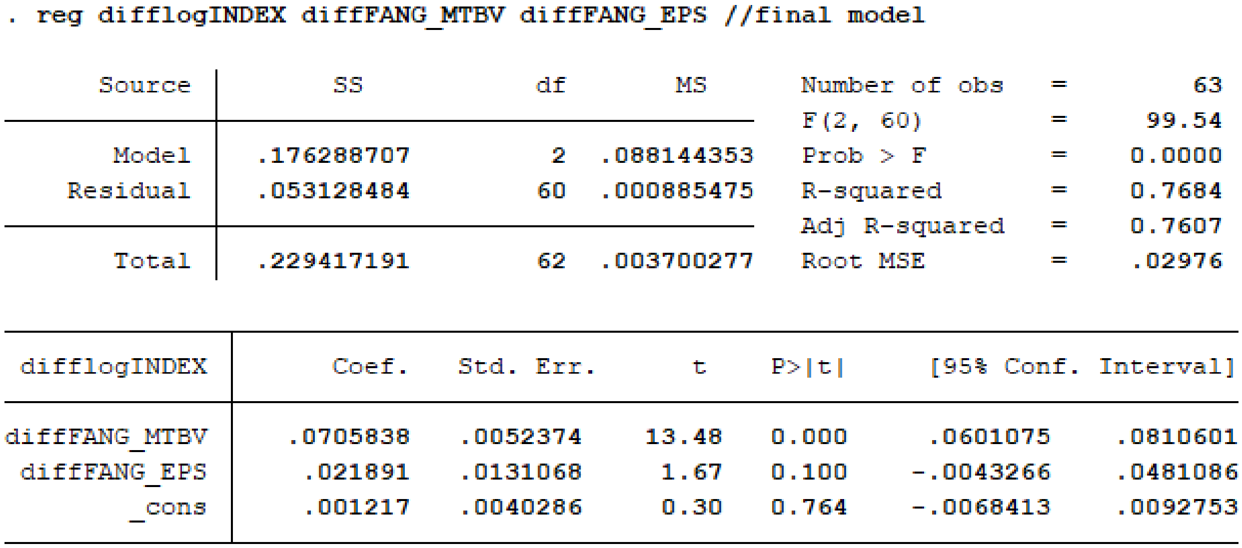
In order to assess whether our model is likely to suffer from a misspecification due to omitted variables, we carry out a specification link test and Ramsey’s RESET test.

The specification link test assumes that there are no additional independent variables that are significant when the regression is specified and the Regression Specification Error Test (RESET) indicates whether the model is linear in the original variables, by assuming that the remaining independent variables are exogenous and have a linear relationship with the independent variable.

Based on both tests, there are clear indications that our model suffers from omitted variable bias.

4.3.2 Analyzing model results

As all the assumptions for applying OLS are met, we estimate H_2 and find as follows:



```
. reg difflogINDEX diffFANG_MTBV diffFANG_EPS //final model

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(2, 60)        =     99.54
       Model |  .176288707         2  .088144353   Prob > F        =    0.0000
    Residual |  .053128484        60  .000885475   R-squared       =    0.7684
-------------+----------------------------------   Adj R-squared   =    0.7607
       Total |  .229417191        62  .003700277   Root MSE        =    .02976

------------------------------------------------------------------------------
 difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
diffFANG_MTBV |   .0705838   .0052374    13.48   0.000     .0601075    .0810601
 diffFANG_EPS |    .021891   .0131068     1.67   0.100    -.0043266    .0481086
        _cons |    .001217   .0040286     0.30   0.764    -.0068413    .0092753
------------------------------------------------------------------------------
```

The adjusted R-squared has increased, compared to H_1, as a result of deleting the outliers. With an adjusted R-squared value of 0.7607 and a p-value from the F-test of 0.0000, the variables are found to be jointly significantly different from 0. By examining the P-value of both diffFANG_MTBV and diffFANG_EPS, we learn that diffFANG_MTBV is significant both at a 5% and a 1% significance level, while diffFANG_EPS is only significant at 10% level. The signs of both coefficients are of the expected sign i.e. if the market to book value increases, the price increases and if earnings per share increases the price also increases.

From the regression coefficients, we find that an increase in market-to-book-value (diffFANG_MTBV) by one unit results in a 7,3%[17] increase in the index price (INDEX) while an increase by one USD in earnings per share results in a 2,2%[18] increase in the index price (INDEX).

By standardizing the coefficients, we are able to evaluate whether market-to-book-value (diffFANG_MTBV) or earnings per share (diffFANG_EPS) has a greater impact on the index price (INDEX). The regression including standardized coefficients can be seen in the appendix "8.2 FANG model - final". It is found that market-to-book-value's impact on the index price is considerably higher than that of earnings per share. When market-to-book value increase by one standard deviation, the log index price increases by 0.85 standard deviations, while a one standard deviation increase in earnings per share only increases the log index price by 0.11 standard deviations.

The estimated model of H_2 looks as follows:

H_2: $difflogINDEX = 0,0012 + \ 0,0706 * diffFANG_{MTBV} + \ 0,0218 * diffFANG_{EPS} + \varepsilon$

4.4 Testing model on US sample

As outlined in our problem statement, the model developed for the FANG stocks will be applied and tested for a sample containing the biggest internet companies leading up to the dotcom bubble. In the following section, the application and testing will take place for the US sample. The testing of the OLS assumptions as well as the examination of the data for unusual and influential observations will be carried out as vigorously as in the previous sections. The full results are available in the appendix "8.3 The US model" and in the "8.6 Log files". The written section will however focus on the main findings during the data examination and testing rather than elaborating on each individual step.

---

[17] $\%\Delta y = 100 * \left(e^{\beta} - 1\right) = 100 * (e^{0,0705838} - 1) = 7,31$

[18] $\%\Delta y = 100 * \left(e^{\beta} - 1\right) = 100 * (e^{0,021891} - 1) = 2,21$

The analysis of the model results will be the main focus of this section as it is an essential part when discussion the application of the model developed for the FANG stocks on the US sample.

By applying the model, we developed for the FANG stocks on the US sample, we will test the following hypothesis:

H_3: $difflogINDEX\_US = \beta_0 + \beta_1 diffUS\_MTBV + \beta_2 diffUS\_EPS + \varepsilon$

4.4.1 Unusual and influential observations

Based on the various diagnostic tools, it is found that the observations with date 1999-02, 1999-06 and 1999-08 are outliers and that they also have a significant impact on the estimation of the coefficients. Thus, they are removed.

4.4.2 Stationarity

The differenced variables are all found to be stationary.

4.4.3 Testing assumptions

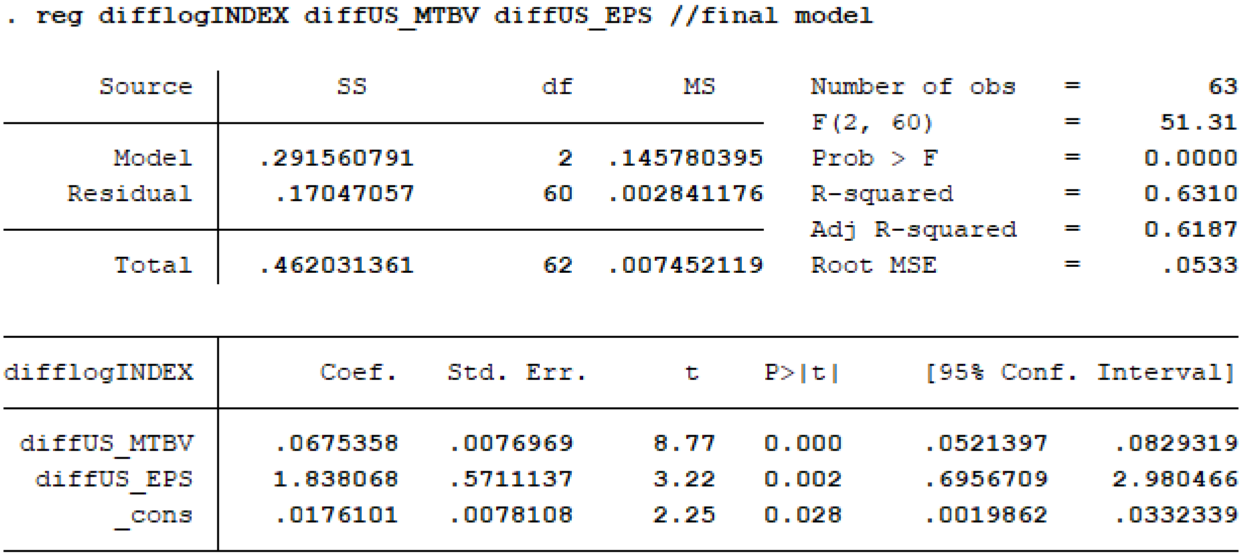
Based on both formal tests and visual inspections, it can be concluded that the data meets the assumptions for linearity, homoscedasticity, normality of residuals, multicolliniarity and independence of the error term. When the residuals are plotted against difflogINDEX, there might seem to be indications of a linear relationship. However, both formal tests and the plot of the error term over time indicate that the assumption of independent error terms is met.

4.4.4 Omitted variable bias

Based on both tests for omitted variable bias, no omitted variable bias exists.

4.4.5 Analyzing model results

As all the assumptions for applying OLS are met, we estimate H_3 and find as follows:



```
. reg difflogINDEX diffUS_MTBV diffUS_EPS //final model

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(2, 60)        =     51.31
       Model |  .291560791         2  .145780395   Prob > F        =    0.0000
    Residual |   .17047057        60  .002841176   R-squared       =    0.6310
-------------+----------------------------------   Adj R-squared   =    0.6187
       Total |  .462031361        62  .007452119   Root MSE        =     .0533

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 diffUS_MTBV |   .0675358   .0076969     8.77   0.000     .0521397    .0829319
  diffUS_EPS |   1.838068   .5711137     3.22   0.002     .6956709    2.980466
       _cons |   .0176101   .0078108     2.25   0.028     .0019862    .0332339
------------------------------------------------------------------------------
```

With an R-squared value of 0,6187 and a P-value for the F-test of 0,0000, we conclude that the coefficients are jointly significantly different from zero and that the regression has significant explanatory power. The P-values for both diffUS_MTBV and diffUS_EPS indicate that both measures are significant at a 1% significance level. As expected, the sign of diffUS_MTBV is positive i.e. when the market capitalization increases relative to the book value, the index price (difflogINDEX) increases. The magnitude indicates that for each unit increase in the market to book ratio, the index price increases by 7,0%[19]. The sign of the coefficient for diffUS_EPS is also positive as expected. The positive sign indicates that an increase in earnings per share will lead to an increase in index price. This is in line with what we would expect to observe in the real world. Furthermore, the magnitude of the coefficient indicates that for each one unit increase in the earnings per share, the index price will increase by 528,4%[20].

The estimation of the standardized coefficients can be found in appendix "8.3 The US model". From the standardized coefficients, we find that the impact of market to book value, diffUS_MTBV, on the index price, difflogINDEX, is considerably greater than the impact of earnings per share, diffUS_EPS. From the interpretation of the standardized coefficients, it is seen that if diffUS_MTBV increases by one standard deviation, difflogINDEX increase by 0,70

---

[19] $\%\Delta y = 100 * \left(e^{\beta} - 1\right) = 100 * (e^{0,0675358} - 1) = 6,987$

[20] $\%\Delta y = 100 * \left(e^{\beta} - 1\right) = 100 * (e^{1,838068} - 1) = 528,44$

standard deviations and a one standard deviation increase in diffUS_EPS will cause a 0,26 standard deviation decrease in difflogINDEX.

The estimated model of H_3 looks as follows:

H_3: $difflogINDEX = 0{,}0176 + 0{,}0675 * diffFANG_{MTBV} - 1{,}8381 * diffFANG_{EPS} + \varepsilon$

4.5 Testing model on CH sample

As for the US sample, the test of assumptions will only be reported in brief for the CH sample. However, the full analysis has been carried out, and the data can be found in “8.6 Log files” and "8.3 The US model”.

By applying the model we developed for the FANG stocks on the CH sample, we will test the following hypothesis:

H_4: $difflogINDEX\_CH = \beta_0 + \beta_1 diffCH\_MTBV + \beta_2 diffCH\_EPS + \varepsilon$

4.5.1 Unusual and influential observations

Based on the various diagnostic tools, it is found that the observations with date 2013-01, 2014-01 and 2017-04 are outliers and that they also have significant impact on the estimation of the coefficients. Thus, they are removed. Based on the tests and the plots, it seems that other observations exist where one could argue that they resemble outliers. In the decision on how many observations to delete, the limited number of observation in the dataset has been taken into consideration.

4.5.2 Stationarity

The differenced variables are all found to be stationary.

4.5.3 Testing assumptions

Based on both formal tests and visual inspections, it can be concluded that the data meets the assumptions of linearity, multicolliniarity and independence of the error term. White’s test for homoscedasticity indicates that our model suffers from heteroskedasticity. However, the rvf plot and the Breusch-Pagan test do not support this concern.

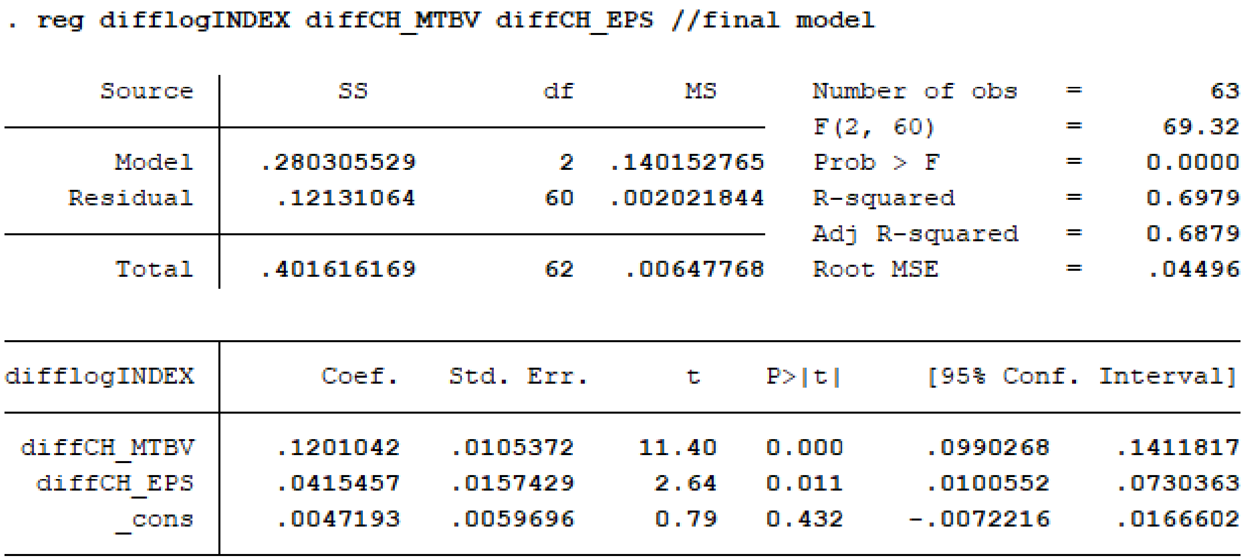
The assumption of normality of residuals is not met. This is seen both by the formal tests and by the visual inspections. The fact that the model for H_4 does not meet the assumption of normality of residuals implies that the calculation of P-values might be affected. However, it does not contribute to bias nor does it cause inefficiency of the regression model (Newbold, 2017).

4.5.4 Omitted variable bias

Based on both tests for omitted variable bias, the model suffers from omitted variable bias.

4.5.5 Analyzing model results

As all the assumptions for applying OLS are met, we estimate H_4 and find as follows:



```
. reg difflogINDEX diffCH_MTBV diffCH_EPS //final model

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(2, 60)        =     69.32
       Model |  .280305529         2  .140152765   Prob > F        =    0.0000
    Residual |   .12131064        60  .002021844   R-squared       =    0.6979
-------------+----------------------------------   Adj R-squared   =    0.6879
       Total |  .401616169        62   .00647768   Root MSE        =    .04496

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 diffCH_MTBV |   .1201042   .0105372    11.40   0.000     .0990268    .1411817
  diffCH_EPS |   .0415457   .0157429     2.64   0.011     .0100552    .0730363
       _cons |   .0047193   .0059696     0.79   0.432    -.0072216    .0166602
------------------------------------------------------------------------------
```

As the P-value of the F-test is 0,0000, we conclude that the coefficients of the independent variables are found to be jointly significantly different from zero. Furthermore, an adjusted R-squared of 0,6879 indicates that the model has a high explanatory power. diffCH_MTBV is found to be statistically significant at a 1% level, while diffCH_EPS is significant at a 5% level. The sign of diffCH_MTBV is as expected and the magnitude of 0,1201 indicates that for each one unit increase in market to book value the index price increases by 12,8%[21]. The sign of the coefficient for diffCH_EPS is also as expected. The magnitude of the coefficient of diffCH_EPS

---

[21] $\%\Delta y = 100 * \left(e^{\beta} - 1\right) = 100 * (e^{0,1201042} - 1) = 12,76$

indicates that for each one USD increase in the earnings per share, the index price increase by 4,24%[22].

The estimation of the standardized coefficients can be found in appendix “8.4 The CH model”. From the standardized coefficients, we find that the impact on difflogINDEX is considerably higher from diffCH_MTBV than from diffCH_EPS. For each increase in diffCH_MTBV by one standard deviation, the difflogINDEX increases by 0,8090 standard deviation while a one standard deviation increases in diffCH_EPS will lead to an increase in difflogINDEX by 0,1873 standard deviation.

The estimated model of H_4 looks as follows:

H_2: $difflogINDEX = 0{,}0047 + \ 0{,}1201 * diffFANG_{MTBV} - \ 0{,}0415 * diffFANG_{EPS} + \varepsilon$

4.6 Coefficient comparison across models

In order to facilitate the discussion (see section “5. Discussion”) where the model’s application on the samples for the FANGS, the US and CH will be discussed, the following section will provide supporting calculations.

In the following section, the results that have already been computed and reported in the previous section will be compared by displaying the relevant measures together. The most significant and interesting findings will be further discussed in chapter "5. Discussion”.

As can be seen from *Table 4*, the adjusted R2 indicates that the variance in INDEX is well explained by the model in all of the three samples.

---

[22] $\%\Delta y = 100 * \left(e^{\beta} - 1\right) = 100 * (e^{0{,}0415457} - 1) = 4{,}24$

Table 4 – Variance, Source: Personal production

Overall model comparison			
	FANG	*US*	*CH*
Adjusted R^2	0,7607	0,6187	0,6879
F-test	0,0000	0,0000	0,0000

From *Table 5*, it is clear that the variable MTBV is significant for all three samples. While the magnitude of the coefficient is quite similar for the FANG and the US sample, the magnitude of the coefficient in the Chinese sample is considerably higher. This is further supported by the Z-test[23] which fails to reject that the magnitude of FANG and the US are different, while it rejects that the magnitude of the models for the FANG and the CH sample are similar.

Table 5 - Coefficients related to MTBV testing, Source: Personal production

Coefficient analysis – MTBV			
	FANG	*US*	*CH*
P-value of t-test	0,000	0,000	0,000
Standard error	0,0052	0,0077	0,0105
Magnitude	0,0706	0,0675	0,1201

Based on *Table 6*, it can be seen that EPS is only significant at a 10% level in the FANG sample, while it is significant at a 5% level in both the US and the CH sample. At first sight, the coefficient of EPS for the US sample might look considerably off compared to the coefficient of

---

[23] The computation and an overview of the actual Z-scores can be found in " 8.5 Comparison of models". The use of z-scores for comparison of coefficients are discussed in Paternoster et al, 1998 and Clogg et al, 1995.

the two other samples. However, when evaluating the coefficient, it is relevant to consider how much an increase in EPS by one USD represents relative to the total EPS for each sample. As a simple illustration of this, the percentage increase that a one USD increase in the weighted average of EPS, for the last observation in each sample, is calculated. Based on the calculations, it is clear that a one USD increase in EPS in the US sample represents an increase of 149% while it for the FANG sample only represents a 10% increase . Needless to say, the magnitude of the coefficient, which represents a percentage change in INDEX, must be expected to be different across the samples.

Table 6 - Coefficients related to EPS testing, Source: Personal production

Coefficient analysis – EPS			
	FANG	*US*	*CH*
P-value of t-test	0,100	0,002	0,011
Standard error	0,0131	0,5711	0,0157
Magnitude	0,0219	1,8380	0,0415

5. Discussion

The price development of the FANGs and the Chinese equivalents differs from that of the pre-dotcom companies

In the initial part of the analysis, “4.1 Index creation”, three price indexes were developed. One for the FANGs, one for the leading Chinese Internet stocks and one for the leading pre-dotcom stocks in the US. While the purpose of developing these price indexes were to generate the dependent variables that were to be used when developing the econometric model, the generated price indexes proved to be interesting on their own. The price development of the indexes was plotted over time in order to facilitate a comparison of the development. The plots can be seen on *Figure 9*.

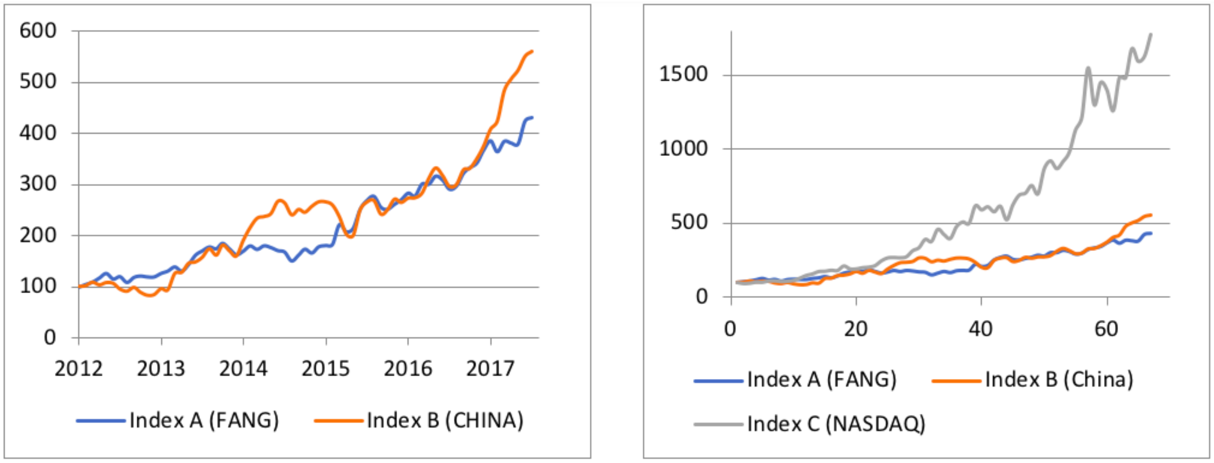


Figure 11 - (A) Plot of Index A and B; (B) Plot of Index A, B and C combined

Source: Personal production, data: Datastream, Bloomberg

From the plots, some very destictive patterns can be seen. For the majority of the time period, May 2012 – December 2017, the FANG stocks and their Chinese equivalents follow each other very closely. The two indexes do deviate from each other, but after a period they return to follow each other very closely. One possible explanation for this deviation could be associated with the fact that the overall Chinese economic outlook has steadily accelerated since 2013, thanks to a robust expansion of R&D expenditures that rapidly enhanced the technologic manufacturing capacity (National Bureau of Statistics of China, 2014).

In the comparison of all three indexes, the price index of the pre-dotcom stocks displays a trend where it can rightfully be stated that the prices are *skyrocketing* relative to the two other indexes, in particular during the last three years that are included. Based on the plots, it is not possible to determine what exactly causes the difference in the development of the index prices. However, the underlying drivers might be explained by concepts that were discussed in the literature review (see 2.1 Internet stocks; a special kind of tech stocks belonging to the growth stock family).

A main driver of the *skyrocketing* trend of the US pre-dotcom stocks, as well as the very positive development of the FANG stocks and the Chinese stocks, could very well be *investors' expectations*. Keynes (1936) represents the first economist who emphasizes the role of investors' beliefs in stock valuations. According to Keynes, investors are very likely to become too optimistic during boom periods and tend to expect that an increase in growth will continue systematically (Keynes, 1936). Keynes' prediction; that investors are likely to expect growth to continue systematically, seems to be true for both the Chinese stocks and the FANGs at this stage. Based on data obtained on May 6th 2018, the stock price growth forecasts of the FANGs are as follows: Facebook Inc 25-48%, Amazon.com Inc 17-33%, Netflix 8,5-31% and Alphabet 26-43% (CNN Money, 2018). The growth expectations are even higher for the Chinese stocks where analysts expect the stock prices to grow as follows: JD.com 39-63%, Tencent Holding Ltd 37-66%, Alibaba Group Holding Ltd approximately 37% and Baidu Inc approximately 20% (CNN Money, 2018).

Animal spirit – an influence on Internet stock prices that cannot be ignored

When consulting existing literature, there is broad support for investors being irrational, which in turn cause the market to be inefficient. Keynes (1936), Smidt (1968), Shiller (1980), Buffet (1989) and Samuelson (1998), all point out the lack of market efficiency as a result of irrational investors. As history showed, the stock prices of the pre-dotcom bubble were driven by *animal spirit* as defined in (Keynes, 1936). The theory of animal spirit is primarily based on the concept of *investors' confidence* regarding the securities in the market. Today, many investors and analysts also show great confidence in Internet stocks, and it must be considered if the market in general is overly optimistic, and therefore irrational, when evaluating the performance of Internet stocks. Similarily, to the pre-dotcom period, the perception that Internet companies can continiously provide extraordinary growth because of innovation and new technologies could be a driver of the market today. From history, we know that the skyrocketing price development of the pre-dotcom bubble proved to be driven by animal spirit and while the price development of the FANGs and their Chinese competitors are nowhere as crazy, it is fair to assume that a part of the increase in stock prices is due to animal spirit.

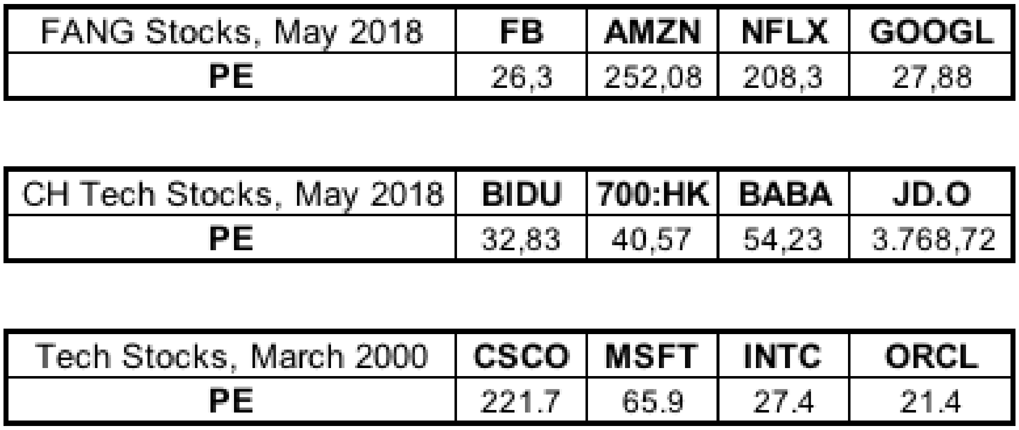
In short, it can be argued that the overall performance of the Internet stocks is to a large extent *mental*, and not necessarily rational. Therefore, investors' beliefs and emotions play a significant role in the investor's decision making, which ultimately affects the price levels of the stock.

Historically, the presence of animal spirit is most clear when the market is either booming or crashing. During boom periods, investors' confidence is irrationally optimistic. In this context, investors' expectations of stock prices are biased and maximized by the *bullish sentiment* (Risager, 2016). The pre-dotcom period represents a good example of bullish sentiment in the market. In the 1990s, the fact that the profit share started declining before the bubble bursted in spring 2000 did not bother the market at all in 1998 and 1999. In fact, investors' confidence and expectations did not shift and the investors continued to be over-optimistic regarding the Internet stocks. The declining profits did not alarm the investors who persisted to believe in the "technology dream" (Risager, 2016)(pp. 42). This scenario is reflected by the trend shown by the Index C on *Figure 9*.

Despite P/E values indicate that today's market resembles the pre-dotcom era there are indications that investors have become less irrational

Several similarities exist between the tecnology market outlook of the the 1990s and the current technology market outlook (Cohan, 2016). in the 1990s, a group of "big name stocks with life-changing technologies" (Shell, 2017) are again outperforming the leading economic indictor (LEI) indexes and other important indexes in the U.S. market. For instance, both the FANGs and the Chinese companies present P/E ratios similar to the top-four dotcom companies. *Table 7* shows the 12 month trailing P/E values for the stocks in Index A, B and C (Thomson Reuters, 2018) (Bloomberg L.P., 2018) (Seeking Alpha, 2015).

Table 7 - Overview of P/E values, Source: Personal production



Index A

FANG Stocks, May 2018	FB	AMZN	NFLX	GOOGL
PE	26,3	252,08	208,3	27,88

Index B

CH Tech Stocks, May 2018	BIDU	700:HK	BABA	JD.O
PE	32,83	40,57	54,23	3.768,72

Index C

Tech Stocks, March 2000	CSCO	MSFT	INTC	ORCL
PE	221.7	65.9	27.4	21.4

Based on the P/E values clear similarities between the pre-dotcom Internet companies and today's Internet companies exists, despite that the index price development indicated otherwise.

However, several differences also exist between the pre-dotcom index and the other two indexes. One of the main differences consists in the fact that today's technology companies are generating greater revenues which mean that today's growth is, to a greater extend, supported by financial facts (Shell, 2017). While companies like Noosh Inc[24] and Pets.com[25] went public declaring no revenues in the 1990s, nowadays it is mandatory to demonstrate to possess at least $100 milion in revenues prior the IPO announcement (Cohan, 2016). Furthermore, the current technology market presents fewer technology categories but an increased number of players compared to the market of the 1990s. The prevailing sectors are *E-Commerce* companies e.g. Alibaba, JD.com and Amazon, *social networks* e.g. Tendenc Holdings and Facebook, *conumers devices* e.g. Apple and *netwrok companies* like Google and Baidu (Cohan, 2016).. By contrast, in the 1990s, the industry was much more fragmented as additional classifications existed, including Web Portals, Web Marketing, Internet venture capitals and Web Consulting (Cohan, 2016). Overall, the modern technology industry tends to differ from the dotcom era to the extent that the market looks more mature and better regulated than it was in the past. In this regard, the dotcom

---

[24] Noosh Inc. is an American company founded in 1998 and provides a marketing platform that produces data-rich reports (Bloomberg L.P., 2018). Now it is known as NewLineNoosh Inc.

[25] Pets.com was a dot-com company that operated in the e-commerce, by selling pet supplies to retailers. It was founded in 1998 and declared default in 2000 (Goldman, 2010).

investors witnessed the rise of the first technology giants, i.e. Microsoft, Cisco, Intel and Oracle, and a rise of start-ups that based their financing policies uniquely on promises of revolutionary innovations often associated with poor supporting financial figures (Sharma, 2017). By contrast, the modern technlogy boom exhibits innovations (e.g. Internet of Things, Artificial Intelligence, Blockchain and machine learning) that are already expanding and difficult to specifically label. What is unique about the modern industry is what Jack Ma, CEO and founder of Alibaba, calls " electricity of the 21st century" (Johnson, 2017). Jack Ma also states that "[The] internet should be treated equally as important as electricity of the last century. And everything will be online, and everything online will have data. And data will be the energy for the innovation." (Johnson, 2017). Therefore, what is characteristic to this technology era is the huge volume of "harness data" (Sharma, 2017) generated on the web.

Moreover, the modern investors might be considered slightly more *consciously irrational* compared to the investors of the 1990s. The term *consciously irrational* is based on the assumption that although irrational, investors may have *moderated* their expectations of limitless growth opportunities as a consequence of the two recent crises. That is, while investors still might be irrational, there might exist more awareness in the market than in the past; for this reason, the market is more consciously irrational. This could be explained by the two recent financial recessions (the dotcom bubble in 2000 and the financial crisis in 2008) which might have taught something to the current investors (Say & Miller, 2016).

Generally speaking, it seems that there might exist greater awareness regarding the technological stock market. Additionally, some very traditional value investors have recently changed their opinions concerning the the Internet stocks. For instance, Warren Buffet has openly admitted to his shareholders at the Berkshire Hathaway 2017 event that not investing in Google and Amazon was a mistake. In this regard, Buffett has reported that "If I was forced to buy Alphabet or short it, I'd buy it; same way with Amazon." (Crippen & Cheng, 2017). The reason why Buffet has avoided investing in Internet stocks is because he did not understand the business model of these technologic companies. In other words, he did not understand how these companies could make profits and how they could sustain their business in the long-run (Setin, 2018). In fact, the only Internet stock that the Berkshire Hathaway portfolio includes is Apple Inc., which represents perhaps the most traditional (manufacturing) business model among the Internet stocks and the FANGs.

However, it also seems that a greater skepticism and criticism exist towards the pricing of technology Internet securities (The Economist, 2017). Two amongst the most valuable lessons learnt from the recent past crises are "sell the hype, don't buy it" and "don't feel big times" (Say & Miller, 2016). In other words, investors should be ready to critically question every optimistic forecasts related to their portfolio. Specifically, an alarm bell should ring inside investors' head anytime technology companies present forecasts based on for instance "user usage" or other non-financial figures (Hoium, 2014).

Looking at the technology market in the 1990s, it is now obvious that the technology stocks in the pre-dotcom era were widely overvalued by traditional financial measures (Risager, 2016). What is also obvious is that the overvaluation persisted in the years prior 2000, specifically from 1997 to 1999. In fact, since the early 1997, investors started buying the technology stocks driven by excessive optimism and euphoria regarding the new technology era (Risager, 2016). Also analysts got caught by this overconfidence; for instance, Glassmann and Hasset (2000) predicted that the Dow-Jones would have jumped from 11,497 to 36,000 in four years, predicting an increase for the Dow-Jones index equal to approximately 214%. In reality, in September 2002 the index reached the bottom at 7,701.45, which represented its lowest value ever reached (CNBC.com, 2007). Generally speaking, the authors can observe an insanely high irrationality in assessing tech stock valuation prior the Dotcom bubble.

Not only the less extreme price development of the FANGs and their Chinese competitors indicate that investors have become more rational. This is further supported by research which find that investors' portfolios are more diversified, both in terms of companies and financial indexes that they invest in. In the 1990s, investors interested in the technology stocks primarily invested in NASDAQ's top players. Now, investors tend to diversify by investing across multiple financial indexes, e.g. Fidelity Nasdaq Composite Index Tracking Stocks (ONEQ), or PowerShares S&P SmallCap Info Tech (PSCT) (Glassman, 2015). Furthermore, today's technology market consist of a wider range of companies than it did in the 1990s. Therefore, the investors' possibilities of diversifying investments within the Internet and technology sector has increased compared to twenty years ago.

The choice of stock exchange can impact which factors that are driving prices

When examining the correlation matrix for the stocks in each index, as displayed in section "4.1 Index creation", the low correlations seen for the CH sample in the period 2015-2017 is striking. The correlation among the stocks in the FANG and the US sample are all very high, with the lowest correlation being of 0.91. As for the CH sample, a correlation as low as 0.52 is observed. When stock prices do not follow each other, an obvious explanation will often be that the nature and the drivers of the companies are different. However, the companies are selected based on being the biggest Internet companies of the region i.e. it would be expected that the stocks are driven by the same factors. The correlation matrixes provide an interesting comparison between the stocks in the CH sample and the FANG stocks.

The business type of the individual FANG stocks and the business type of the most dominant Chinese Internet companies is striking. To a great extent, direct parallels of the services that the FANG companies provide can be drawn to those of the Chinese companies. Tencent is active within the entertainment section and one of its main products is WeChat, which share similarities with Facebook. Furthermore, Tencent offers a video platform that resemble Netflix' service (Tencent 腾讯, 2018). Alibaba and JD.com both engage in the 'new retail' segment and resembles Amazon (JD.com, 2018) (Alibaba.com, 2018). In addition, Alibaba offer their Youku streaming service which resembles Netflix and Amazon's Prime[26]. Baidu's main product is a search engine which is the Chinese equivalent to Google, and it also owns the majority of iQiyi which is a video and TV streaming service (Pham, 2018). Furthermore, Baidu is also running a project on driverless cars which are considered a serious competitor to Alphabet's Waymo project (Bloomberg News, 2017). As the FANGS and the internet stocks of the CH sample are very similar in the business areas in which they operate, differences in the nature of business cannot be considered as an explanation for the difference in homogeneity of the stock prices within the CH sample.

---

[26] Prime contains Amazon's TV and video streaming service, as well as other services.

The strong domestic customer focus of the companies in the CH sample exposes the Chinese companies to less global competition. While the Chinese companies are huge companies on a global scale, the market they focus on is to a great extend domestic rather than global. With a domestic market that is highly regulated by the government and where main competitors such as Facebook, Twitter, Youtube and Google are banned, the companies in the CH sample are exposed to much less competition than companies in the two other samples (Banjo, 2018). This creates an environment with oligarch-like conditions, and the stock price might be less driven by general market conditions due to the role of the government. If it is indeed the case that the Chinese companies are less impacted by broad market conditions, a natural result would be that companies can suddenly do particular well within an area, without fundamental market drivers have changed. In turn, this means that the stock price can suddenly develop positively (or negatively) without the broader market having changed. In effect, this could mean that the correlation of the companies is reduced as a considerable part of the share price is driven by other factors than the general market conditions.

While focusing on domestic customers, the Chinese companies are seeking for investors globally. In order to be in the most attractive financial markets most of the Chinese internet stocks are listed outside China. Both Baidu and JD.com are listed on NASDAQ, while Alibaba is listed on NYSE (The New York Stock Exchange) and Tencent at HKG (Hong Kong Stock Exchange). While NASDAQ and NYSE are the two biggest stock exchanges in the world, HKG is only the 6th largest with a size that is approximately half of the NASDAQ and one sixth of the NYSE (Desjardins, 2017). With the smaller size of HKG, it can be assumed that the market is less efficient. Tencent, which is listed on HKG, can expect to attract less international investors, in particular European and American investors, as investing in foreign markets implies extra fees and both private and professional investors are subject to other regulations than what they are accustomed to. The lower correlation for the CH stocks could be a result of the stocks being listed at different stock exchanges whereby different investors are attracted. For comparison, the companies included in the US sample and the FANG companies are all listed on NASDAQ.

Macro-economic variables are non-significant as a result of a short time period and possibly also of the type of service offered by Internet companies

The approach of this paper is based on previous research which has been combined to form the research design. First, the share-weighted index method is used for creating an index consisting of the stocks of interest. Then, an econometric model is developed where the independent variables consists of both financial accounting factors as well as macro-economic factors. In other terms, the authors have formulated and tested a model that includes both macro-economic and micro-economic variables. The purpose of this section is to outline and discuss the econometric model with a critical emphasis on the differences between this model and past research.

In the development of the econometric model, six independent variables are included: two financial accounting factors i.e. earnings-per-share and market-to-book value, and four macroeconomic factors i.e. money supply (M1 and M2), CPI and the oil price. The purpose was to build a holistic model which could explain the factors that affect Internet stocks. As a result of this, it was decided to include both financial accounting and macro-economic factors in order to develop a model that could be used to find which factors drive the prices of Internet stocks. While the existing literature on the impact of accounting factors and macro-economic factors on stock prices is plentiful, no literature that combines these factors in an effort to forecast stock prices exist, to the best of the authors knowledge. In other terms, starting from the problem statement that sets the goal of the research, a broad econometric model has been developed based on the combination of macro-economic and financial accounting variables. This approach is based on the conviction that both areas could contain variables that have a significant impact on the stock price of the FANG stocks.

In the analysis, none of the macro-economic variables were found to be significant. This finding could be a result of the limited period of time that is analyzed. It is clear that the time period taken into consideration in this paper is quite short compared to what existing research that utilizes macro-economic factors normally use (see “2.3.3 The macro-economic factors”). Most other studies based on macro-economic variables are found to use periods between 10 and 50 years. The short time period used in this paper is dictated by the earliest date where data is available for all four FANG stocks. As the FANG stocks are the main focus of this paper, data must be available for all stocks throughout the period that is examined. The earliest month where all relevant data for all FANG stocks is available is June, 2012. Among the FANGs, Facebook is the latest stock to be publicly listed, thus the IPO of Facebook on May 18th 2012 is a restrictive factor (Nasdaq.com, 2018). As the samples for the Chinese stocks and the pre-dotcom stocks are used for comparison to the findings for the FANG sample, the time duration for these samples are selected to match that of the FANG sample.

The short time period where data is available has an impact both when testing the impact of the accounting factors as well as when testing the impact of the macro-economic factors. However, based on how previous research has used macro-economic factors to explain stock price development, it indicates that the short time period has a greater impact when evaluating the impact of macro-economic factors. All of the related work concerning macro-economic factors that are discussed in “2.3.3 The macro-economic factors” considers the long-term influences of macro-economic variables on a specific set of stock prices. Furthermore, in the academic world there is a general consensus that macro-economic variables are valid predictors of stock prices in the long run. Moreover, the general view is that macro-economic changes do not usually have an effect in the short-run. For this reason, macro-economic indicators i.e. money supply, CPI and oil prices, are typically used to detect change in the outlook of the general economy over a longer period of time.

Taking the above into account, a natural question is whether the chosen macro-economic variables would have been significant if the time period had been longer. As there is no possibility of extending the time period for the research of the FANG stocks, this question can only be answered based on assumptions. In the light of the considerations reported in the literature review (see “2.3.3 The macro-economic factors”), it is found to be likely that the macro-economic variables would not be significant even if a longer time period had been available. Very little literature exists that support long-term relationships between technology stocks and macro-economic variables. A natural explanation for this could very well be that the explanatory power of macro-economic variables mainly applies to more traditional stocks such as e.g. transportation and manufacturing stocks. This explanation is in line with several financial text books that all emphasize the predictive power of macro-variables on stock prices. However, they also point out that this connection is mainly valid over longer time periods and for stocks within e.g. manufacturing and logistics (Risager, 2016) (Yamarone, 2016) (Mankiw, 2005).

From a logical point of view, it also seems to hold true that the stocks in our sample are less affected by the overall economic trends than other traditional stocks. When considering all the stocks, two underlying factors seem to link these companies: *technology* and *service-provision*. When considering the business models of these companies, it is clear that the services offered by these companies are generally accessible and aim to reach a broad customer segment of the market. In general, their products and services are relatively cheap. Saying that they offer cheap services is perhaps too simplistic, but all these companies provide very accessible services which in general allow both high-income people as well as low-income people to get access. As a consequence of the limited costs of the services, it can be claimed that the general wealth of the economy, which is measured by the major macro-economic indicators (Yamarone, 2016), is perhaps not as significant in the context of Internet stocks as it is for the more traditional stocks.

From a customer and user point of view, the usage of the services that are offered might not be significantly impacted by the general economy. However, other factors might cause the general well-being of the economy to impact some of the Internet companies. While users might not stop using Facebook or stop streaming content from Amazon Prime and iQiyi, the profitability of the Internet companies might still be impacted by macro-economic factors to some extent. The main revenue driver for Facebook and Google is income generated from advertising which is an area where companies might cut their spending during challenging times (Alphabet 2017 yearly statement). Furthermore, Amazon and Alibaba’s main business is to be an online market place where manufacturers can sell their products. In other words, if manufacturers’ sales are negatively impacted by the condition of the general economy, it must lead to lower sales activity through the online platforms as well. In summary, there is support that the Internet stocks are less likely to be impacted by macro-economic variables. However, a relation cannot be completely ignored as some of the Internet stocks still rely on income streams that could have a clear link to the general wellbeing of the economy.

When researchers seek for *long-term* relationships between stock prices in a specific market and a set of different *macroeconomic* factors cointegration is normally found to exist between the variables. As discussed in the section “2.5 Statistical analysis of stock performance”, the concept of *cointegration* refers the presence of a linear combination between a set of non-stationary time series (Granger, 1981). The cointegration condition is measured by two tests: Engle-Granger and Johansen cointegration tests (Hobdari, 2016). The lack of cointegration for the variables examined in this paper conflicts with the findings of previous research. In fact, the majority of the studies that explore the relationships between security prices and macro-economic factors, present cointegration. For instance, Mukherjee and Naka (1995) assess and discuss the cointegration between the Japanese stock market and macro-economic variables such as inflation, money supply and industrial production. Similarly, Nasseh and Strauss (2000) find that a high degree of cointegration exists between stock prices and the economic activity in the European sample they have decided to select. Likewise, Al-Majali and Al-Assaf (2014) examine the cointegration of variables between the Amman Stock Exchange and a group of macroeconomic factors, such as gross domestic product (GDP) and consumer price index (CPI). Since this research does not find cointegration in the analysis, it could be concluded that there exists a conceptual conflict between the existing literature deliberated in chapter “2.5 Statistical analysis of stock performance” and this paper. One of the potential explanatory reasons could be linked to the fact that the timeframe of this work is too short.

Internet stock prices are driven by expectations of future earnings rather than current ones

In opposition to the macro-economic variables, both of the accounting variables are found to be significant as can be seen in “4.6 Coefficient comparison across models”. However, as previously discussed, the findings for EPS are considered to be unstable due to outliers. The short time period for our samples has a negative impact on the reliability of our results. However, as already discussed this restriction could not be mitigated. Another factor that might have impacted the results is the nature of the stocks. This will be discussed in the subsequent section.

From a pure theoretical point of view, the technology Internet stocks are classified as growth stocks (see “2.1.3 Internet Growth Stocks”). As it is stated the literature review, technology Internet stocks do share the same characteristics as the growth stocks e.g. high market value and great growth expectations (Hand, 2001). In light of this fact, the authors believe that although they belong to the growth stock category, technology Internet stocks present features that belong purely to their own exceptional nature. In the authors' opinions, growth expectations affecting the stock values of technology Internet stocks highly exceed the actual expectations on earnings. In other terms, the investors are so positive and optimistic about the growth prospectuses of these technology companies that the fact that the actual earnings-per-share for Facebook Inc., Amazon Inc. and Netflix Inc. are low and in rare cases negative does not affect the positive investors' confidence over these stocks[27]. And at the end of the day, what primarily affects the prices of the FANGs (and also CH and US stocks) are the investors' expectations of future growth. Therefore, the authors conclude that the current earnings do not necessarily have the same weight as they would have if the technology Internet stocks were *traditional* growth stocks.

Expectations of high future earnings drive the price of innovative and disruptive companies

As could be seen from the analysis, the significance of the EPS variable for the model of the FANG sample is very different from the significance observed in the models of the US and CH sample. While EPS is only significant at a 10% level in the FANG model, it is significant at a 5% level in the two other samples. During the analysis of the data, the significance and the coefficient of EPS were found to be greatly impacted by outliers across all models. During the modeling process both significance and magnitude of the EPS variable were changing dramatically based on the exclusion of just a few observations. Based on the sensitivity to the inclusion or removal of individual observations, the predictions of the significance and the coefficient of the EPS variable must be considered unstable, and based on our samples, we cannot conclude that EPS is a significant factor in determining the INDEX price.

The EPS variable was included in the model, as it has been found to have a significant impact on stock prices. There exists broad literature that have documented the relationship between earnings-per-share and the stock price Modigliani and Miller (1958), Foster (1973), Kormendi and Lipe (1987), Chang et al. (2008) McAnally et al. (1997). Preinreich (1932) argued early on that the strongest fundamental factor associated to growth, along with market-to-book value, is

---

[27] Unlike the other FANGs,Alphabet Inc. reports extremely high earnings-per-share values over the whole period of the analysis.

earnings-per-share (EPS). As previously discussed, the model of this paper found EPS to be significant but when considering the instability of the coefficient and the significance, the paper cannot claim to support the finding of past literature.

The unclear pattern concerning the significance of EPS might be partly attributed to the nature of the companies that we are examining. As the companies in each index represent some of the most innovative and disruptive companies of their time, investors can be expected to be driven by the expectations of future growth rather than current earnings. In the early phase of internet companies' existence, it is not uncommon that current results are negative while a positive stock price development is taking place. This is possible as the current negative results are more than outweighed by the potential to achieve high and lasting future earnings opportunities, which is what investors expect from growth stocks (Hillier, et al., 2013). The fact that the stock price of innovative and disruptive growth companies is often under greater influence from future expected earnings, rather than from current actual earnings, can be a part of the explanation of why our model did not find a stable relationship between the INDEX price and EPS.

Difference in market conditions lead to a different effect of MTBV on stock prices

The market-to-book ratio consists of two components: The market cap and the book value. The market cap is calculated as the price multiplied by the outstanding number of shares. The price, and thereby the market cap, is an outcome of both historical information and expectations for future earnings and cash flows. On the other hand, the book value only contains historical information that is recorded on the balance sheet i.e. there is no impact from future expectations. As the coefficient of MTBV for the CH sample exceed that of the FANG and the US sample, it means that a change in the market-to-book value has a greater impact on the price for the Chinese companies than for the American companies.

Studies of American growth stocks have showed that some of the most important drivers are the expectations of future earnings and future cash flow (Lintner, 1965) (Kaplan & Ruback, 1995). Based on the above, one could develop the hypothesis that as the Chinese stock market is less developed, there is a higher focus on other metrics than on future earnings and cash flows. However, this hypothesis does not hold as most of the stocks from the CH sample are in fact listed in the US. As the stocks are listed in the US, it can rightfully be assumed that they are faced by the same potential investors, and that these investors will use similar metrics when they analyze stocks unless they perceive the market where the companies operate as considerably different.

The Chinese market that Internet companies operate in can be seen as considerably different from the European and the American market. By imposing bans on specific companies and imposing very strict requirements on how the businesses can operate in China, the Chinese government has created an Oligarchic business environment. In the Oligarchic business environment, a few domestic companies thrive, and their foreign competitors such as Facebook, Twitter, YouTube and Google are kept out (Bloomberg News, 2017). The lower competition could lead investors to attribute greater power to the market-to-book ratio when pricing stocks as it could be used to gauge whether the big Chinese companies have a proven track record of working well under the Chinese regulations.

MTBV is found to be significant for all our samples which is in line with the existing literature (see “2.3.2.2 Price-Earnings (P/E) ratio and market-to-book (P/B) ratio”).

6. Conclusion

The following section will summarize the main findings of the paper as well as point out the practical implications that the findings might have for business. Finally, suggestions for future research will be provided.

6.1 Summary of Main Findings

The purpose of this paper was to find macro-economic and accounting factors that impact the stock prices of the FANGs, which represent today’s leading Internet companies. The generalizability of the findings was then to be tested on a sample consisting of the leading Chinese Internet stocks of today as well as a sample of the leading Internet companies of the pre-dotcom bubble era.

First, a price index and the weights of each share were established by using the share-weighted index method for each of the three samples. Subsequently, an initial econometric model was formulated based on findings of existing literature. The initial model was then tested on the sample containing the main object of interest, the FANGs. Based on the findings from this test, a final model was developed and tested on all three samples.

One of the main findings when testing the initial model was that none of the macro-economic factors were significant. The non-significance of the macro-economic variables is mainly attributed to the fact that the time duration of the analysis is well below what is normally used when testing the impact of macro-economic variables. Another main finding was that both market-to-book-value and earnings-per-share were found to be significant. However, both the significance and the magnitude of the coefficient of earnings-per-share were found to be very unstable and depended on the inclusion or exclusion of specific observations. Thus, the only factor that was documented to have a significant impact on the stock price of the leading Internet companies was the market-to-book value.

When applying the model on the sample for the Chinese stocks and the pre-dotcom stocks, the findings are similar to those for the FANG sample. This indicates that the model is generalizable across current stocks with different origin as well as across time periods. The higher coefficient of the market-to-book value that is found for the Chinese stocks, relative to the two other samples, might be a result of the Chinese market being considerably different due to the role that the state plays by setting up strict regulation which limits competition.

During the development and testing of the econometric model, other interesting insights also occurred. When only consulting the price development, the situation of the leading pre-dotcom companies and the FANGs differ significantly. This is in line with the belief that although animal spirit is stills a big player in driving Internet stock prices, investors have learned from the past and are now *consciously irrational*. However, other indicators such as the price-earnings ratio do blur this picture.

Based on the correlation among stocks, another interesting finding is that the stock price development is impacted by which stock exchange a stock is listed on.

As, to the best of the authors' knowledge, no literature exists on which macro-economic and accounting factors that are impacting Internet stock prices of today. Therefore, the research of this paper is an important contribution to a niche area.

As the area that has been investigated does not contain any existing literature, the results are compared to existing literature within the general area of factors that impact stock prices. The findings that Internet stock prices are impacted by accounting variables are in line with previous research, while the findings of macro-economic variables being non-significant conflicts with the existing research. Furthermore, it is not in line with current research that no cointegration is found among the variables.

6.2 Business Implications

Some of the findings of this paper may have direct business and managerial implications and are worth to be aware of both as an investor and as a manager.

The fact that the market-to-book value has a significant impact on the stock prices in all samples indicates that this is a factor that should not be ignored neither by investors nor managers. The importance attributed to the market-to-book value might however differ across markets as indicated by the higher coefficient of market-to-book value for the Chinese companies.

Another important finding of this paper is the indication that the stock price development might be impacted by the choice of exchange to list at. This has broad support in the existing literature (Edelman et al, 1990)(Biddle et al, 1991)(Doidge et al, 2004).

6.3 Future Research

During the work on this paper, several other interesting aspects that could be researched has surfaced. Furthermore, there has been a clear learning process. The intent of this section is to share some of the learning points and the areas that could be interesting to further develop. This is done to support further research regarding the topic.

What drove the curiosity and emergence of this thesis was the interest in the FANGs. Focusing and testing a model on just four stocks served the purpose of this thesis, but it would be interesting to see a similar model development for a broader segment of Internet stocks or for technology stocks in general. If research is carried out on a broader segment, it might allow substitution of stocks during the time period which would make it possible for the researcher to examine a considerably longer time period than that of this study. By investigating a longer time period one of the most considerable limitations of this study would be mitigated.

While this study, based on previous research, decided to include two accounting factors, it would be interesting to see the result of research which would be testing a broader set of accounting variables. The testing of the accounting factors is of particular interest as this research tested both accounting variables to be significant although the result for earnings-per-share was unstable.

If a ratio is chosen as an independent variable, the authors of this paper will recommend testing the individual components of the ratio as well. This should be done as the coefficient of a ratio is not as easily interpreted as it is for an individual component. Furthermore, by including the individual components, it allows the researcher to identify if the component is the main driver of the stock price and not the ratio itself.

7. Bibliography

Aboura, S. and Chevallier, J. (2017) ‘A new weighting-scheme for equity indexes’, International Review of Financial Analysis. Elsevier Inc., 54, pp. 159–175. doi: 10.1016/j.irfa.2016.11.004.

Adelman, M. A., 1993. *The Economics of Petroleum Supply: Papers by M. A. Adelman,* Cambridge, MA (USA): The MIT Press.

Alibaba.com, 2018. *Alibaba Home Page.* [Online]
Available at: https://www.alibaba.com/
[Accessed 8 May 2018].

Allisaon, P., 2014. *Prediction vs. Causation in Regression Analysis.* [Online]
Available at: https://statisticalhorizons.com/prediction-vs-causation-in-regression-analysis
[Accessed 8 April 2018].

Allison, P. D., 1999. *Multiple Regression.* 1st Edition ed. Thousands Oaks, California: Pine Forge Press.

Al-Majali, A. A. & Al-Assaf, G. I., 2014. LONG-RUN AND SHORT-RUN RELATIONSHIP BETWEEN STOCK MARKET INDEX AND MAIN MACROECONOMIC VARIABLES PERFORMANCE IN JORDAN. *European Scientific Journal* , 10(10), pp. 156-171.

Altinbas, H. & Biskin, O. T., 2015. Selecting Macroeconomic Influencers on Stock Markets by Using Feature Selection Algorithms. *Procedia Economics and Finance,* Volume 30, pp. 22-29.

Amazon.com, 2018. *Our communities.* [Online]
Available at: https://www.aboutamazon.com/
[Accessed 8 May 2018].

Ansotegui, C. & Esteban, M. V., 2002. Cointegration for market forecast in the Spanish stock market. *Applied Economics,* Volume 34, pp. 843-857.

Arrow, K. J., 1982. Risk Perception in Psychology and Economics. *Economic Inquiry,* Volume 20, pp. 1-9.

AXA Equitable Life Insurance Company, 2018. *Growth vs. value: Two approaches to stock selection,* New York: AXA Equitable Life Insurance Company.

Ball, R., 1978. Anomalies in relationships between securities' yields and yield-surrogates. *Journal of Financial Economics,* Volume 6, pp. 103-126.

Banjo, S., 2018. *China Protectionism Creates Tech Billionaires Who Protect Xi.* [Online]
Available at: https://www.bloomberg.com/news/articles/2018-03-06/how-china-protectionism-creates-tech-billionaires-who-protect-xi
[Accessed 8 May 2018].

Barefield, R. M. & Comiskey, E. E., 1975. The Accuracy of Analysts' forecasts of Earnings per Share. *Journal of Business Research,* 3(3), pp. 241-252.

Basu, S., 1977. Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis. *The Journal of Finance,* 32(3), pp. 663-682.

Bauman, S. W., Conover, M. C. & Miller, R. E., 1998. Growth versus Value and Large-Cap versus Small-Cap Stocks in International Markets. *Financial Analysts Journal,* 54(2), pp. 75-89.

Beaver, W. H., Lambert, R. A. & Ryan, S. G., 1987. The Information Content of Security Prices. *Journal of Accounting and Economics,* Volume 9, pp. 139-157.

Becketti, S., 2013. *Introduction to Time Series Using Stata.* 1st Edition ed. New York: Stata Pres.

Bhargava, A., 2014. Firm's fundamentals, macroeconomic variables and quarterly stock prices in the US. *Journal of Econometrics,* Volume 183, pp. 241-250.

Biddle, G. C. and Saudagaran, S. M. (1991) ‘Foreign stock listings: Benefits, costs, and the accounting policy dilemma’, Accounting Horizons, 5(3), pp. 69–80.

Bilgili, F., 1998. *Stationarity and cointegration tests: Comparison of Engle - Granger and Johansen methodologies,* s.l.: MPRA Paper No. 75967.

Blackman, A. (2003) Is There a Better Way to Build Indexes?, Wall Street Journal. Available at: https://www.wsj.com/articles/SB10001424127887323844804578527611352978912 (Accessed: 20 March 2018).

Bloomberg L.P. , 2018. *Bloomberg Markets, China.* [Online]
Available at: https://www.bloomberg.com/search?query=china&sites=markets
[Accessed 21 April 2018].

Bloomberg L.P., 2018. *Internet Software and Services, Company Overview of NewlineNoosh, Inc.*. [Online]
Available at: https://www.bloomberg.com/research/stocks/private/snapshot.asp?privcapId=32207
[Accessed 8 May 2018].

Bloomberg News, 2017. *Baidu Snags 50-Plus Partners for its Apollo Driverless Car.* [Online]
Available at: https://www.bloomberg.com/news/articles/2017-07-05/baidu-signs-50-plus-partners-for-apollo-driverless-car-project
[Accessed 8 May 2018].

Bollen, K. A. & Jackman, R. W., 1990. Regression diagnostics: An expository treatment of outliers and influential cases.. *Sociological Methods and Research,* 13(4), pp. 257-291.

Brandt, M. W. & Kang, Q., 2004. On the relationship between the conditional mean and volatility of stock returns: A latent VAR approach. *Journal of Financial Economics,* Volume 72, pp. 217-257.

Brockwell, P. J. & Davis, R. A., 2016. Introduction. In: S. I. P. Switzerland, ed. *Introduction to Time Series and Forecasting* . Zurich: Springer International Publishing Switzerland, pp. 1-34.

Brodie, L., 2013. *Cramer: does your portfolio have FANGs?.* [Online]
Available at: https://www.cnbc.com/id/100436754
[Accessed 6 May 2018].

Bryman, A., Bell, E., Mills, A. J. & Yue, A. R., 2011. *Business Research Methods.* 3rd Edition ed. Oxford: Oxford University Press .

Buffett, W. E., 1989. *To the Shareholders of Berkshire Hathaway Inc.,* New York: BERKSHIRE HATHAWAY INC..

Burbidge, J. & Harrison, A., 1984. Testing for the Effects of Oil-Price Rises Using Vector Autoregressions. *International Economic Review,* 25(2), pp. 459-484.

Campbell, H. R., 1989. Forecast of Economic Growth from the Bond and Stock Markets. *Financial Analyst Journal,* 45(5), pp. 38-45.

Campbell, J. Y. & Shiller, R. J., 1988. Stock Prices, Earnings, and Expected Dividends. *The Journal of Finance,* Volume 43, pp. 661-676.

Chan, K., Chan, K. C. and Karolyi, A. (2018) 'Intraday Volatility in the Stock Index and Stock Index Futures Markets', Review of Financial Studies, 4(4), pp. 657–684.

Chang, H.-L., Chen, Y.-S., Su, C.-W. & Chang, Y.-W., 2008. The Relationship between Stock Price and EPS: Evidence Based on Taiwan Panel Data. *Economics Bulletin,* 3(30), pp. 1-12.

Chishom, A. M., 2009. *An Introduction to International Capital Markets: Products, Strategies, Participants.* 2nd Edition ed. London: John Wiley and Sons Ltd.

CNBC, 2015. *Cramer: FANG is bad, here's why.* [Online]
Available at: https://www.cnbc.com/2015/07/27/cramer-fang-is-bad-heres-why.html
[Accessed 20 April 2018].

Cohen, J., Cohen, P., West, S. G. & Aiken, L. S., 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Science.* 3rd Edition ed. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc..

Copenhagen Business School, 2018. *Datastream.* [Online]
Available at: https://www.cbs.dk/en/library/databases/datastream
[Accessed 21 April 2018].

Crippen, A. & Cheng, E., 2017. *Warren Buffett: I was wrong on Google and 'too dumb' to appreciate Amazon.* [Online]

Available at: https://www.cnbc.com/2017/05/06/warren-buffett-admits-he-made-a-mistake-on-google.html
[Accessed 8 May 2018].

Cutler, D. M., Poterba, J. M. & Summers, L. H., 1988. *What moves stock prices?,* Cambridge, MA: U.S. National Buereau of Economic Research.

Damodaran, A., 2012. *Investment Valuation: Tools and Techniques for Determining the Value of Any Asset.* 3rd Edition ed. New York: Wiley Finance.

De Bondt, W. F. M. & Thaler, R., 1985. Does the Stock Market Overreact?. *The Journal of Finance,* 40(3), pp. 793-805.

Dean, T. J., Brown, R. L. . and Bamford, C. E. . (1998) 'Differences in Large and Small Firm Responses to Environmental Context', Strategic Management Journal, 19(8), pp. 709–729.

Degiannakis, S., Filis, G. & Arora, V., 2017. Oil Prices and Stock Markets. *Independent Statistics & Analysis,* pp. 1-51.

Deloitte, 2018. *2018 Technology Industry Outlook,* London: Deloitte.

Desjardins, J., 2017. *Here are the 20 biggest stock exchanges in the world.* [Online]
Available at: http://www.businessinsider.com/here-are-the-20-biggest-stock-exchanges-in-the-world-2017-4?r=US&IR=T&IR=T
[Accessed 8 May 2018].

Dickey, D. A. & Fuller, W. A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Associatio,* Volume 74, pp. 427-431.

Doidge, C., Karolyi, G. A. and Stulz, R. M. (2004) 'Why are foreign firms listed in the U.S. worth more?', Journal of Financial Economics. North-Holland, 71(2), pp. 205–238. doi: 10.1016/S0304-405X(03)00183-1.

Durand, D., 1957. Growth Stocks and the Petersburg Paradox. *The Journal of Finance,* 12(3), pp. 348-363.

Durbin, J. & Watson, G., 1950. Testing for serial correlation in least squares regression.. *Biometrika,* Volume 37, pp. 409-428.

Edelman, R. B. and Baker, H. K. (1990) 'Liquidity and Stock Exchange Listing', The Financial Review. Wiley/Blackwell (10.1111), 25(2), pp. 231–249.

Elliott, G. R., Rothenberg, T. J. & Stock, J. H., 1996. Efficient tests for an autoregressive unit root. *Econometrica,* Volume 64, pp. 813-836.

Engle, R. & Granger, C., 1987. Co-integration and Error Correction: Representation, Estimation and Testing. *Econometrica,* Volume 55, pp. 251-276.

FACEBOOK INC (FB) IPO - NASDAQ.com (2018). Available at: https://www.nasdaq.com/markets/ipos/company/facebook-inc-673740-69138 (Accessed: 22 March 2018).

Facebook, 2018. *Recent News and Announcements.* [Online]
Available at: https://newsroom.fb.com/
[Accessed 8 May 2018].

Fama, E. & Schwert, W. G., 1977. Asset Returns and Inflation. *Journal of Financial Economics,* Volume 5, pp. 115-146.

Fama, E. F. & French, K. R., 1988. Permanent and temporary components of stock prices. *Journal of Political Economy,* 96(2), pp. 246-273.

Fama, E. F. & French, K. R., 1992. The Cross-Section of Expected Stock Returns. *The Journal of Finance,* XLVII(2), pp. 427-465.

Fama, E. F. & French, K., 1993. Common risk factors in the returns on stocks and bonds. *The Journal of Financial Economics,* Volume 33, pp. 3-56.

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. *Journal of Finance,* Volume 25, pp. 383-417.

Fama, E. F., 1981. Stock returns, real activity, inflation and money. *American Economic Review,* Volume 71, pp. 545-565.

Fisher, I., 1907. *The Rate of Interes. Its nature, determination and relation to economic phenomena.* 1st Edition ed. New York: The Macmillan Company.

Foretti, P., 2007. Testing for Granger causality between stock prices and economic growth. *Munich Personal RePEc Archive* , Volume 4, pp. 1-10.

Foster, G., 1973. Stock Market Reaction to Estimates of Earnings per Share by Company Officials. *Journal of Accouting Research,* 11(1), pp. 25-37.

Freeman, E., 1990. Stock returns, expected returns and real activity. *Journal of Finance,* Volume XLV, pp. 1089-1108.

Friedman, M. & Schwartz, A. J., 1963. Money and Business Cycle. *Review of Economics and Statistics, Supplement,* pp. 32-64.

Fuller, W., 1996. *Introduction to Statistical Time Series.* 2nd Edition ed. New York: Wiley.

Garcia, J. A., 2013. *A Bit About the St. Petersburg Paradox,* Houston: Texas A&M University.

Gelman, A. & Hill, J., 2006. Causal inference using regression on the treatment variable. In: C. U. Press, ed. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge: Cambridge University Press, pp. 167-198.

Geske, R. & Roll, R., 1983. The monetary and fiscal linkage between stock returns and inflation. *Journal of Finance,* Volume 38, pp. 1-33.

Golafshani, N., 2003. Understanding Reliability and Validity in Qualitative Research. *The Qualitative Repor,* 8(4), pp. 597-606.

Goldman, D., 2010. *10 big dot.com flops.* [Online]
Available at: http://money.cnn.com/galleries/2010/technology/1003/gallery.dot_com_busts/
[Accessed 9 May 2018].

Google Inc., 2018. *Our Commitments.* [Online]
Available at: https://www.google.com/about/our-company/
[Accessed 8 May 2018].

Gottwald, R., 2011. The Dependence Between Stock Price and Intrinsic Value of a Stock. *Journal of Interdesciplinary Research,* 1(1), pp. 38-40.

Graham, B. & Dodd, D. L., 1988. *Security Analysis.* 5th Edition ed. New York: McGraw Hill.

Graham, B. & Dodd, D. L., 2009. *Security Analysis.* 6th Edition ed. New York: McGraw Hill.

Granger, C. & Engle, R., 1987. Co-Integration and Error Correction: Representation, Estimation, and Testing. *The Econometric Society,* 55(2), pp. 251-276.

Granger, C., 1981. Some Properties of Time Series Data and their use in Econometric Model Specificaiton. *Journal of Econometrics,* Volume 16, pp. 121-130.

Hamilton, J. D., 1983. Oil and the Macroeconomy since World War II. *The Journal of Political Economy,* 91(2), pp. 228-248.

Hammersley, M., Lewis-Beck, M. S., Bryman, A. & Liao, T., 2011. Literature Review. *The SAGE Encyclopedia of Social Science Research Methods,* pp. 1-4.

Hand, J. R. M., 2001. The Role of Book Income, Web Traffic, and Supply and Demand in the Pricing of U.S. Internet Stocks. *European Finance Review,* Volume 5, pp. 295-317.

Heale, R. & Twycross, A., 2015. *Validity and reliability in quantitative studies,* Sudbury, Canada: School of Nursing, Laurentian University.

Healy, P. M., Hutton, A. P. & Palepu, K. G., 1999. Stock Performance and Intermediation Changes Surrounding Sustained Increases in Disclosure. *Contemporary Accounting Research,* 16(2), pp. 485-520.

Hillier, D. et al., 2013. A Description of Efficient Capital Markets. In: M. Education, ed. *Corporate Finance.* Maidenhead (UK): McGraw-Hill Education, pp. 353-377.

Hillier, D., Ross, S. W. R., Jaffe, J. & Jordan, B., 2013. *Corporate Finance.* 2nd Edition ed. New York: McGraw Hill.

Hjalmarsson, E. & Österholm, P., 2007. *Testing for Cointegration Using the Johansen Methodology when Variables are Near-Integrated* , Uppsala: International Monetary Fund .

Hobdari, B., 2016. *Time Series and Forecasting.* Copenhagen: Copenhagen Business School.

Hoium, T., 2014. *The Motley Fool.* [Online]
Available at: https://www.fool.com/investing/general/2014/03/30/what-you-can-learn-from-warren-buffett-about-inter.aspx
[Accessed 8 May 2018].

Holt, C. C., 1962. The Influence of Growth Duration on Share Prices. *The Journal of Finance,* 17(3), pp. 465-475.

Huang, R. D., Masulis, R. W. & Stoll, H. R., 1996. Energy shocks and financial markets. J. Futures Markets. *Journal of Future Markets,* 16(1), pp. 1-27.

Hulbert, M. (2003) STRATEGIES; 4 Ways to Build an Index (None of Them Ideal) - The New York Times, NY Times. Available at: http://www.nytimes.com/2003/08/31/business/strategies-4-ways-to-build-an-index-none-of-them-ideal.html (Accessed: 20 March 2018).

Index Definition - NASDAQ.com (2018). Available at: https://www.nasdaq.com/investing/glossary/i/index (Accessed: 20 March 2018).

iResearch (2017) Top 100 Internet Companies in China by Revenue in 2016, iResearch. Available at: http://www.iresearchchina.com/content/details7_38543.html (Accessed: 22 March 2018).

Jagannathan, R. and Wang, Z. (1996) ‘The Conditional CAPM and the Cross-Sectionof’, Journal of Finance, 51(1), pp. 3–53.

James, C., Koreisha, S. & Partch, M., 1985. A VARMA analysis of the causal relations among stock returns, real output, and nominal interest rates. *Journal of Finance,* Volume 40, pp. 1375-1384.

JD.com, 2018. *JD.com Home page.* [Online]
Available at: http://www.jd.com/
[Accessed 8 May 2018].

Jensen, M. C., 1978. Some Anomalous Evidence Regarding Market Efficiency. *Journal of Financial Economics,* Volume 6, pp. 95-101.

Johnson, D., 2017. *News in Brief 13 September 2017 - Geneva (AM).* [Online]
Available at: http://www.un.org/en/index.html
[Accessed 8 May 2018].

Johnson, P. & Clark, M., 2006. *Mapping the terrain : an overview of business and management research methodologies.* 6th Edition ed. Thousand Oaks, United States: SAGE Publications Inc.

Jones, C. M. & Kaul, G., 1996. Oil and the stock markets. *The Journal of Finance,* 51(2), pp. 463-491.

Kaplan, S. N. & Ruback, R., 1995. The Valuation of Cash Flow Forecasts: An Empirical Analysis. *The Journal of Finance,* 1(4), pp. 1059-1093.

Kaul, G., 1987. Stock returns and inflation: the role of the monetary sector. *Journal of Financial Economics,* Volume 18, pp. 253-276.

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

Kilian, L. and Park, C. (2009) ‘The impact of oil price shocks on the U.S. stock market’, International Economic Review, 50(4), pp. 1267–1287.

King, S. L., 1997. *NEURAL NETWORKS VS. MULTIPLE LINEAR REGRESSION FOR ESTIMAT_NG PREVIOUS DHAIViETER.*

Kjersti, A. & Dimakos, X. K., 2004. *Statistical modelling of financial time series: An introduction,* Oslo: Norwegian Computing Center .

Koller, T., Goedhart, M. & Wessels, D., 2010. *Valaution: Measuring and Managing the Value of Companies.* 5th Edition ed. New York: John Wiley & Sons, Inc..

Kormendi, R. & Lipe, R., 1987. Earnings Innovations, Earnings Persistence, and Stock Returns. *The Journal of Business,* 60(3), pp. 323-345.

Kumar, N. P. & Puja, P., 2012. *The Impact of Macroeconomic Fundamentals on Stock Prices revisited: An Evidence from Indian Data.* [Online]
Available at: https://mpra.ub.uni-muenchen.de/38980/
[Accessed 8 January 2018].

Lakonishok, J. & Shapiro, A. C., 1986. SYSTEMATIC RISK, TOTAL RISK AND SIZE AS DETERMINANTS OF STOCK MARKET RETURNS. *Journal of Banking and Finance,* Volume 10, pp. 115-132.

Lakonishok, J., Shleifer, A. & Vishny, R. W., 1994. Contrarian Investment, Extrapolation and Risk. *The Journal of Finance,* XLIX(5), pp. 1541-1578.

Lamont, O. & Thaler, R. H., 2003. Can the Market Add and Subtract? Mispricing in Tech Stock Carve-Outs. *Journal of Political Economy,* Volume 111, pp. 227-268.

Lamoureux, C. G. & Zhou, G., 1996. Temporary Components of Stock Returns: What Do the Data Tell Us?. *The Review of Financial Studies,* 9(4), pp. 1033-1059.

Langevoort, D. C., 1992. Assumptions and Securitites Regulation: Market Efficiency Revised. *University of Pennsylvania Law Review,* 140(3), pp. 851-920.

Lewis, M. M., 1989. *Liar's Poker: Rising Through the Wreckage on Wall Street.* Norton ed. New York: Norton.

Lintner, J., 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics,* 47(1), pp. 13-37.

Litzenberger, R. H. & Rao, C. U., 1971. Estimates of the Marginal Rate of Time Preference and Average Risk Aversion of Investors in Electric Utility Shares: 1960-66. *The Bell Journal of Economics and Management Science,* 2(1), pp. 265-277.

Loungani, P., 1986. Oil Price Shocks and the Dispersion Hypothesis. *The Review of Economics and Statistics,* 68(3), pp. 536-539.

MacKinnon, J. G., 1990. *Critical Values for Cointegration Tests,* Ontario, Canada: Department of Economics Queen's University.

Maio, P. & Philip, D., 2015. Macro variables and the components of stock returns. *Journal of Empirical Finance,* Volume 33, pp. 287-308.

Malkiel, B. G. & Cragg, J. G., 1970. Expectations and the Structure of Share Prices. *The American Economic Review,* 60(4), pp. 601-617.

Mallows, C. L., 1986. Augmented Partial Residuals. *Technometric,* 28(4), pp. 313-319.

Mankiw, G. N., 2005. *Macroeconomics.* 7th Edition ed. Boston: Harvard University.

Mauldin, J., 2018. *The Dark Side Of Social Networks Could Destroy FAANGs.* [Online]
Available at: https://www.forbes.com/sites/johnmauldin/2018/03/28/the-dark-side-of-social-networks-could-destroy-faangs/2/#6f08f1216674
[Accessed 20 April 2018].

McAnally, M. L., Beaver, W. H. & Stinson, C. H., 1997. The information content of earnings and prices: A simultaneous equations approach. *Journal of Accounting and Economics,* Volume 23, pp. 53-81.

McKinsey&Company, 2015. *Valuing high-tech companies,* New York: McKinsey&Company.

Miller, R. E. & Bauman, S. W., 1997. Investor Expectations and the Performance of Value Stocks versus Growth Stocks. *Journal of Portfolio Management,* 23(3), pp. 57-68.

Modigliani, F. & Cohn, R., 1979. Inflation, Rational Valuation and the Market. *Financial Analysis Journal,* Volume 35, pp. 24-44.

Modigliani, F. & Miller, M. H., 1958. The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review,* 48(3), pp. 261-297.

Modigliani, F. & Miller, M. H., 1961. Dividend Policy, Growth, and the Valuation of Shares. *The Journal of Business* , 34(4), pp. 411-433.

Montgomery, D. C., Peck, E. A. & Vining, G. G., 2012. *Introduction to linear regression analysis.* 5th Edition ed. Toronto: John Wiley & Sons, Inc..

Mukherjee, T. K. & Naka, A. N., 1995. Dynamic Relations between Macroeconomic Variables and the Japanese Stock Market: An Application of a Vector Error Correction Model. *Journal of Financial research,* 18(2), pp. 223-237.

Nasdaq, 2013. *Reverse DCF: Another way to evaluate stocks,* New York: Nasdaq.

Nasdaq.com, 2018. *FACEBOOK INC (FB) IPO.* [Online]
Available at: https://www.nasdaq.com/markets/ipos/company/facebook-inc-673740-69138
[Accessed 8 May 2018].

Nasseh, A. & Strauss, j., 2000. Stock prices and domestic and international macroeconomic activity: a cointegration approach. *The Quarterly Review of Economics and Finance,* Volume 40, pp. 229-245.

National Business Research Institute (NBRI) (2018) Validity of Your Survey Results. Available at: https://www.nbrii.com/our-process/validity/ (Accessed: 14 May 2018).

Newbold, P., 2017. *Statistics for Business and Economics.* s.l.:Pearson.

OECD, 2005. *Data and Metadata Reporting and Presentation Handbook,* Paris: OECD .

Oyedele, A., 2017. *BUSINESS INDISER Nordics.* [Online]
Available at: http://nordic.businessinsider.com/faamg-tech-stocks-market-goldman-2017-6
[Accessed 20 April 2018].

Palmer, M., 1970. Money Supply, Portfolio Ad justments and Stock Price. *Financial Analysts Journal,* pp. 19-22.

Pearce, D. K. & Roley, V. V., 1983. The reaction of stock prices to unanticipated changes in money: a note. *The Journal of Finance,* 38(4), pp. 1323-1333.

Pearce, D. K., 1984. *Stock Prices and Economic News,* Cambridge, MA: U.S. National Bureau of Economic Reserch.

Peirò, A., 2016. Stock Prices and Macroeconomic Factors: some European evidence. *International Review of Economics and Finance,* Volume 41, pp. 287-294.

Penman, S. H. & Sougiannis, T., 1998. A Comparison of Dividend, Cash Flow, andEarnings Approaches to Equity Valuation. *COntemporary Accounting Research,* 15(3), pp. 343-383.

Penman, S. H., 1996. The Articulation of Price-Earnings Ratios and Market-to-Book Ratios and the Evaluation of Growth. *Journal of Accouting Research,* 34(2), pp. 235-259.

Pethe, A. & Karnik, A., 2000. Do Indian Stock Markets Matter? Stock Market Indices and Macro-Economic. *Economic and Political Weekly* , 35(5), pp. 349-356.

Petkove, R. & hang, L., 2005. Is value riskier than growth?. *The Journal of Financial Economics,* Volume 78, pp. 187-202.

Pham, S., 2018. *iQiyi, the Netflix of China, is going public in the US.* [Online]
Available at: http://money.cnn.com/2018/02/28/technology/baidu-iqiyi-netflix-china/index.html
[Accessed 8 May 2018].

Phillips, P. C. B. & Perron, P., 1988. Testing for a unit root in time series regression. *Biometrika,* Volume 75, pp. 335-346.

Prakash, G. and Loungani, V. (2000) 'THE DIFFERENTIAL IMPACT OF UNCERTAINTY ON INVESTMENT IN SMALL AND LARGE BUSINESSES', Review of Economics & Statistics, 82(2), p. 338.

Preinreich, G. A. D., 1932. Stock Yields, Stock Dividends and Inflation. *The Accounting Review,* 7(4), pp. 273-289.

PWC, 2017. *Global Top 100 Companies by market capitalisation,* London: Price Waterhouse Coopers LLP..

Ram, R. & Spencer, D. E., 1983. Stock returns, real activity, inflation and money: Comment. *American Economic Review,* Volume 74, pp. 463-470.

Reggiani, F. & Penman, S., 2010. *Returns to Buying Earnings and Book Value: Accounting for Growth and Risk,* Milano: SDA Bocconi.

Reilly, F. K. & Lewis, J. E., 1971. *Monetary Variables and Stock Price,* Lawrence, KS: School of Business, University of Kansas.

Risager, O., 2016. *Global Macroeconomics for Business and Investors.* 3rd Edition ed. New York: McGraw-Hill Education .

Risager, O., 2016. Mr Market and the efficient market hypothesis. In: M. Education, ed. *Global Macroeconomics for Business and Investors.* New York: McGraw-Hill Education, pp. 144-165.

Robson, C. & McCartan, K., 2015. *Real World Research.* 4th Edition ed. Cambridge, MA: John Wiley & Sons Inc.

Roll, R., 1984. Orange Juice and Weather. *American Economic Review,* Volume 74, pp. 861-880.

Rosenberg, B., Reid, K. & Lanstein, R., 1985. Persuasive evidence of market inefficiency. *Journal of Portfolio Management,* Volume 11, pp. 9-17.

Rozeff, M. S., 1975. The Money Supply and the Stock Market: The Demise of a Leading Indicator. *Financial Analyst Journal,* 31(5), pp. 18-76.

Sadorsky, P., 1999. Oil price shocks and stock market activity. *Energy Economics,* Volume 21, pp. 449-469.

Salgado-Ugarte, I. H., Shimizu, M. & Taniuchi, T., 1993. snp6: Exploring the shape of univariate data using kernel density estimators. *Stata Technical Bulletin,* Volume 3, pp. 155-173.

Samuelson, P. A., 1998. Summing up on business cycles: opening address. *Federal Reserve Bank of Boston,* Volume 42, pp. 33-36.

Saunders, M., Lewis, P. & Thornhill, A., 2016. Chapter 4 "Understanding research philosophies and approaches". In: P. E. Limited, ed. *Research methods for business students* . Harlow: Pearson Education Limited, pp. 106-134.

Schonfeld, E., 2000. How Much Are Your Eyeballs Worth? Placing a value on a Website's customers may be the best way to judge a Net stock. It's not perfect, but on the Net, what is?. *Fortune Magazine,* pp. 1-3.

Schwert, W. G., 1981. e adjustment of stock prices to information about inflation. *The Journal of Finance,* 36(1), pp. 15-29.

Serven, L. & Solimano, A., 1989. *Private Investment and Macroeconomic Ad justment,* Washington DC: The World Bank.

Setin, C., 2018. *Top 5 Tech Stocks Based On Warren Buffett's Strategy.* [Online]
Available at: https://www.forbesmiddleeast.com/en/top-tech-stocks-warren-buffett/
[Accessed 8 May 2018].

Sharma, R., 2017. *When Will the Tech Bubble Burst? - The New York Times.* [Online]
Available at: https://www.nytimes.com/2017/08/05/opinion/sunday/when-will-the-tech-bubble-burst.html
[Accessed 8 May 2018].

Shell, A., 2017. *Record-setting Nasdaq: What's different today vs. 2000 - USA Today.* [Online]
Available at: https://www.usatoday.com/story/money/markets/2017/05/16/nasdaq-tech-boom-then-and-now/101748912/
[Accessed 9 May 2018].

Shiller, R. J., 1980. Do Stock Prices Move Too Much To Be Justified By Subsequent Changes in Dividends?. *National Bureau Of Economic Research,* Volume 71, pp. 421-436.

Shleifer, A., 2000. *Inefficient Merkets An Introduction to Behavioral Finance.* 1st Edition ed. Oxford: Oxford University Press.

Sing, T. F., H., L. K. & Chiang, W. J., 2002. Mean reversion of Singapore property stock prices towards their fundamental values. *Journal of Property Investment and Finance,* Volume 20, pp. 374-387.

Smidt, S., 1968. A New Look at the Random Walk Hypothesis. *The Journal of Financial and Quantitative Analysis,* 5(2), pp. 235-262.

Speth, W. M. (2003) 'Patent Application Publication (10) Pub. No.: US 2003/0136915 A1 (43)'. United States.

Sprinkel, B. W., 1964. *Money and Stock Prices.* 4th Edition ed. Ann Arbor, MI: R.D. Iewin .

Stata.com, 2013. *dfgls — DF-GLS unit-root test,* s.l.: Stata.com.

Stata.com, 2013. *regress postestimation — Postestimation tools for regress,* s.l.: Stata.com.

Stata.com, 2013. *regress postestimation time series — Postestimation tools for regress with time series,* s.l.: Stata.com.

Stattman, D., 1980. Book values and stock returns. *The Chicago MBA: A Journal of Selected Papers,* Volume 4, pp. 25-45.

Stoll, H. R. and Whaley, R. E. (1990) 'The Dynamics of Stock Index and Stock Index Futures', Journal of Financial & Quantitative Analysis, 25(4), pp. 441–468.

Stulz, R., 1986. Asset pricing and expected inflation. *Journal of Finance,* Volume 41, pp. 209-223.

Summers, L. H., 1986. Does the Stock Market Rationally REflect Fundamental Values?. *The Journal of Finance,* 41(3), pp. 591-601.

Sydney, A. S., 1967. Price Movements in Speculative Markets: Trends or Random Walks. *Industrial Management Review,* 2(2), pp. 7-26.

Székely, G. J. & Richards, D. S., 2012. The St. Petersburg Paradox and the Crash of High-Tech Stocks in 2000. *The American Statistician* , 58(3), pp. 225-231.

Tencent 腾讯, 2018. *About Tencent.* [Online]
Available at: https://www.tencent.com/en-us/abouttencent.html
[Accessed 8 May 2018].

The Financial Times, 2018. *Financial Times.* [Online]
Available at: http://lexicon.ft.com/Term?term=earnings-per-share--EPS
[Accessed 3 April 2018].

Thomson Reuters Datastream, 2018. *Thomson Reuters Datastream.* [Online]
Available at: https://financial.thomsonreuters.com/en/products/tools-applications/trading-investment-tools/datastream-macroeconomic-analysis.html
[Accessed 21 April 2018].

Trueman, B., Wong, M. F. & Zhang, X.-T., 2000. The Eyeballs Have It: Searching for the Value in Internet Stocks. *Journal of Accounting Research,* Volume 38, pp. 137-162.

Tverszy, A. & Kahneman, D., 1981. The Framing of Decisions and the Psychology of Choice. *Science,* Volume 211, pp. 453-458.

Wasserfallen, W., 1989. Macroeconomics news and the stock market. *Journal of Banking and Finance,* Volume 13, pp. 613-626.

Watsham, T. J., 1993. *International Portfolio Management: A Modern Approach.* Longman ed. London: Longman.

Wilcox, J. W., 1984. The P/B-ROE Valuation Model. *Financial Analysts Journal,* 40(1), pp. 58-66.

Yamarone, R., 2016. Gross Domestic Product. In: B. Press, ed. *The trader's guide to key economic indicators.* New York: John Wiley & Sons, pp. 1-46.

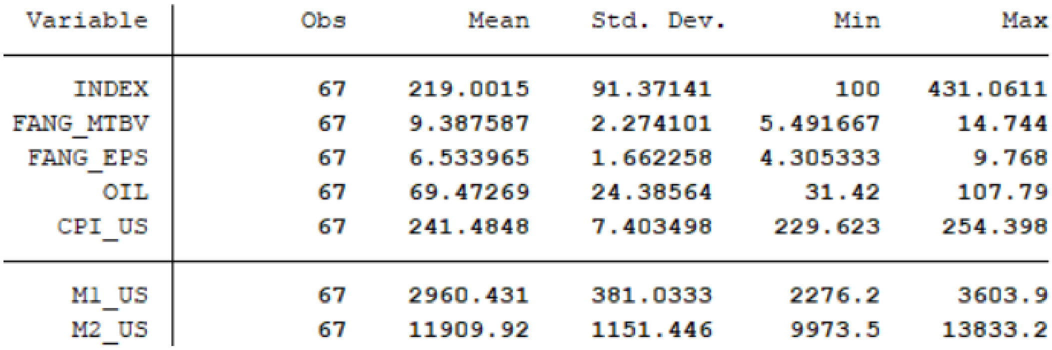
Yamarone, R., 2016. Indexes of Leading, Lagging, and Coincident Indicators. In: B. Press, ed. *The Trader's Guide to Key Economic Indicators.* New York: John Wiley & Sons, pp. 12-47.

Yin, R. K., 2013. *Case Study Research: Design and Method.* 5th Edition ed. Thousand Oaks: SAGE Publications.

8. Appendix

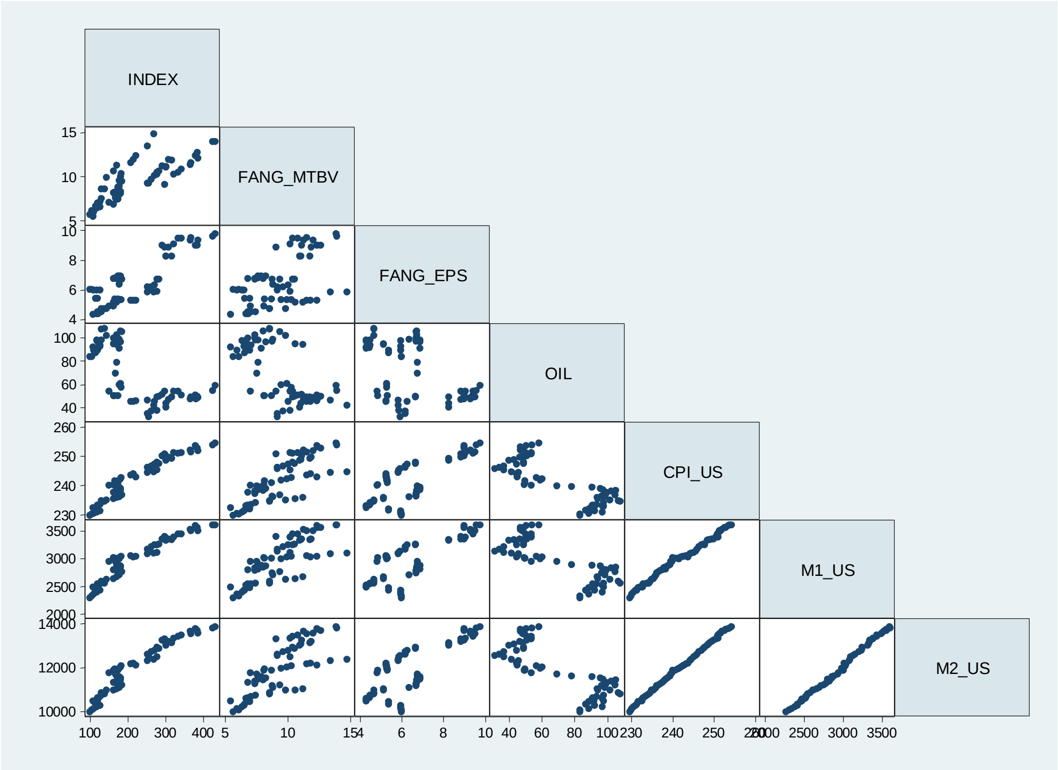
8.1 FANG model – all variables

Data Overview

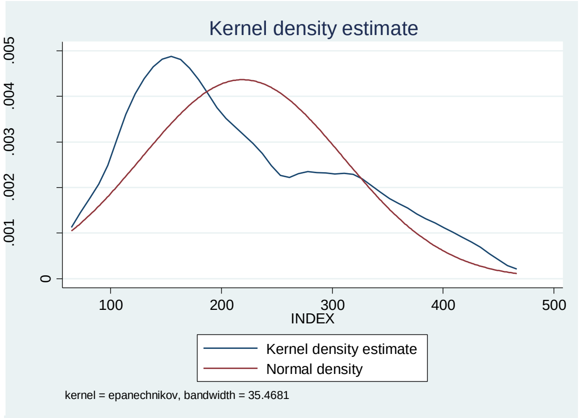


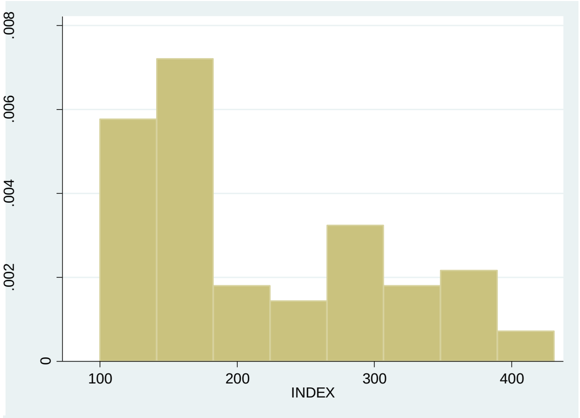
```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       INDEX |         67    219.0015    91.37141        100   431.0611
   FANG_MTBV |         67    9.387587    2.274101   5.491667     14.744
    FANG_EPS |         67    6.533965    1.662258   4.305333      9.768
         OIL |         67    69.47269    24.38564      31.42     107.79
      CPI_US |         67    241.4848    7.403498    229.623    254.398
-------------+---------------------------------------------------------
       M1_US |         67    2960.431    381.0333     2276.2     3603.9
       M2_US |         67    11909.92    1151.446     9973.5    13833.2
```

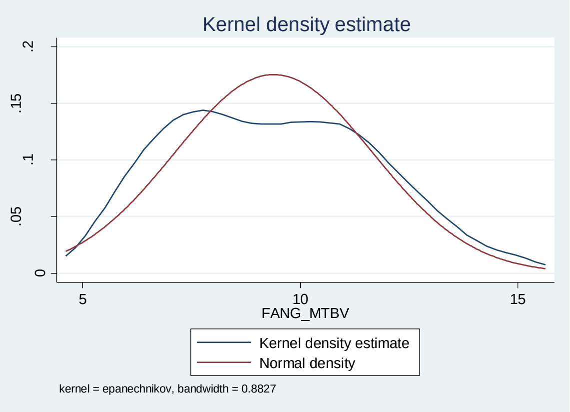
Graph Matrix

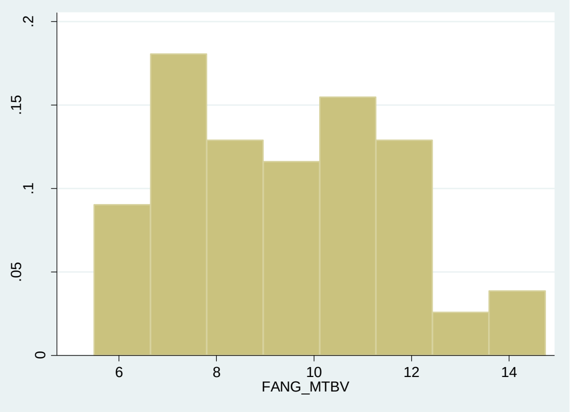


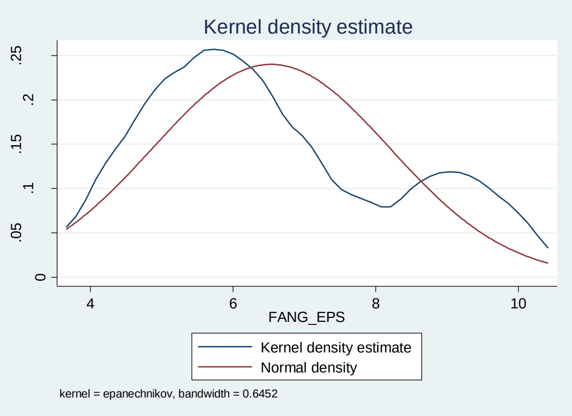
Data distribution – Variables

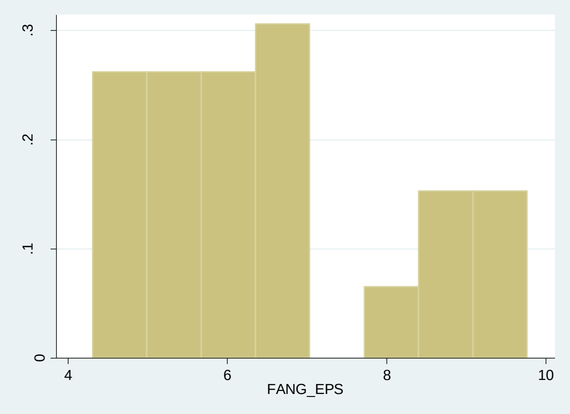


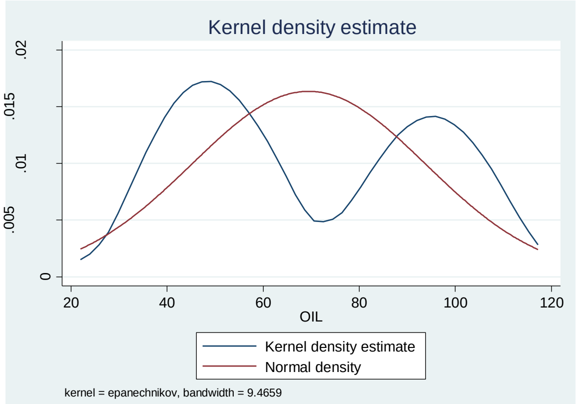


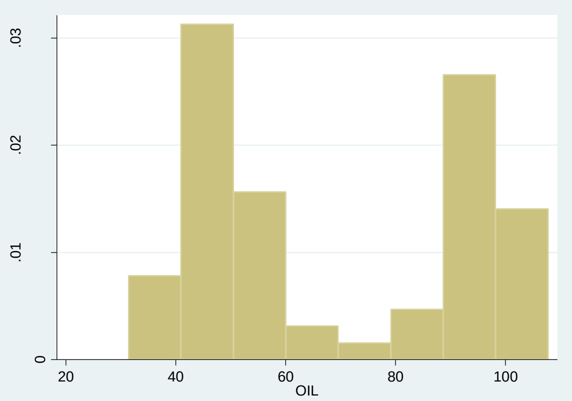


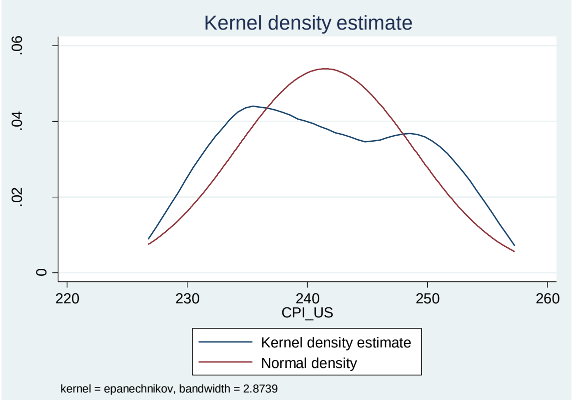


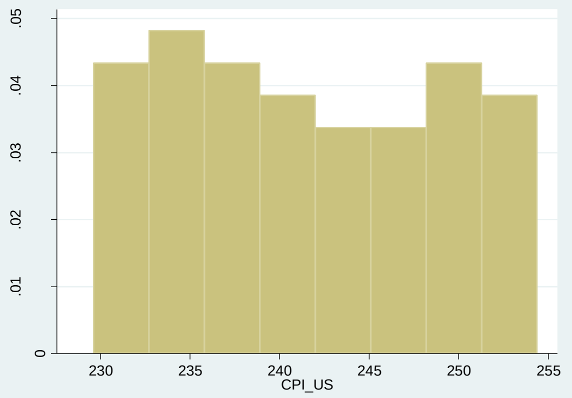


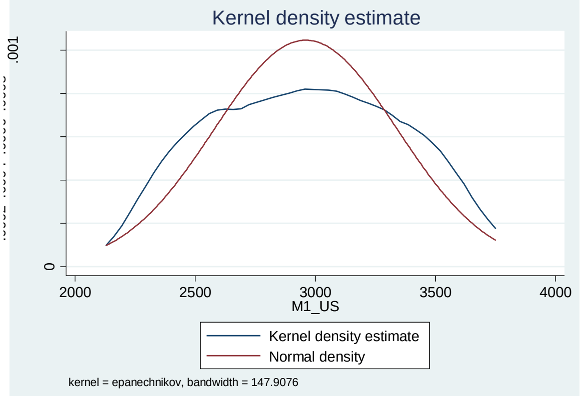


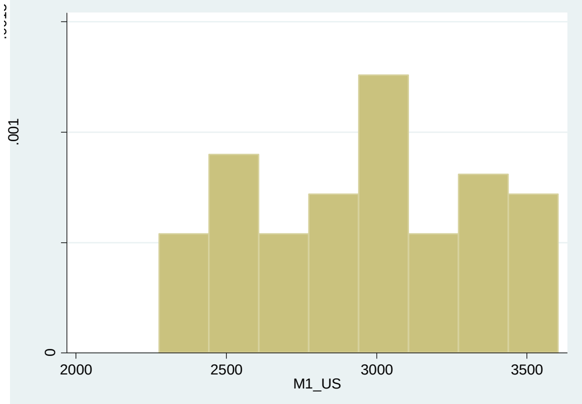

Kernel density estimate
0
.005
.01
.015
.02
20
40
60
80
100
120
OIL
Kernel density estimate
Normal density
kernel = epanechnikov, bandwidth = 9.4659

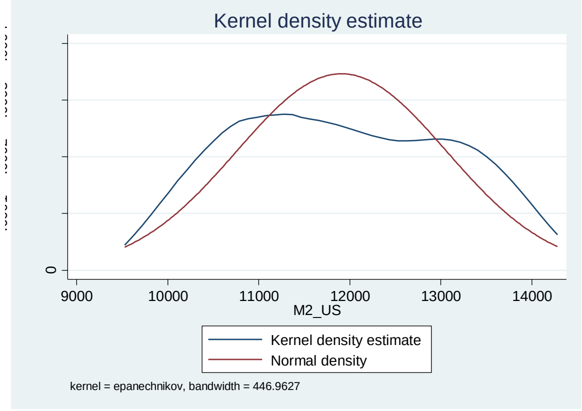

0
.01
.02
.03
20
40
60
80
100
OIL

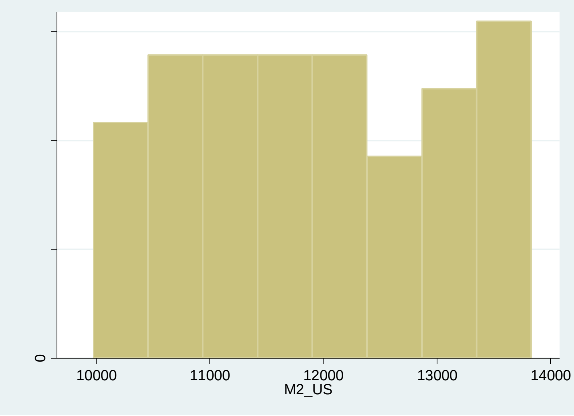

Kernel density estimate
0
.02
.04
.06
220
230
240
250
260
CPI_US
Kernel density estimate
Normal density
kernel = epanechnikov, bandwidth = 2.8739


0
.01
.02
.03
.04
.05
230
235
240
245
250
255
CPI_US

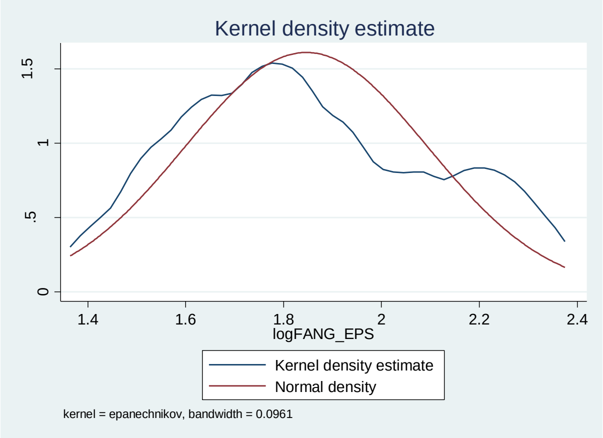

Kernel density estimate
0
.001
2000
2500
3000
3500
4000
M1_US
Kernel density estimate
Normal density
kernel = epanechnikov, bandwidth = 147.9076

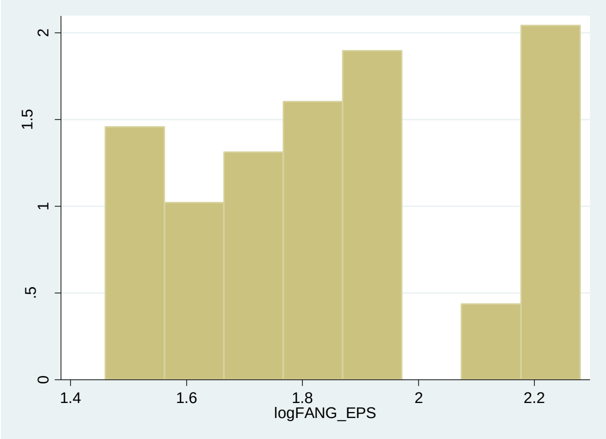

0
.001
2000
2500
3000
3500
M1_US

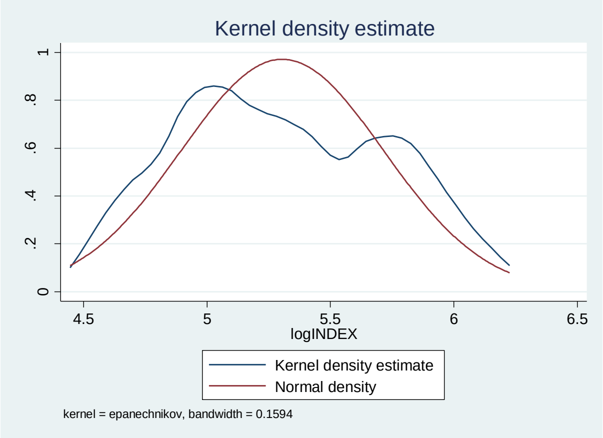

Kernel density estimate
0
9000
10000
11000
12000
13000
14000
M2_US
Kernel density estimate
Normal density
kernel = epanechnikov, bandwidth = 446.9627

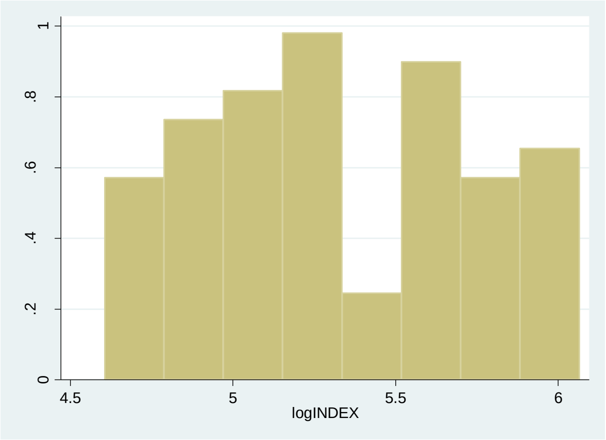

0
10000
11000
12000
13000
14000
M2_US

Data distribution – Log Variables

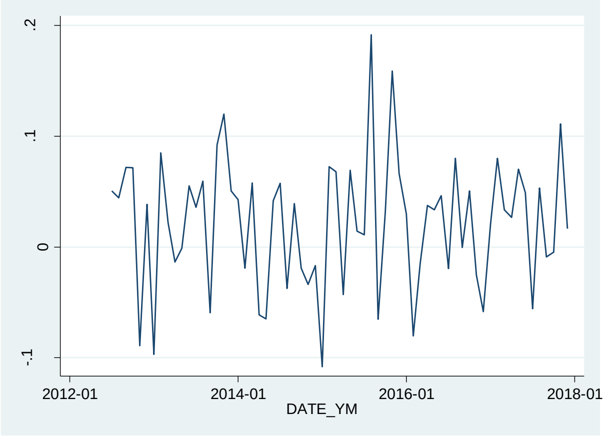


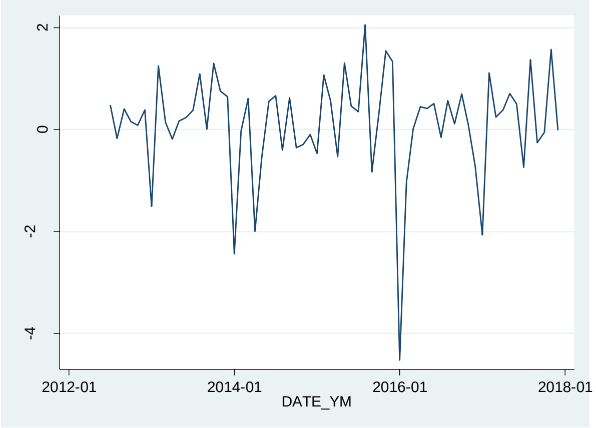


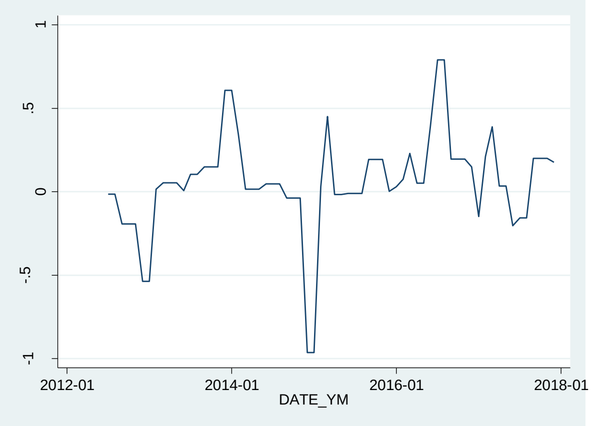


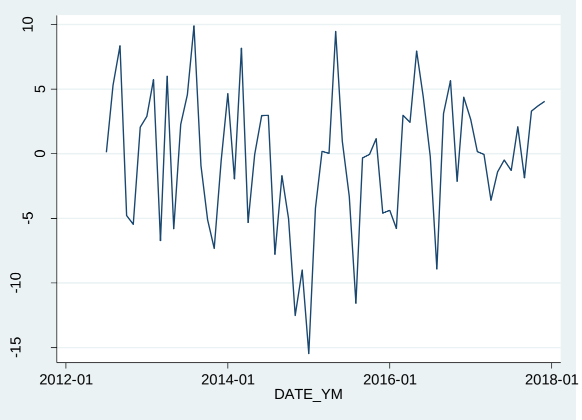


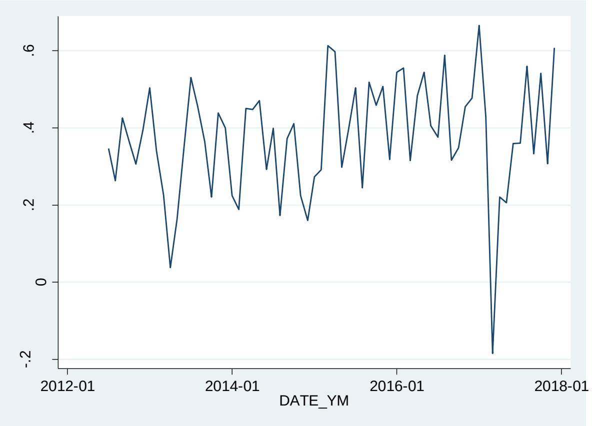
Stationarity

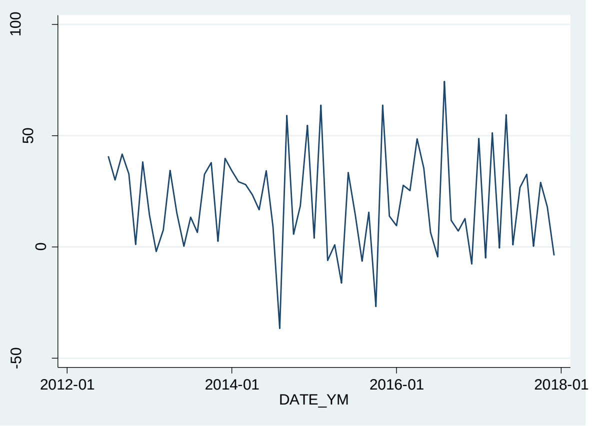


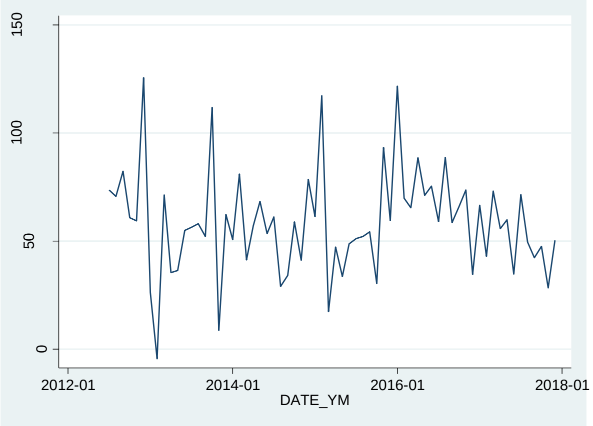



1
.5
0
-.5
-1
2012-01
2014-01
2016-01
2018-01
DATE_YM

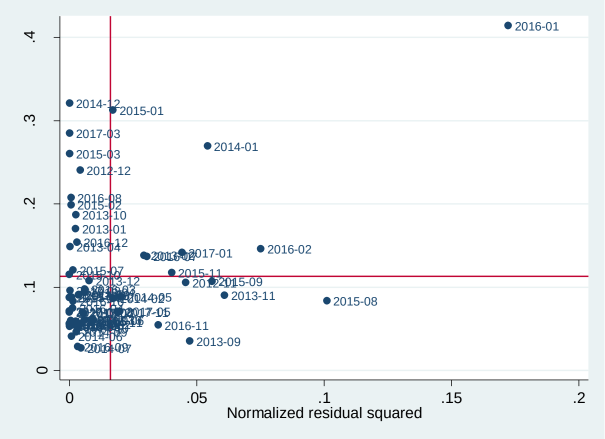

10
5
0
-5
-10
-15
2012-01
2014-01
2016-01
2018-01
DATE_YM

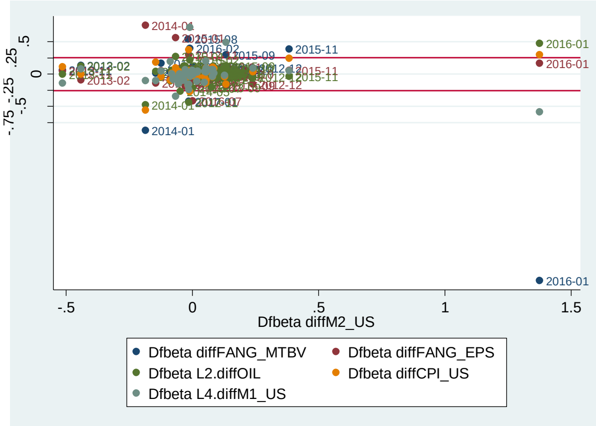

.6
.4
.2
0
-.2
2012-01
2014-01
2016-01
2018-01
DATE_YM

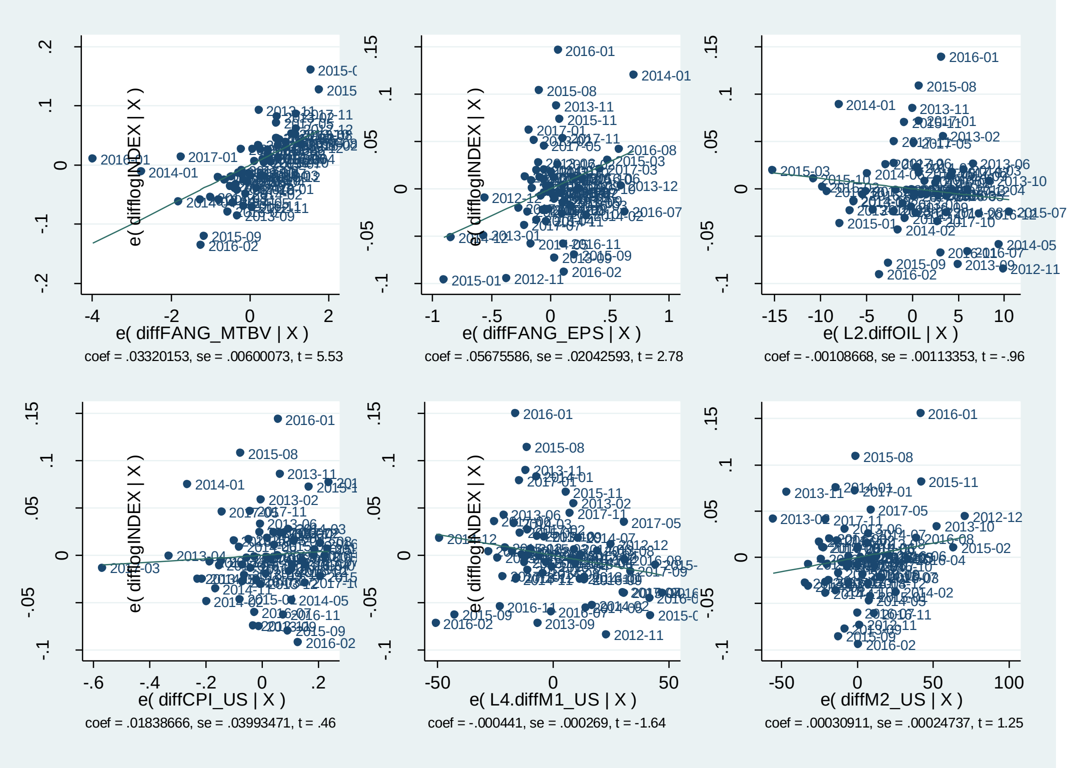

100
50
0
-50
2012-01
2014-01
2016-01
2018-01
DATE_YM


150
100
50
0
2012-01
2014-01
2016-01
2018-01
DATE_YM

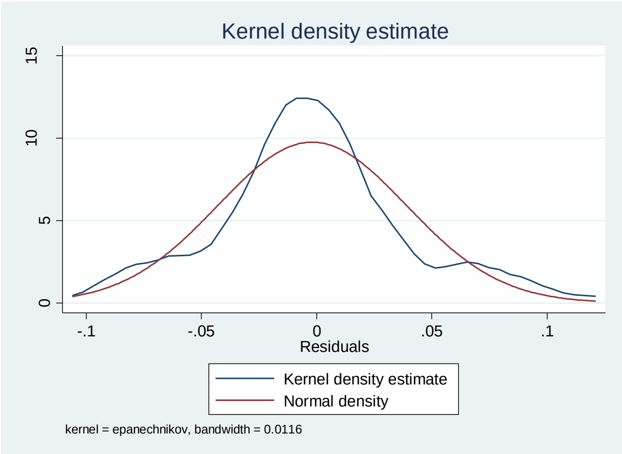
Unusual and influential observations

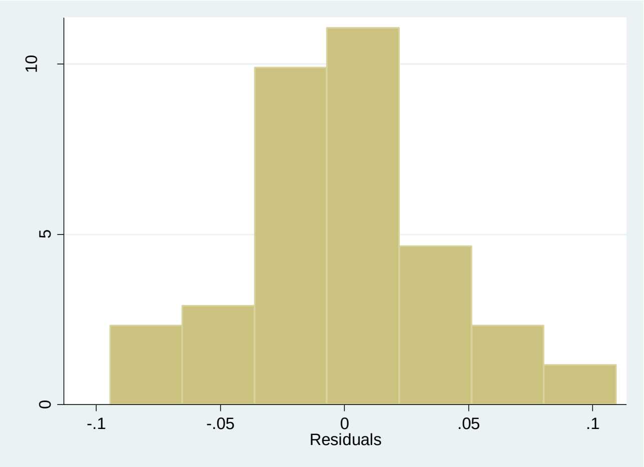


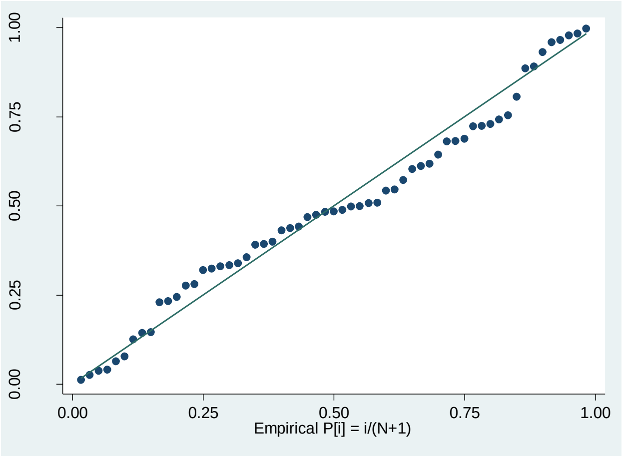


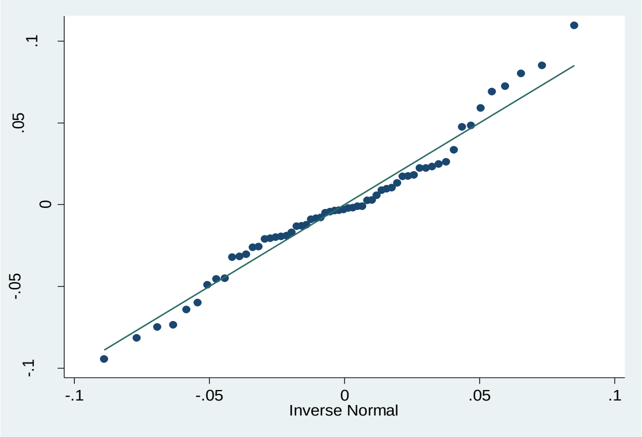


Normality of residuals

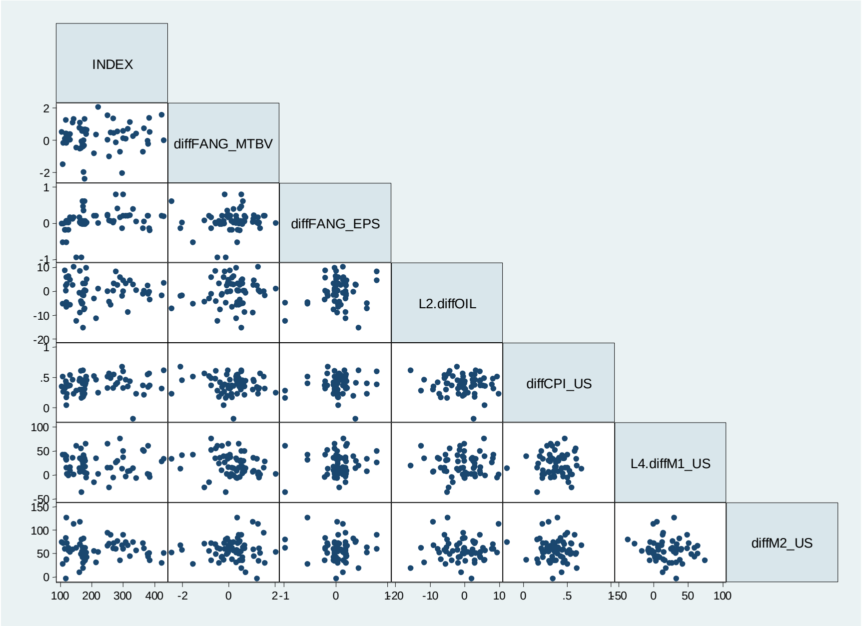


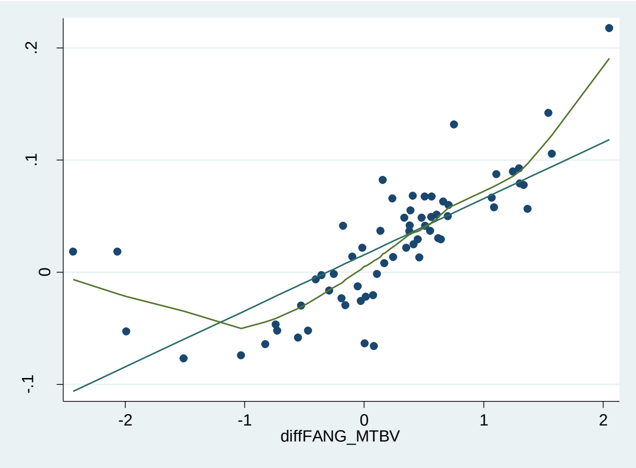


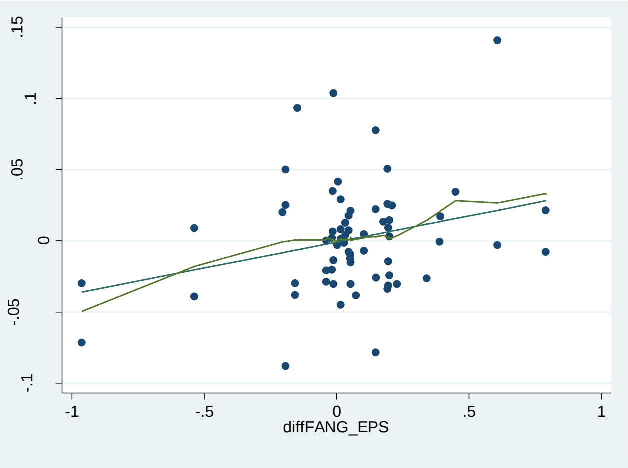


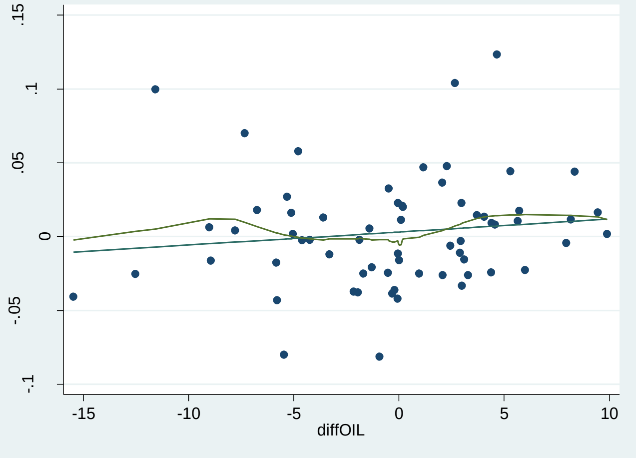


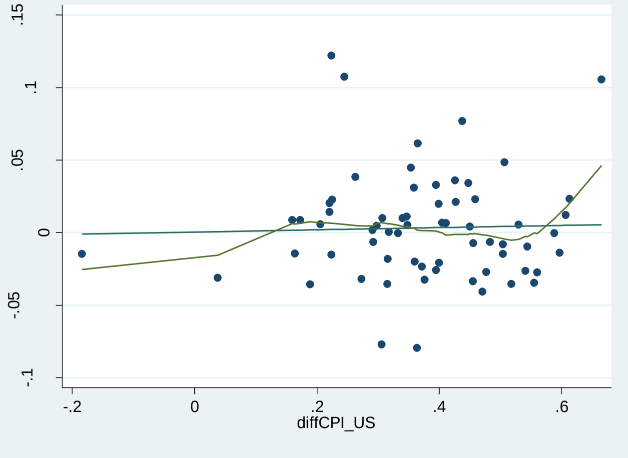
Linearity

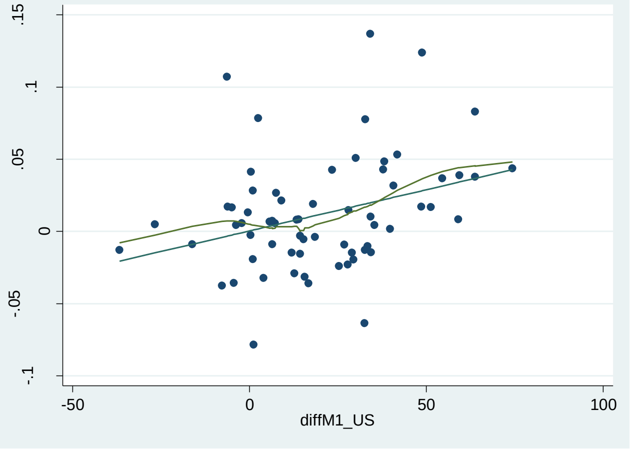


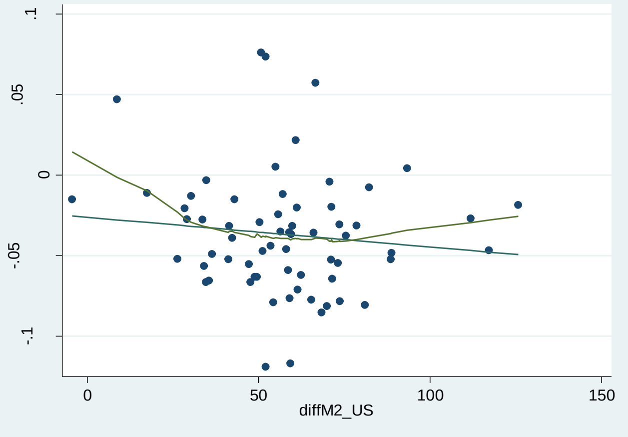

.2
.1
0
-.1
-2
-1
0
1
2
diffFANG_MTBV


.15
.1
.05
0
-.05
-.1
-1
-.5
0
.5
1
diffFANG_EPS

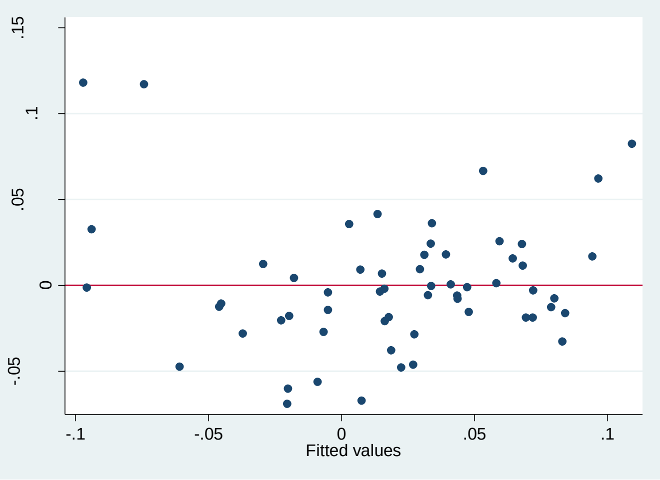

.15
.1
.05
0
-.05
-.1
-15
-10
-5
0
5
10
diffOIL


.15
.1
.05
0
-.05
-.1
-.2
0
.2
.4
.6
diffCPI_US

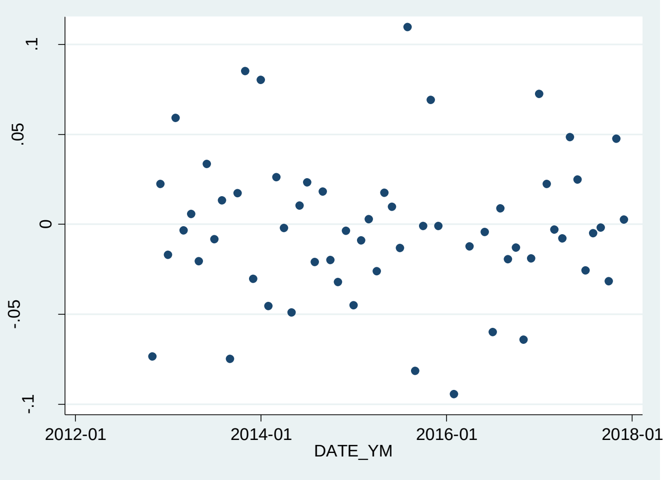

.15
.1
.05
0
-.05
-.1
-50
0
50
100
diffM1_US


.1
.05
0
-.05
-.1
0
50
100
150
diffM2_US

Homoskedasticity

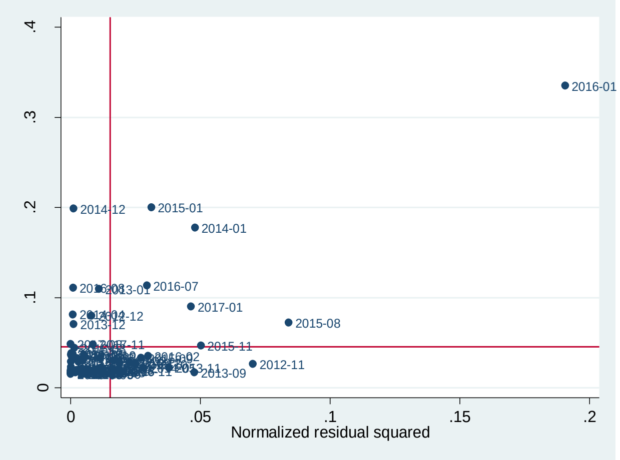


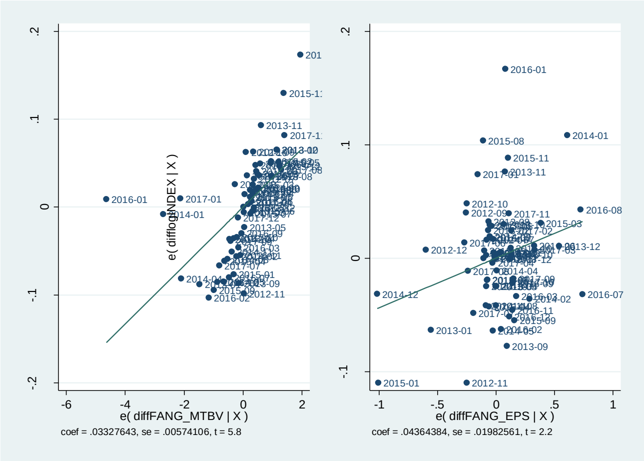
Independence of error term

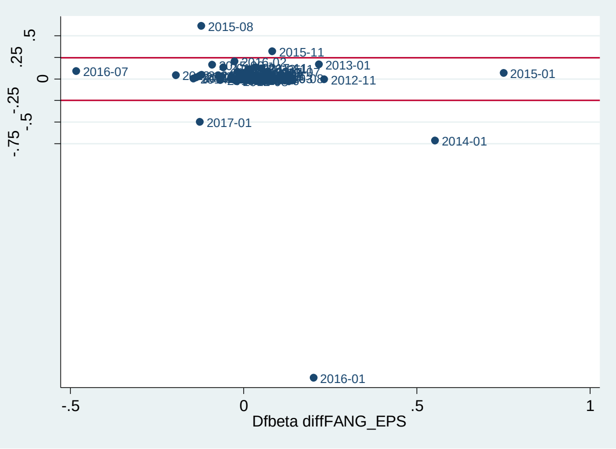


8.2 FANG model - final

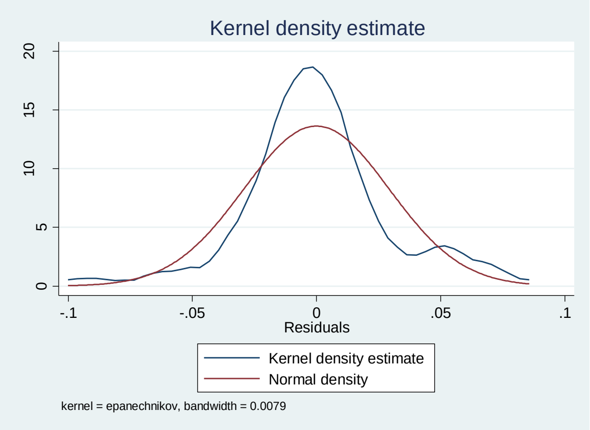
Unusual and Influential observations

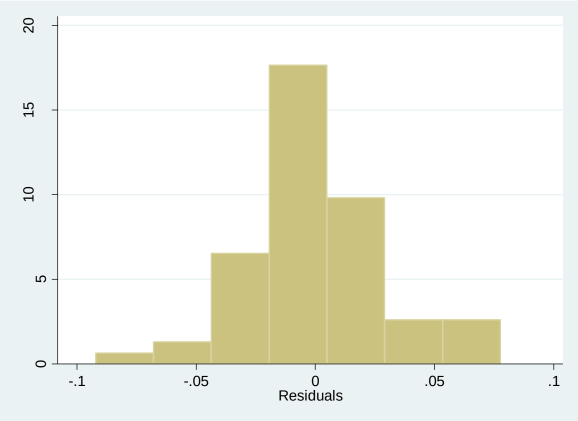


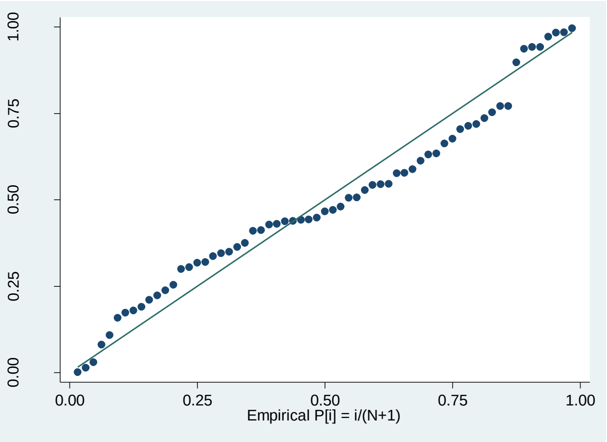


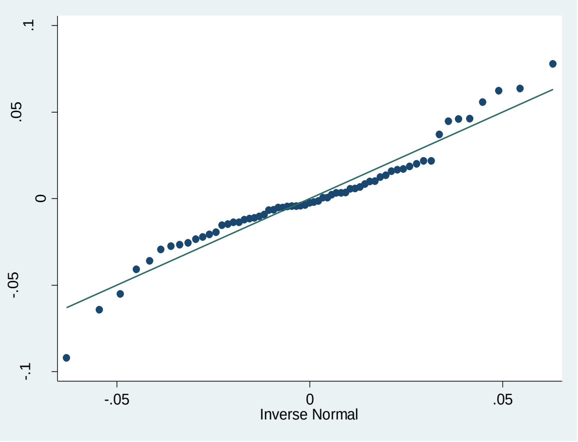


Normality of residuals

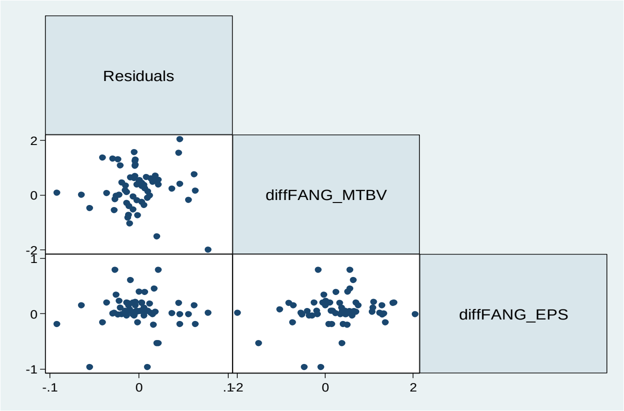


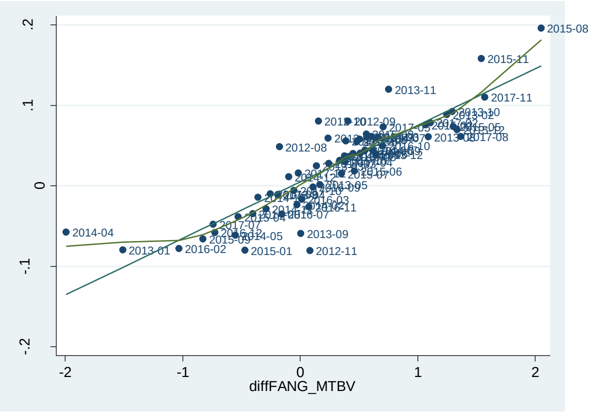


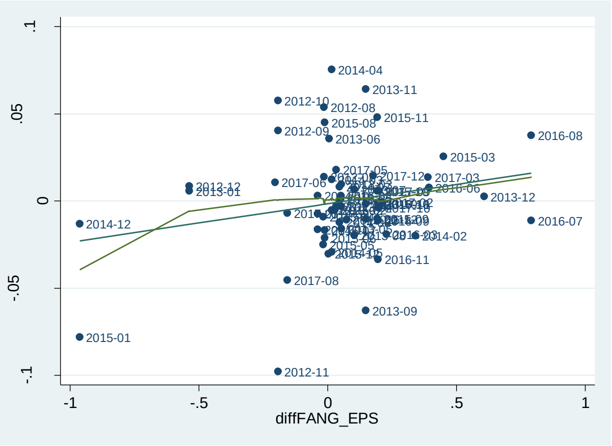

1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
Empirical P[i] = i/(N+1)


.1
.05
0
-.05
-.1
-.05
0
.05
Inverse Normal

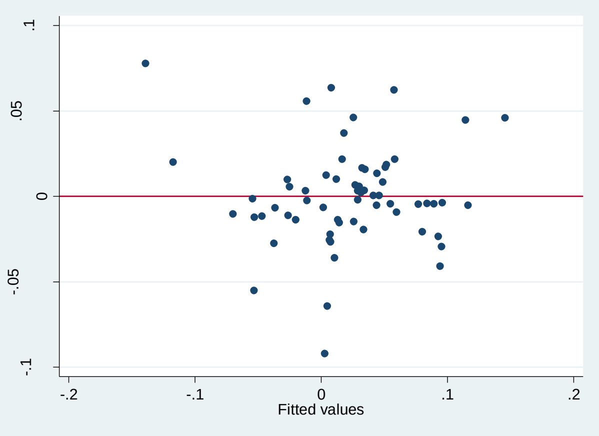
Linearity


Residuals
diffFANG_MTBV
diffFANG_EPS
2
0
-2
1
0
-1
-.1
0
.1
-2
0
2

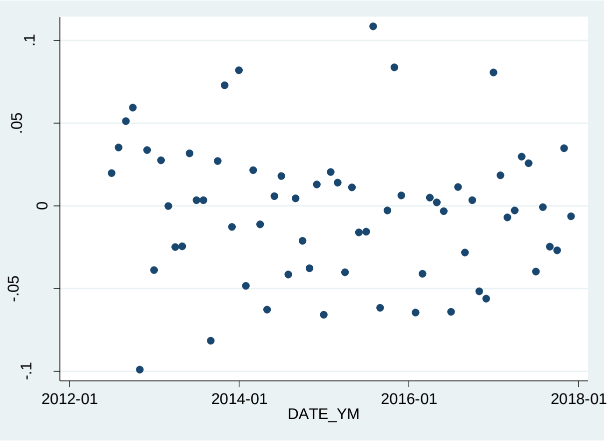

.2
.1
0
-.1
-.2
-2
-1
0
1
2
diffFANG_MTBV
2015-08
2015-11
2013-11
2017-11
2012-09
2013-10
2017-05
2012-08
2017-08
2016-10
2016-06
2016-03
2016-11
2016-07
2014-04
2017-07
2015-09
2014-05
2013-09
2013-01
2016-02
2015-01
2012-11

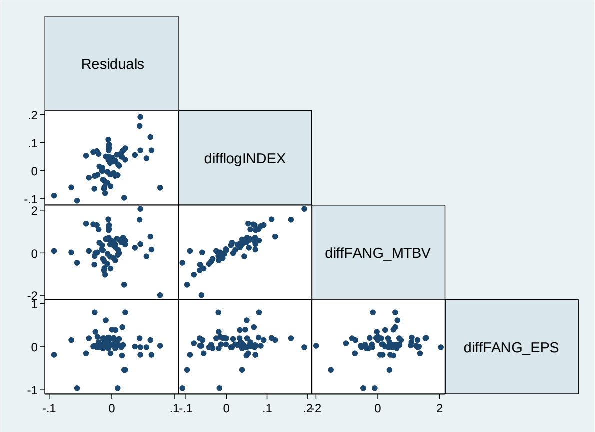

.1
.05
0
-.05
-.1
-1
-.5
0
.5
1
diffFANG_EPS
2014-04
2013-11
2012-10
2012-08
2015-08
2015-11
2012-09
2013-06
2016-08
2015-03
2017-05
2017-06
2017-12
2017-03
2012-12
2013-01
2016-06
2013-12
2014-12
2016-07
2014-02
2016-11
2015-05
2017-08
2013-09
2015-01
2012-11

Homoskedasticity



Independent error term



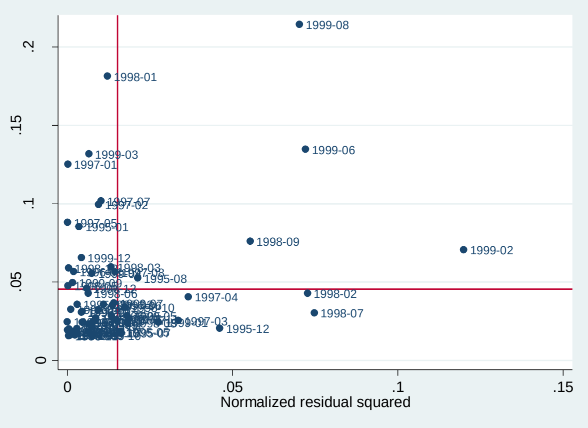


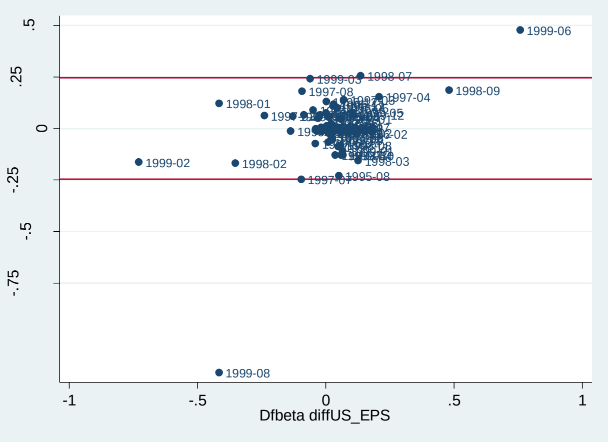
Standardized coefficients

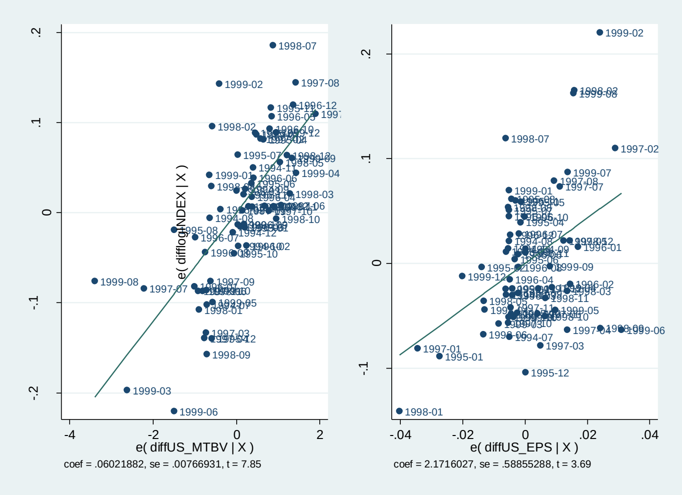
```
. reg difflogINDEX diffFANG_MTBV diffFANG_EPS, beta

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(2, 60)        =     99.54
       Model |  .176288707         2  .088144353   Prob > F        =    0.0000
    Residual |  .053128484        60  .000885475   R-squared       =    0.7684
-------------+----------------------------------   Adj R-squared   =    0.7607
       Total |  .229417191        62  .003700277   Root MSE        =    .02976

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
diffFANG_MTBV |   .0705838   .0052374    13.48   0.000                 .8513357
 diffFANG_EPS |    .021891   .0131068     1.67   0.100                 .1055057
        _cons |    .001217   .0040286     0.30   0.764                        .
------------------------------------------------------------------------------
```

8.3 The US model

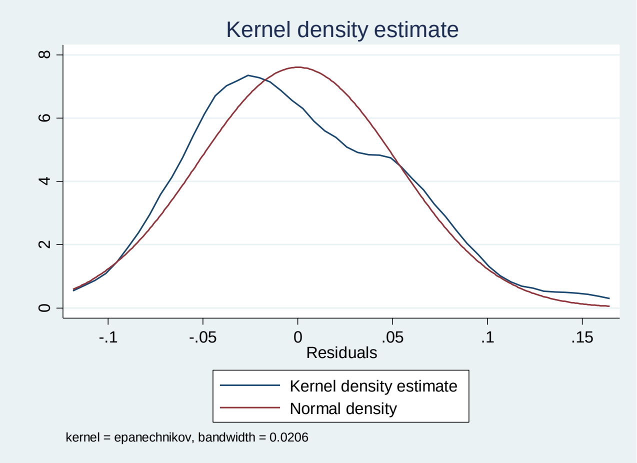
Unusual and influential data

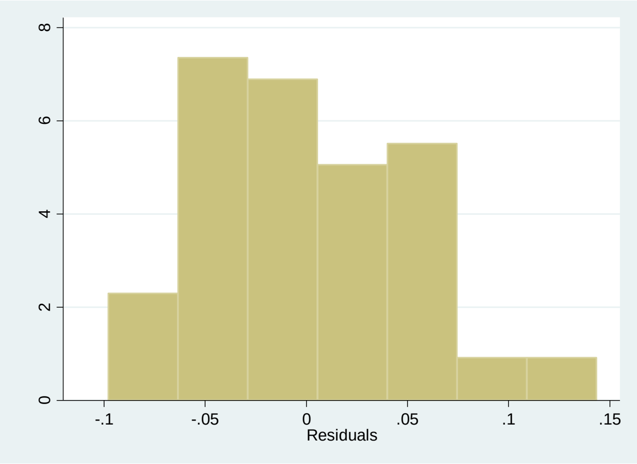


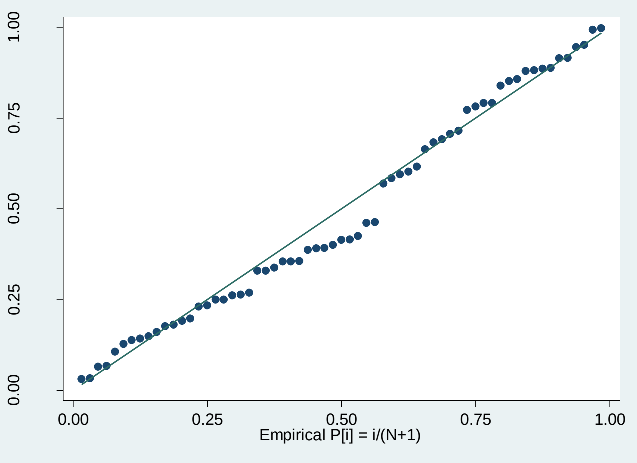


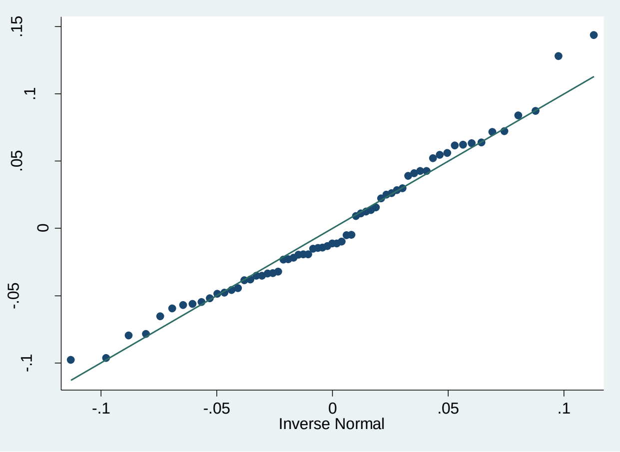


Normality of residuals

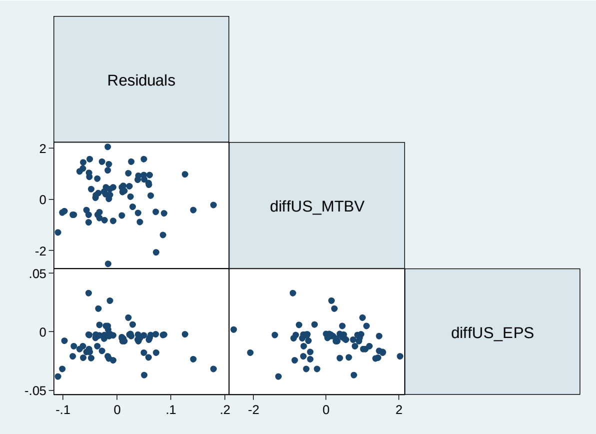


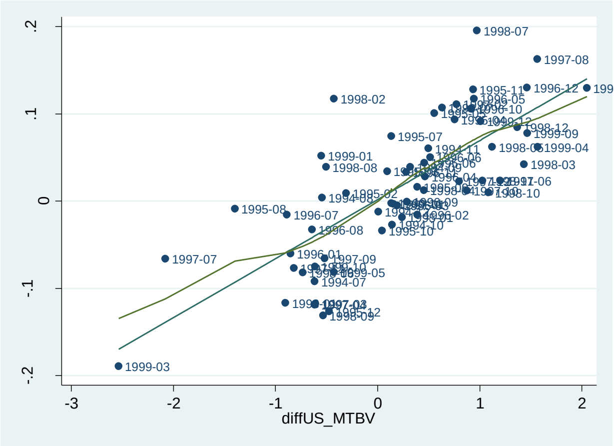


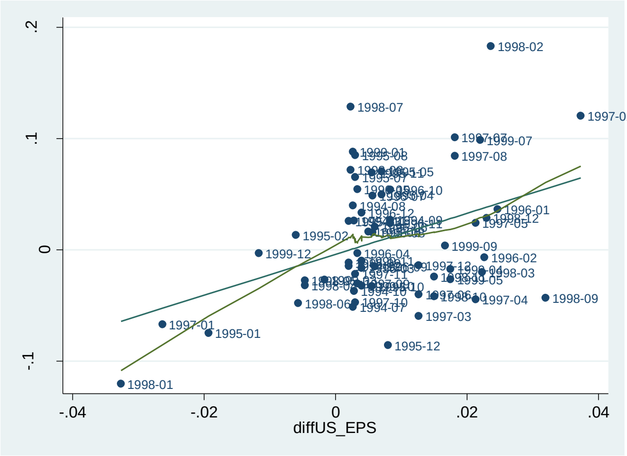




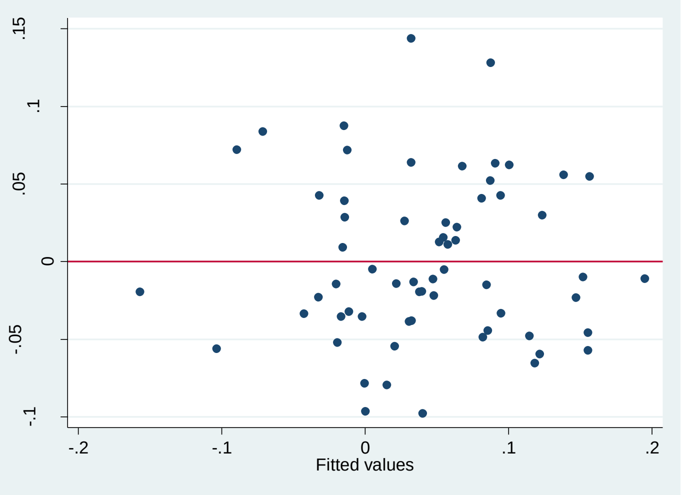
Linearity



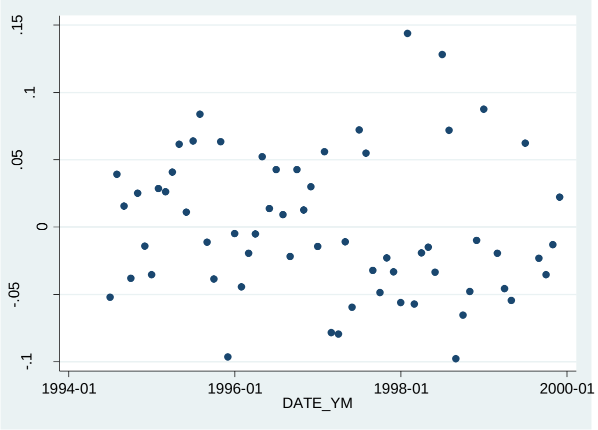


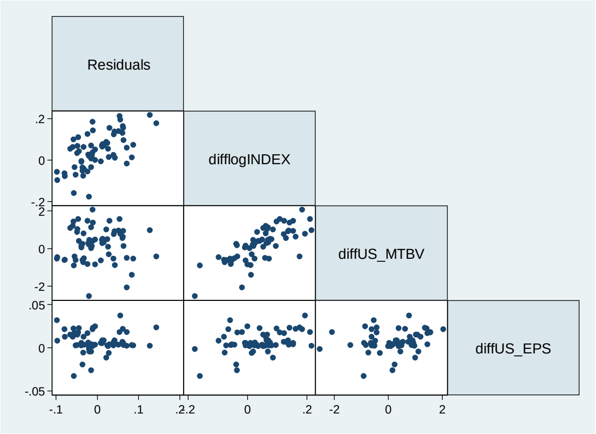


Homoskedasticity



Independence of error term



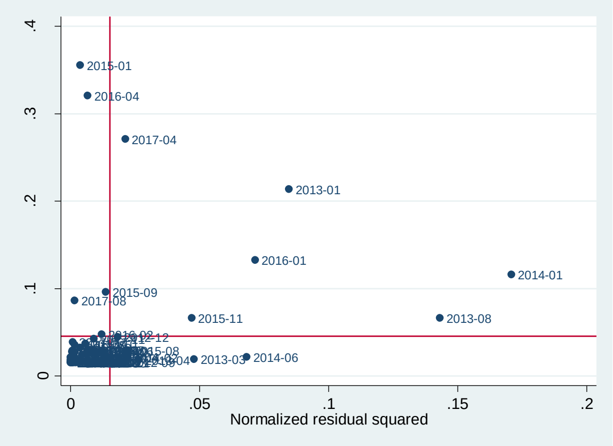


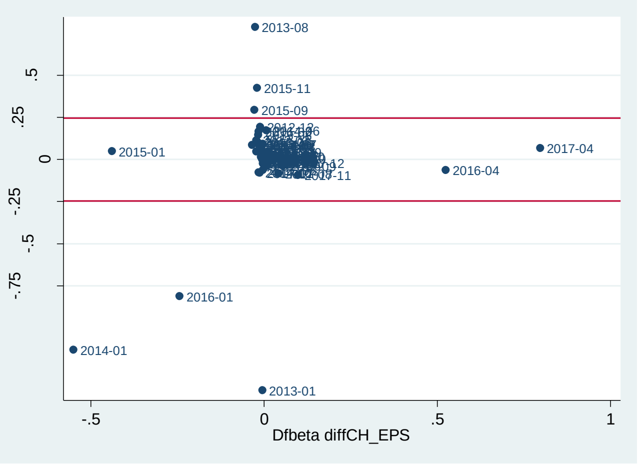
Standardized coefficients

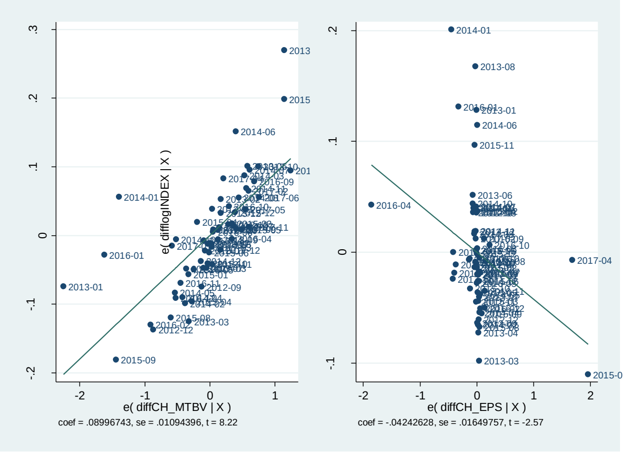
```
. reg difflogINDEX diffUS_MTBV diffUS_EPS, beta

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(2, 60)        =     51.31
       Model |  .291560791         2  .145780395   Prob > F        =    0.0000
    Residual |   .17047057        60  .002841176   R-squared       =    0.6310
-------------+----------------------------------   Adj R-squared   =    0.6187
       Total |  .462031361        62  .007452119   Root MSE        =     .0533

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
 diffUS_MTBV |   .0675358   .0076969     8.77   0.000                 .7020638
  diffUS_EPS |   1.838068   .5711137     3.22   0.002                 .2575117
       _cons |   .0176101   .0078108     2.25   0.028                        .
------------------------------------------------------------------------------
```

8.4 The CH model

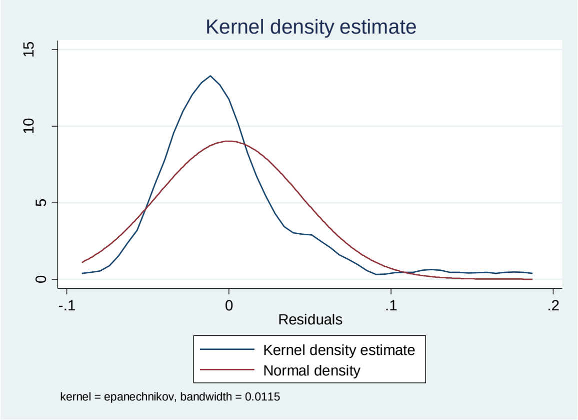
Unusual and influential data

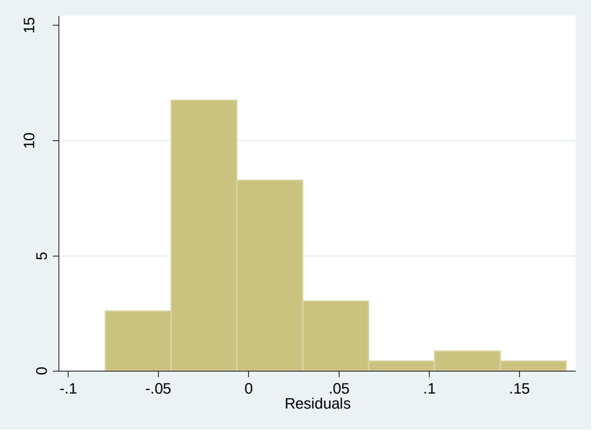


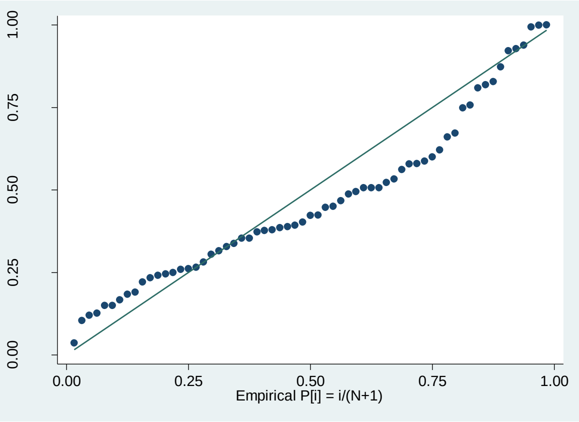


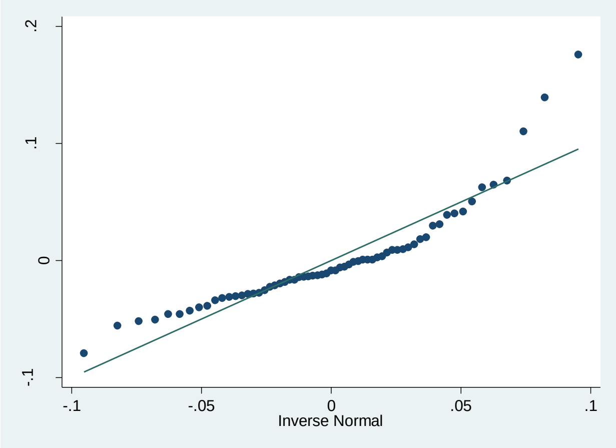


Normality of residuals

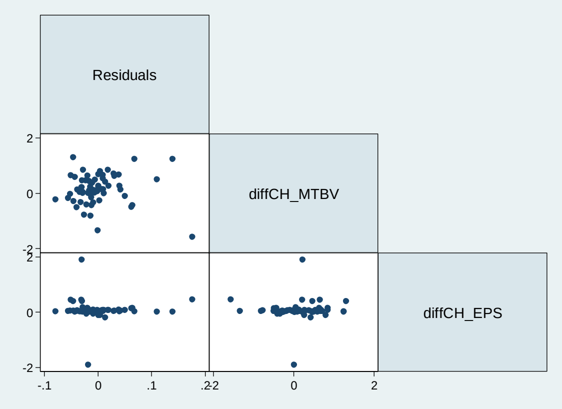


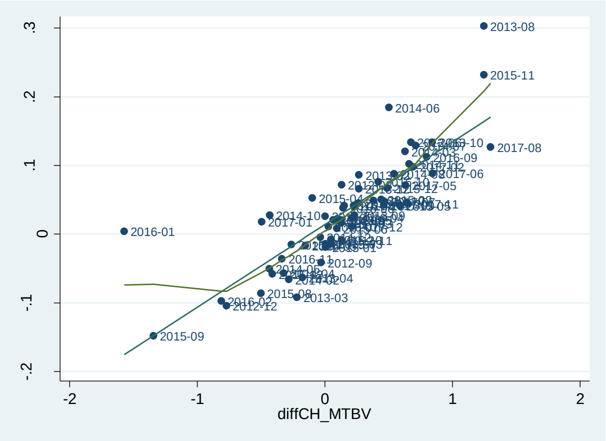


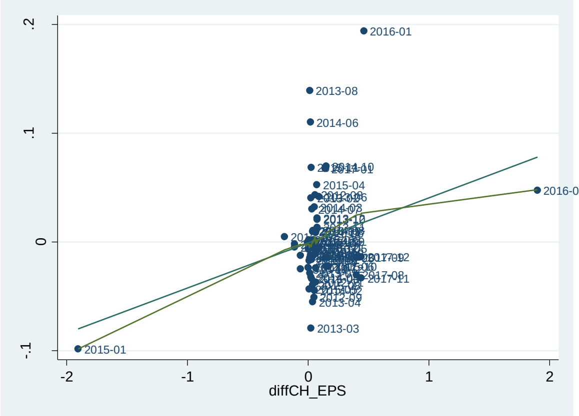

0.00
0.25
0.50
0.75
1.00
0.00
0.25
0.50
0.75
1.00
Empirical P[i] = i/(N+1)


-.1
0
.1
.2
-.1
-.05
0
.05
.1
Inverse Normal

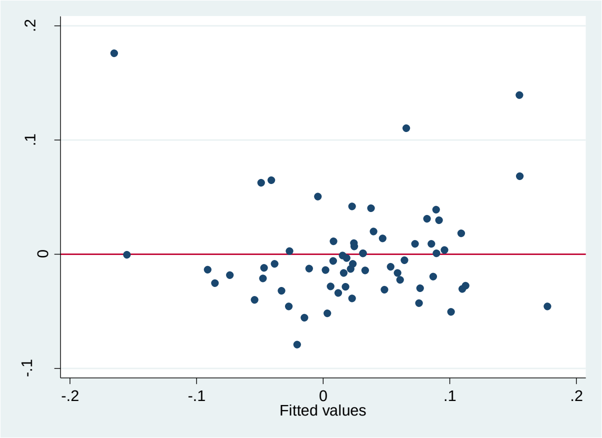
Linearity


Residuals
diffCH_MTBV
diffCH_EPS
2
0
-2
2
0
-2
-.1
0
.1
.2
-2
0
2

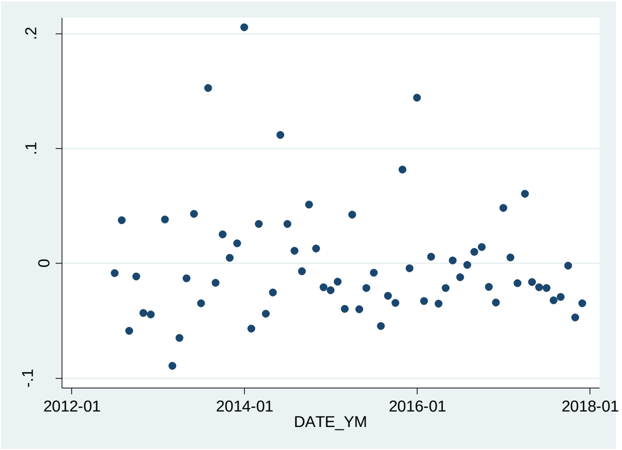

-.2
-.1
0
.1
.2
.3
-2
-1
0
1
2
diffCH_MTBV
2013-08
2015-11
2014-06
2017-08
2016-09
2017-06
2017-05
2015-04
2014-10
2017-01
2016-01
2016-11
2012-09
2015-08
2013-03
2016-02
2012-12
2015-09

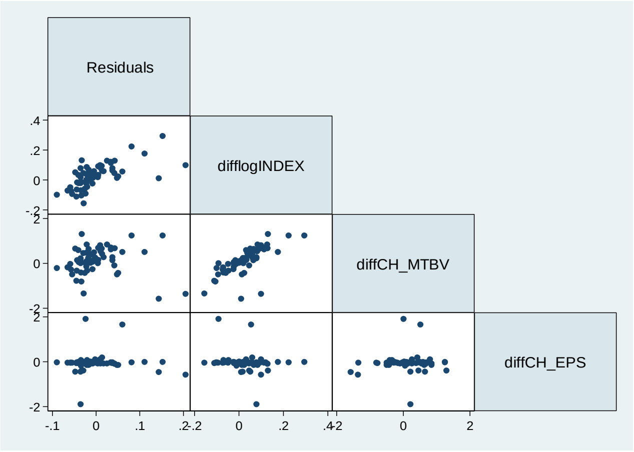

-.1
0
.1
.2
-2
-1
0
1
2
diffCH_EPS
2016-01
2013-08
2014-06
2015-04
2016-04
2013-04
2013-03
2015-01

Homoskedasticity

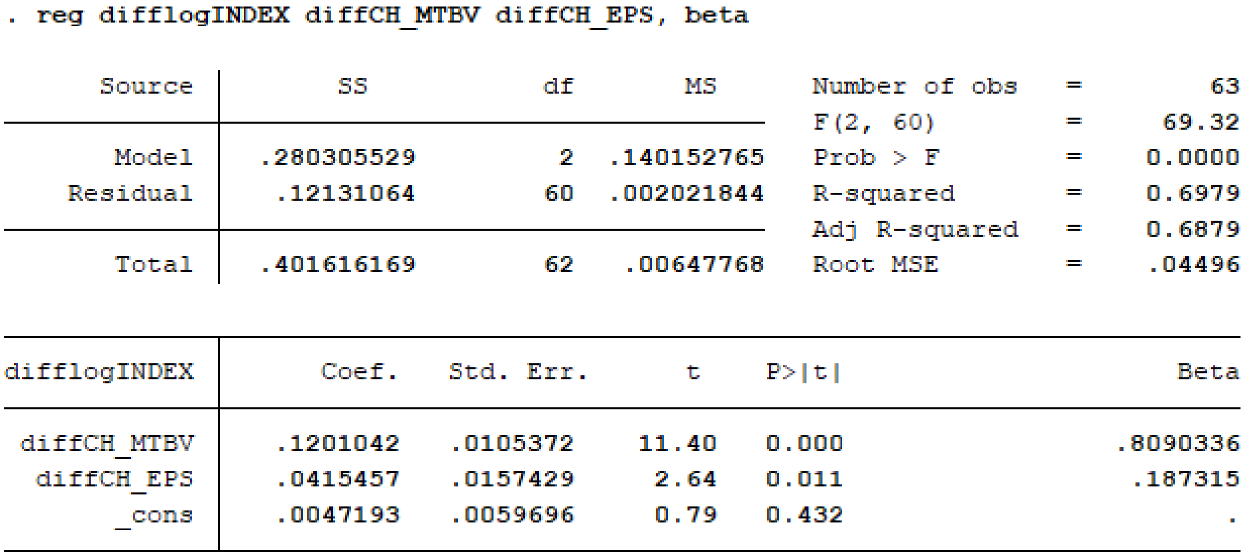


Independence of error term





Standardized coefficients



```
. reg difflogINDEX diffCH_MTBV diffCH_EPS, beta

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(2, 60)        =     69.32
       Model |  .280305529         2  .140152765   Prob > F        =    0.0000
    Residual |   .12131064        60  .002021844   R-squared       =    0.6979
-------------+----------------------------------   Adj R-squared   =    0.6879
       Total |  .401616169        62   .00647768   Root MSE        =    .04496

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
 diffCH_MTBV |   .1201042   .0105372    11.40   0.000                 .8090336
  diffCH_EPS |   .0415457   .0157429     2.64   0.011                  .187315
       _cons |   .0047193   .0059696     0.79   0.432                        .
------------------------------------------------------------------------------
```

8.5 Comparison of models

Comparison of coefficients across samples can be done by the use of a Z-test as discussed by Paternoster et al, 1998 and Clogg et al, 1995.

The calculation of the Z-score can be done by the following formula:

$Z = \frac{\beta_1 - \beta_2}{\sqrt{SE\beta_1{}^2 + SE\beta_2{}^2}}$, where β refers to the coefficient and Seβ refers to the standard error of the coefficient.

FANG vs US

MTBV

$$Z = \frac{\beta_1 - \beta_2}{\sqrt{SE\beta_1{}^2 + SE\beta_2{}^2}} = \frac{0{,}0706 - 0{,}0675}{\sqrt{0{,}0052^2 + 0{,}0077^2}} = 0{,}3336$$

FANG vs CHINA

MTBV

$$Z = \frac{\beta_1 - \beta_2}{\sqrt{SE\beta_1{}^2 + SE\beta_2{}^2}} = \frac{0{,}0706 - 0{,}1201}{\sqrt{0{,}0052^2 + 0{,}0105^2}} = -4{,}2246$$

8.6 Log files

Please see the following pages for the three log files from STATA

1. FANG
2. US
3. CH

```
      name:  <unnamed>
       log:  C:\Users\rinem\Dropbox\IB\master thesis\STATA\FANG_log.smcl
  log type:  smcl
 opened on:  27 Apr 2018, 12:55:03

 1 .
 2 .
 3 . //import excel data
 4 . import exc STATA_FANG.xlsx, firstrow clear

 5 .
 6 . //Creates date variable with YYYY/MM
 7 . generate int DATE_YM = mofd(DATE)

 8 . format DATE_YM %tmCY-N

 9 . drop DATE

10 . tsset DATE_YM
          time variable:  DATE_YM, 2012-06 to 2017-12
                  delta:  1 month

11 .
12 . //overview of the data
13 . sum INDEX FANG_MTBV FANG_EPS OIL CPI_US M1_US M2_US

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       INDEX |         67    219.0015    91.37141        100   431.0611
   FANG_MTBV |         67    9.387587    2.274101   5.491667     14.744
    FANG_EPS |         67    6.533965    1.662258   4.305333      9.768
         OIL |         67    69.47269    24.38564      31.42     107.79
      CPI_US |         67    241.4848    7.403498    229.623    254.398
-------------+---------------------------------------------------------
       M1_US |         67    2960.431    381.0333     2276.2     3603.9
       M2_US |         67    11909.92    1151.446     9973.5    13833.2

14 . graph matrix INDEX FANG_MTBV FANG_EPS OIL CPI_US M1_US M2_US, half

15 .
16 . //checking distribution of all data
17 . kdensity INDEX, nor

18 . histogram INDEX
   (bin=8, start=100, width=41.382641)

19 . kdensity FANG_MTBV, nor

20 . histogram FANG_MTBV
   (bin=8, start=5.4916667, width=1.1565417)

21 . kdensity FANG_EPS, nor

22 . histogram FANG_EPS
   (bin=8, start=4.3053333, width=.68283333)

23 . kdensity OIL, nor

24 . histogram OIL
   (bin=8, start=31.42, width=9.54625)
```

```
25 . kdensity CPI_US, nor

26 . histogram CPI_US
   (bin=8, start=229.623, width=3.096875)

27 . kdensity M1_US, nor

28 . histogram M1_US
   (bin=8, start=2276.2, width=165.9625)

29 . kdensity M2_US, nor

30 . histogram M2_US
   (bin=8, start=9973.5, width=482.4625)

31 .
32 . /*From above it seems like INDEX should be log transformed, which also would
   > make more sense when we have to interpret the model*/
33 .
34 . //generating logINDEX
35 . gen logINDEX = log(INDEX)

36 . gen logFANG_EPS = log(FANG_EPS)

37 .
38 . kdensity logFANG_EPS, nor

39 . histogram logFANG_EPS
   (bin=8, start=1.4598546, width=.10240713)

40 . kdensity logINDEX, nor

41 . histogram logINDEX
   (bin=8, start=4.6051702, width=.18263495)

42 .
43 . /*As our data are time series we suspect that they might be non-stationary. Hence, we start o
   > by testing that*/
44 .
45 . *************
46 . **Stationarity
47 . *************
48 . /*dfuller H0:random walk with a possible drift, Ha: stationarity around a linear time trend
   > p-value >0,05 => fail to reject H0*/
49 . dfuller logINDEX, trend

   Dickey-Fuller test for unit root                   Number of obs   =        66

                                ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(t)             -2.731            -4.115            -3.484            -3.170
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.2234

50 . dfuller FANG_MTBV, trend

   Dickey-Fuller test for unit root                   Number of obs   =        66

                                ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(t)             -3.254            -4.115            -3.484            -3.170
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.0741
```

```
51 . dfuller FANG_EPS, trend

   Dickey-Fuller test for unit root                   Number of obs   =        66

                                  ────────── Interpolated Dickey-Fuller ─────────
                     Test         1% Critical       5% Critical      10% Critical
                  Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(t)             -1.828            -4.115            -3.484            -3.170
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.6911

52 . dfuller OIL, trend

   Dickey-Fuller test for unit root                   Number of obs   =        66

                                  ────────── Interpolated Dickey-Fuller ─────────
                     Test         1% Critical       5% Critical      10% Critical
                  Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(t)             -1.411            -4.115            -3.484            -3.170
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.8576

53 . dfuller CPI_US, trend

   Dickey-Fuller test for unit root                   Number of obs   =        66

                                  ────────── Interpolated Dickey-Fuller ─────────
                     Test         1% Critical       5% Critical      10% Critical
                  Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(t)             -1.651            -4.115            -3.484            -3.170
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.7717

54 . dfuller M1_US, trend

   Dickey-Fuller test for unit root                   Number of obs   =        66

                                  ────────── Interpolated Dickey-Fuller ─────────
                     Test         1% Critical       5% Critical      10% Critical
                  Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(t)             -3.813            -4.115            -3.484            -3.170
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0159

55 . dfuller M2_US, trend

   Dickey-Fuller test for unit root                   Number of obs   =        66

                                  ────────── Interpolated Dickey-Fuller ─────────
                     Test         1% Critical       5% Critical      10% Critical
                  Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(t)             -1.666            -4.115            -3.484            -3.170
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.7657
```

```
56 .
57 . pperron logINDEX, trend

   Phillips-Perron test for unit root                Number of obs   =        66
                                                     Newey-West lags =         3

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)       -15.031           -26.244           -20.088           -17.024
    Z(t)          -2.789            -4.115            -3.484            -3.170
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.2009

58 . pperron FANG_MTBV, trend

   Phillips-Perron test for unit root                Number of obs   =        66
                                                     Newey-West lags =         3

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)       -20.881           -26.244           -20.088           -17.024
    Z(t)          -3.359            -4.115            -3.484            -3.170
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.0571

59 . pperron FANG_EPS, trend

   Phillips-Perron test for unit root                Number of obs   =        66
                                                     Newey-West lags =         3

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)        -6.713           -26.244           -20.088           -17.024
    Z(t)          -1.948            -4.115            -3.484            -3.170
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.6294

60 . pperron OIL, trend

   Phillips-Perron test for unit root                Number of obs   =        66
                                                     Newey-West lags =         3

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)        -6.287           -26.244           -20.088           -17.024
    Z(t)          -1.722            -4.115            -3.484            -3.170
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.7411

61 . pperron CPI_US, trend

   Phillips-Perron test for unit root                Number of obs   =        66
                                                     Newey-West lags =         3

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)        -5.213           -26.244           -20.088           -17.024
    Z(t)          -1.748            -4.115            -3.484            -3.170
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.7294
```

```
62 . pperron M1_US, trend

   Phillips-Perron test for unit root                Number of obs   =        66
                                                     Newey-West lags =         3

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)          -24.378           -26.244           -20.088           -17.024
    Z(t)             -3.865            -4.115            -3.484            -3.170
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.0135

63 . pperron M2_US, trend

   Phillips-Perron test for unit root                Number of obs   =        66
                                                     Newey-West lags =         3

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)           -6.330           -26.244           -20.088           -17.024
    Z(t)             -1.766            -4.115            -3.484            -3.170
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.7209

64 .
65 . /*both dfuller and pperron shows that all variables except M1 fail to reject H0. As
   > the data for M1 is of similar nature to the otherdata, time series, we treat all
   > variables in the same way*/
66 .
67 . //Tests for cointegration
68 . egranger logINDEX FANG_MTBV FANG_EPS OIL CPI_US M1_US M2_US //We are unable to reject H0: var

   Engle-Granger test for cointegration                 N (1st step)  =        67
                                                        N (test)      =        66
   ------------------------------------------------------------------------------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(t)             -3.494            -5.976            -5.283            -4.936

   Critical values from MacKinnon (1990, 2010)

69 . vecrank logINDEX FANG_MTBV FANG_EPS OIL CPI_US M1_US M2_US //number of cointegrating equation

                       Johansen tests for cointegration
   Trend: constant                                         Number of obs =      65
   Sample:  2012-08 - 2017-12                                       Lags =       2
   -------------------------------------------------------------------------------
                                                                 5%
   maximum                                             trace    critical
     rank    parms       LL           eigenvalue  statistic     value
      0       56     -736.24631           .        128.7003    124.24
      1       69     -718.81961       0.41504       93.8469*    94.15
      2       80     -703.81818       0.36971       63.8441     68.52
      3       89     -691.36304       0.31835       38.9338     47.21
      4       96     -682.61817       0.23591       21.4441     29.68
      5       101    -676.35455       0.17529        8.9168     15.41
      6       104    -671.95213       0.12669        0.1120      3.76
      7       105    -671.89614       0.00172
   -------------------------------------------------------------------------------
```

```
70 . //vacrank finds one cointegration equation. Need to test for the variable that we end up havi
71 .
72 . //creates differenced variables
73 . gen difflogINDEX = D.logINDEX
   (1 missing value generated)

74 . gen diffFANG_MTBV = D.FANG_MTBV
   (1 missing value generated)

75 . //gen diffFANG_EPS = D.FANG_EPS Manually created to avoid 0 values
76 . gen diffOIL = D.OIL
   (1 missing value generated)

77 . gen diffCPI_US = D.CPI_US
   (1 missing value generated)

78 . gen diffM1_US = D.M1_US
   (1 missing value generated)

79 . gen diffM2_US = D.M2_US
   (1 missing value generated)

80 .
81 . drop in 1/1 //drops the first observation which contains 0 values for the differenced variabl
   (1 observation deleted)

82 .
83 . //tests differenced variables
84 . dfuller difflogINDEX, trend

   Dickey-Fuller test for unit root                   Number of obs   =        65

                               ────────── Interpolated Dickey-Fuller ─────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ───────────────────────────────────────────────────────────────────────────
    Z(t)            -8.975            -4.117            -3.485            -3.171
   ───────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000

85 . dfuller diffFANG_MTBV, trend

   Dickey-Fuller test for unit root                   Number of obs   =        65

                               ────────── Interpolated Dickey-Fuller ─────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ───────────────────────────────────────────────────────────────────────────
    Z(t)            -8.485            -4.117            -3.485            -3.171
   ───────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000

86 . dfuller diffFANG_EPS, trend

   Dickey-Fuller test for unit root                   Number of obs   =        65

                               ────────── Interpolated Dickey-Fuller ─────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ───────────────────────────────────────────────────────────────────────────
    Z(t)            -4.026            -4.117            -3.485            -3.171
   ───────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0081
```

```
87 . dfuller diffOIL, trend

   Dickey-Fuller test for unit root                   Number of obs   =        65

                               ─────────── Interpolated Dickey-Fuller ──────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(t)             -6.103            -4.117            -3.485            -3.171
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000

88 . dfuller diffCPI_US, trend

   Dickey-Fuller test for unit root                   Number of obs   =        65

                               ─────────── Interpolated Dickey-Fuller ──────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(t)             -6.399            -4.117            -3.485            -3.171
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000

89 . dfuller diffM1_US, trend

   Dickey-Fuller test for unit root                   Number of obs   =        65

                               ─────────── Interpolated Dickey-Fuller ──────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(t)            -11.257            -4.117            -3.485            -3.171
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000

90 . dfuller diffM2_US, trend

   Dickey-Fuller test for unit root                   Number of obs   =        65

                               ─────────── Interpolated Dickey-Fuller ──────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(t)             -8.812            -4.117            -3.485            -3.171
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000

91 .
92 . pperron difflogINDEX, trend

   Phillips-Perron test for unit root                 Number of obs   =        65
                                                      Newey-West lags =         3

                               ─────────── Interpolated Dickey-Fuller ──────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(rho)          -70.053           -26.210           -20.070           -17.010
    Z(t)             -9.043            -4.117            -3.485            -3.171
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000
```

```
93 . pperron diffFANG_MTBV, trend

   Phillips-Perron test for unit root                  Number of obs   =        65
                                                       Newey-West lags =         3

                              ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)        -64.466           -26.210           -20.070           -17.010
    Z(t)           -8.560            -4.117            -3.485            -3.171
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.0000

94 . pperron diffFANG_EPS, trend

   Phillips-Perron test for unit root                  Number of obs   =        65
                                                       Newey-West lags =         3

                              ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)        -25.538           -26.210           -20.070           -17.010
    Z(t)           -3.944            -4.117            -3.485            -3.171
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.0106

95 . pperron diffOIL, trend

   Phillips-Perron test for unit root                  Number of obs   =        65
                                                       Newey-West lags =         3

                              ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)        -48.033           -26.210           -20.070           -17.010
    Z(t)           -6.072            -4.117            -3.485            -3.171
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.0000

96 . pperron diffCPI_US, trend

   Phillips-Perron test for unit root                  Number of obs   =        65
                                                       Newey-West lags =         3

                              ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)        -48.836           -26.210           -20.070           -17.010
    Z(t)           -6.311            -4.117            -3.485            -3.171
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.0000

97 . pperron diffM1_US, trend

   Phillips-Perron test for unit root                  Number of obs   =        65
                                                       Newey-West lags =         3

                              ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)        -86.487           -26.210           -20.070           -17.010
    Z(t)          -11.320            -4.117            -3.485            -3.171
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.0000
```

```
 98 . pperron diffM2_US, trend

    Phillips-Perron test for unit root                Number of obs   =        65
                                                      Newey-West lags =         3

                               ---------- Interpolated Dickey-Fuller ----------
                   Test         1% Critical       5% Critical      10% Critical
                Statistic           Value             Value             Value
    ---------------------------------------------------------------------------
     Z(rho)       -78.282           -26.210           -20.070           -17.010
     Z(t)          -8.774            -4.117            -3.485            -3.171
    ---------------------------------------------------------------------------
    MacKinnon approximate p-value for Z(t) = 0.0000

 99 .
100 . //graphical examination of stationarity (mean reverting, constant variance)
101 . twoway (tsline difflogINDEX)

102 . twoway (tsline diffFANG_MTBV)

103 . twoway (tsline diffFANG_EPS)

104 . twoway (tsline diffOIL)

105 . twoway (tsline diffCPI_US)

106 . twoway (tsline diffM1_US)

107 . twoway (tsline diffM2_US)

108 .
109 . *************
110 . **Lags
111 . *************
112 .
113 . dfgls diffFANG_MTBV, maxlag(5)

    DF-GLS for diffFANG_MTBV                                 Number of obs =    60

                 DF-GLS tau      1% Critical       5% Critical      10% Critical
      [lags]     Test Statistic        Value             Value             Value
    ---------------------------------------------------------------------------

        5           -3.367           -3.709            -2.978            -2.692
        4           -3.966           -3.709            -3.018            -2.730
        3           -3.824           -3.709            -3.056            -2.764
        2           -4.612           -3.709            -3.089            -2.795
        1           -6.533           -3.709            -3.118            -2.821

    Opt Lag (Ng-Perron seq t) = 0 [use maxlag(0)]
    Min SC   =  .2305485 at lag  1 with RMSE  1.048159
    Min MAIC =   2.92419 at lag  3 with RMSE  1.046229

114 . dfgls diffFANG_EPS, maxlag(5)

    DF-GLS for diffFANG_EPS                                  Number of obs =    60

                 DF-GLS tau      1% Critical       5% Critical      10% Critical
      [lags]     Test Statistic        Value             Value             Value
    ---------------------------------------------------------------------------

        5           -2.613           -3.709            -2.978            -2.692
        4           -3.315           -3.709            -3.018            -2.730
        3           -3.345           -3.709            -3.056            -2.764
        2           -3.501           -3.709            -3.089            -2.795
        1           -5.371           -3.709            -3.118            -2.821

    Opt Lag (Ng-Perron seq t) =  2 with RMSE  .2102198
    Min SC   = -2.914486 at lag  2 with RMSE  .2102198
    Min MAIC = -2.273571 at lag  2 with RMSE  .2102198
```

```
115 . dfgls diffOIL, maxlag(5)

    DF-GLS for diffOIL                                       Number of obs =    60

                   DF-GLS tau      1% Critical       5% Critical       10% Critical
      [lags]     Test Statistic        Value             Value             Value
    ______________________________________________________________________________

        5             -2.737           -3.709            -2.978            -2.692
        4             -3.201           -3.709            -3.018            -2.730
        3             -3.846           -3.709            -3.056            -2.764
        2             -4.625           -3.709            -3.089            -2.795
        1             -4.221           -3.709            -3.118            -2.821

    Opt Lag (Ng-Perron seq t) =  2 with RMSE  4.994116
    Min SC    =  3.407574 at lag  1 with RMSE  5.132268
    Min MAIC =  4.304912 at lag  1 with RMSE  5.132268

116 . dfgls diffCPI_US, maxlag(5)

    DF-GLS for diffCPI_US                                    Number of obs =    60

                   DF-GLS tau      1% Critical       5% Critical       10% Critical
      [lags]     Test Statistic        Value             Value             Value
    ______________________________________________________________________________

        5             -2.722           -3.709            -2.978            -2.692
        4             -3.846           -3.709            -3.018            -2.730
        3             -4.044           -3.709            -3.056            -2.764
        2             -4.801           -3.709            -3.089            -2.795
        1             -5.276           -3.709            -3.118            -2.821

    Opt Lag (Ng-Perron seq t) = 0 [use maxlag(0)]
    Min SC    = -3.691202 at lag  1 with RMSE  .1475129
    Min MAIC =  -2.21951 at lag  1 with RMSE  .1475129

117 . dfgls diffM1_US, maxlag(5)

    DF-GLS for diffM1_US                                     Number of obs =    60

                   DF-GLS tau      1% Critical       5% Critical       10% Critical
      [lags]     Test Statistic        Value             Value             Value
    ______________________________________________________________________________

        5             -2.188           -3.709            -2.978            -2.692
        4             -2.218           -3.709            -3.018            -2.730
        3             -3.148           -3.709            -3.056            -2.764
        2             -3.691           -3.709            -3.089            -2.795
        1             -5.340           -3.709            -3.118            -2.821

    Opt Lag (Ng-Perron seq t) =  4 with RMSE  21.80282
    Min SC    =  6.396158 at lag  1 with RMSE  22.87031
    Min MAIC =  7.338451 at lag  4 with RMSE  21.80282

118 . dfgls diffM2_US, maxlag(5)

    DF-GLS for diffM2_US                                     Number of obs =    60

                   DF-GLS tau      1% Critical       5% Critical       10% Critical
      [lags]     Test Statistic        Value             Value             Value
    ______________________________________________________________________________

        5             -2.965           -3.709            -2.978            -2.692
        4             -2.880           -3.709            -3.018            -2.730
        3             -2.974           -3.709            -3.056            -2.764
        2             -3.612           -3.709            -3.089            -2.795
        1             -4.715           -3.709            -3.118            -2.821

    Opt Lag (Ng-Perron seq t) = 0 [use maxlag(0)]
    Min SC    =    6.4689 at lag  1 with RMSE  23.71744
    Min MAIC =  7.582867 at lag  3 with RMSE  23.51503
```

```
119 .
120 . /*from theory we would expect CPI, M1 and M2 to have same lag, if any. Maybe even oil.
    >
    >
    > Using theory I think we should argue that we use 0 lags. EMH = market changes
    > emmidiatly on new information. Despite this might not be 100% true there is no
    > argument for why micro factor lags should be incorporated. For macro one might
    > argue that it could take time to impact the stock market and/or the individual stock.*/
121 .
122 . *************
123 . **UNUSUAL AND INFLUENTIAL DATA
124 . *************
125 . reg difflogINDEX diffFANG_MTBV diffFANG_EPS L2.diffOIL diffCPI_US L4.diffM1_US diffM2_US //mo

          Source |       SS           df       MS      Number of obs   =        62
    -------------+----------------------------------   F(6, 55)        =      8.12
           Model |  .104926156         6  .017487693   Prob > F        =    0.0000
        Residual |  .118422235        55  .002153132   R-squared       =    0.4698
    -------------+----------------------------------   Adj R-squared   =    0.4119
           Total |  .223348391        61  .003661449   Root MSE        =     .0464

    ------------------------------------------------------------------------------
     difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
    --------------+----------------------------------------------------------------
    diffFANG_MTBV |   .0332015   .0060007     5.53   0.000     .0211758    .0452273
     diffFANG_EPS |   .0567559   .0204259     2.78   0.007     .0158214    .0976903
                  |
          diffOIL |
              L2. |  -.0010867   .0011335    -0.96   0.342    -.0033583     .001185
                  |
       diffCPI_US |   .0183867   .0399347     0.46   0.647    -.0616443    .0984176
                  |
        diffM1_US |
              L4. |   -.000441    .000269    -1.64   0.107    -.0009801    .0000981
                  |
        diffM2_US |   .0003091   .0002474     1.25   0.217    -.0001866    .0008048
            _cons |  -.0043757   .0233603    -0.19   0.852    -.0511907    .0424393
    ------------------------------------------------------------------------------

126 .
127 . //Identifying outliers
128 . predict r, rstudent //generate studentized residuals (used for finding outliers)
    (4 missing values generated)

129 . stem r

    Stem-and-leaf plot for r (Studentized residuals)

    r rounded to nearest multiple of .01
    plot in units of .01

     -2** | 28
     -1** | 90,71,67
     -1** | 44,40,18,11,02
     -0** | 71,70,69,60,59,57
     -0** | 47,46,44,44,42,40,30,29,28,21,18,18,12,10,09,08,08,07,05,04,02,02
      0** | 05,07,13,21,22,23,29,38,40,41,49
      0** | 50,55,55,58,74
      1** | 06,08,38
      1** | 61,71,97
      2** | 08
      2** | 59
      3** |
      3** |
      4** |
      4** | 74
```

```
130 . sort r

131 . list r DATE  if abs(r) > 2

          +-----------------------+
          |         r   DATE_YM   |
          |-----------------------|
       1. | -2.283872   2016-02   |
      60. |  2.082979   2014-01   |
      61. |   2.58779   2015-08   |
      62. |  4.743855   2016-01   |
      63. |         .   2012-07   |
          |-----------------------|
      64. |         .   2012-08   |
      65. |         .   2012-10   |
      66. |         .   2012-09   |
          +-----------------------+

132 .
133 . //Identifying leverage
134 . sort DATE

135 . predict lev, leverage
    (4 missing values generated)

136 . stem lev

    Stem-and-leaf plot for lev (Leverage)

    lev rounded to nearest multiple of .001
    plot in units of .001

      0** | 27,29,35
      0** | 41,46,51,53,53,54,55,56,56,56,57,58,59
      0** | 60,60,61,61,68,70,70,70,71,71,75
      0** | 83,83,86,88,88,88,90,90,91,94,96,98
      1** | 05,07,08,15,17
      1** | 20,37,38
      1** | 42,46,48,54
      1** | 70
      1** | 87,99
      2** | 08
      2** |
      2** | 40
      2** | 60,70
      2** | 85
      3** | 13
      3** | 21
      3** |
      3** |
      3** |
      4** | 14

137 .
138 . /*a point with leverage greater than (2k+2)/n should be carefully examined.
    > k is the number of predictors and n is the number of observations. */
139 . display (2*6+2)/62
    .22580645

140 . list lev DATE if lev > .22
```

```
     +---------------------+
     |       lev   DATE_YM |
     |---------------------|
  1. |         .   2012-07 |
  2. |         .   2012-08 |
  3. |         .   2012-09 |
  4. |         .   2012-10 |
  6. |  .2403511   2012-12 |
     |---------------------|
 19. |  .2697809   2014-01 |
 30. |   .321018   2014-12 |
 31. |  .3127477   2015-01 |
 33. |  .2603066   2015-03 |
 43. |  .4141394   2016-01 |
     |---------------------|
 57. |  .2849438   2017-03 |
     +---------------------+

141 . lvr2plot, mlabel(DATE)

142 . //plots the leverage versus squared residuals, red lines = avg
143 .
144 . //Overall measure of influence (Outliers+leverage)
145 . sort DATE

146 . predict d, cooksd //Conventional cutoff point = 4/n
    (4 missing values generated)

147 . list d DATE  if d > 4/62

     +---------------------+
     |         d   DATE_YM |
     |---------------------|
  1. |         .   2012-07 |
  2. |         .   2012-08 |
  3. |         .   2012-09 |
  4. |         .   2012-10 |
 19. |  .2158913   2014-01 |
     |---------------------|
 31. |  .0892156   2015-01 |
 38. |   .078793   2015-08 |
 43. |  1.633784   2016-01 |
 44. |  .1185128   2016-02 |
 55. |  .0667646   2017-01 |
     +---------------------+

148 . predict dfit, dfits  //Conventional cutoff point = 2*sqrt(k/n)
    (4 missing values generated)

149 . list dfit DATE  if dfit > 2*sqrt(6/62)

     +---------------------+
     |      dfit   DATE_YM |
     |---------------------|
  1. |         .   2012-07 |
  2. |         .   2012-08 |
  3. |         .   2012-09 |
  4. |         .   2012-10 |
 19. |  1.266088   2014-01 |
     |---------------------|
 38. |  .7801784   2015-08 |
 43. |   3.98848   2016-01 |
 55. |  .6955392   2017-01 |
     +---------------------+
```

```
150 . dfbeta //how much each coefficient changes by leaving out the observation
    (4 missing values generated)
                          _dfbeta_1: dfbeta(diffFANG_MTBV)
    (4 missing values generated)
                          _dfbeta_2: dfbeta(diffFANG_EPS)
    (4 missing values generated)
                          _dfbeta_3: dfbeta(L2.diffOIL)
    (4 missing values generated)
                          _dfbeta_4: dfbeta(diffCPI_US)
    (4 missing values generated)
                          _dfbeta_5: dfbeta(L4.diffM1_US)
    (4 missing values generated)
                          _dfbeta_6: dfbeta(diffM2_US)

151 . //Conventional cutoff value for dfbeta = 2/sqrt(n)
152 . display 2/sqrt(62)
    .25400025

153 . scatter _dfbeta_1 _dfbeta_2 _dfbeta_3 _dfbeta_4 _dfbeta_5 _dfbeta_6, ylabel(-.75(.25).5) ylin

154 . list _dfbeta_1 difflogINDEX diffFANG_MTBV diffFANG_EPS diffOIL diffCPI_US diffM1_US diffM2_US

          +-----------------------------------------------------------------------+
       1. | _dfbeta_1   difflog~X   diffFAN~V   diffFANG~S   diffOIL   diffCP~S |
          |         .    .0505795    .4816667   -.01383333       .11       .347 |
          |-----------------------------------------------------------------------|
          |            diffM1~S            |             diffM2~S            |
          |                40.8            |                 73.6            |
          +-----------------------------------------------------------------------+

          +-----------------------------------------------------------------------+
       2. | _dfbeta_1   difflog~X   diffFAN~V   diffFANG~S   diffOIL   diffCP~S |
          |         .    .0442247   -.1766667   -.01383333       5.3       .263 |
          |-----------------------------------------------------------------------|
          |            diffM1~S            |             diffM2~S            |
          |                30.1            |                 70.7            |
          +-----------------------------------------------------------------------+

          +-----------------------------------------------------------------------+
       3. | _dfbeta_1   difflog~X   diffFAN~V   diffFANG~S   diffOIL   diffCP~S |
          |         .    .0718651    .4073333   -.19266667      8.36       .426 |
          |-----------------------------------------------------------------------|
          |            diffM1~S            |             diffM2~S            |
          |                41.8            |                 82.3            |
          +-----------------------------------------------------------------------+

          +-----------------------------------------------------------------------+
       4. | _dfbeta_1   difflog~X   diffFAN~V   diffFANG~S   diffOIL   diffCP~S |
          |         .    .0715022        .156   -.19266667     -4.79       .365 |
          |-----------------------------------------------------------------------|
          |            diffM1~S            |             diffM2~S            |
          |                32.8            |                 60.8            |
          +-----------------------------------------------------------------------+

          +-----------------------------------------------------------------------+
      19. | _dfbeta_1   difflog~X   diffFAN~V   diffFANG~S   diffOIL   diffCP~S |
          | -.8710017    .0428643   -2.434333    .60766667      4.66       .224 |
          |-----------------------------------------------------------------------|
          |            diffM1~S            |             diffM2~S            |
          |                34.2            |                 50.7            |
          +-----------------------------------------------------------------------+

          +-----------------------------------------------------------------------+
      38. | _dfbeta_1   difflog~X   diffFAN~V   diffFANG~S   diffOIL   diffCP~S |
          |  .5397205    .1916103        2.05   -.01111111    -11.58       .245 |
          |-----------------------------------------------------------------------|
          |            diffM1~S            |             diffM2~S            |
          |                -6.4            |                 52.1            |
          +-----------------------------------------------------------------------+
```

```
     +------------------------------------------------------------------------+
 39. | _dfbeta_1 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |  .3016337 | -.0652614 | -.8266667 |  .19344444 |    -.31 |     .518 |
     |-----------------------------------+------------------------------------|
     |             diffM1~S              |              diffM2~S              |
     |                 15.6              |                  54.3              |
     +------------------------------------------------------------------------+

     +------------------------------------------------------------------------+
 41. | _dfbeta_1 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |  .3887396 |  .1590457 |  1.542667 |  .19344444 |    1.17 |     .507 |
     |-----------------------------------+------------------------------------|
     |             diffM1~S              |              diffM2~S              |
     |                 63.8              |                  93.4              |
     +------------------------------------------------------------------------+

     +------------------------------------------------------------------------+
 43. | _dfbeta_1 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     | -3.197604 |  .0296297 | -4.520333 |       .031 |   -4.38 |     .544 |
     |-----------------------------------+------------------------------------|
     |             diffM1~S              |              diffM2~S              |
     |                  9.6              |                 121.7              |
     +------------------------------------------------------------------------+

     +------------------------------------------------------------------------+
 44. | _dfbeta_1 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |  .3964604 |  -.080224 |    -1.031 |  .07366667 |   -5.78 |     .555 |
     |-----------------------------------+------------------------------------|
     |             diffM1~S              |              diffM2~S              |
     |                 27.8              |                  69.9              |
     +------------------------------------------------------------------------+

     +------------------------------------------------------------------------+
 55. | _dfbeta_1 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     | -.4199172 |  .0209394 |    -2.063 |      -.148 |    2.66 |     .665 |
     |-----------------------------------+------------------------------------|
     |             diffM1~S              |              diffM2~S              |
     |                 48.8              |                  66.6              |
     +------------------------------------------------------------------------+

155 . list _dfbeta_2 difflogINDEX diffFANG_MTBV diffFANG_EPS diffOIL diffCPI_US diffM1_US diffM2_US

     +------------------------------------------------------------------------+
  1. | _dfbeta_2 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |         . |  .0505795 |  .4816667 | -.01383333 |     .11 |     .347 |
     |-----------------------------------+------------------------------------|
     |             diffM1~S              |              diffM2~S              |
     |                 40.8              |                  73.6              |
     +------------------------------------------------------------------------+

     +------------------------------------------------------------------------+
  2. | _dfbeta_2 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |         . |  .0442247 | -.1766667 | -.01383333 |     5.3 |     .263 |
     |-----------------------------------+------------------------------------|
     |             diffM1~S              |              diffM2~S              |
     |                 30.1              |                  70.7              |
     +------------------------------------------------------------------------+

     +------------------------------------------------------------------------+
  3. | _dfbeta_2 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |         . |  .0718651 |  .4073333 | -.19266667 |    8.36 |     .426 |
     |-----------------------------------+------------------------------------|
     |             diffM1~S              |              diffM2~S              |
     |                 41.8              |                  82.3              |
     +------------------------------------------------------------------------+
```

```
     +----------------------------------------------------------------------+
  4. | _dfbeta_2 | difflog~X | diffFAN~V |  diffFANG~S | diffOIL | diffCP~S |
     |         . |  .0715022 |      .156 |  -.19266667 |   -4.79 |     .365 |
     |-----------------------------------+----------------------------------|
     |             diffM1~S              |             diffM2~S             |
     |                 32.8              |                 60.8             |
     +----------------------------------------------------------------------+

     +----------------------------------------------------------------------+
  5. | _dfbeta_2 | difflog~X | diffFAN~V |  diffFANG~S | diffOIL | diffCP~S |
     |  .2971583 | -.0892739 |      .083 |  -.19266667 |   -5.45 |     .306 |
     |-----------------------------------+----------------------------------|
     |             diffM1~S              |             diffM2~S             |
     |                  1.2              |                 59.3             |
     +----------------------------------------------------------------------+

     +----------------------------------------------------------------------+
 19. | _dfbeta_2 | difflog~X | diffFAN~V |  diffFANG~S | diffOIL | diffCP~S |
     |  .7535058 |  .0428643 | -2.434333 |   .60766667 |    4.66 |     .224 |
     |-----------------------------------+----------------------------------|
     |             diffM1~S              |             diffM2~S             |
     |                 34.2              |                 50.7             |
     +----------------------------------------------------------------------+

     +----------------------------------------------------------------------+
 31. | _dfbeta_2 | difflog~X | diffFAN~V |  diffFANG~S | diffOIL | diffCP~S |
     |  .5653062 | -.1081252 | -.4693333 |       -.963 |  -15.47 |     .273 |
     |-----------------------------------+----------------------------------|
     |             diffM1~S              |             diffM2~S             |
     |                    4              |                 61.4             |
     +----------------------------------------------------------------------+

     +----------------------------------------------------------------------+
 49. | _dfbeta_2 | difflog~X | diffFAN~V |  diffFANG~S | diffOIL | diffCP~S |
     | -.4170964 | -.0192909 | -.1556667 |        .791 |    -.21 |     .376 |
     |-----------------------------------+----------------------------------|
     |             diffM1~S              |             diffM2~S             |
     |                 -4.4              |                 59.1             |
     +----------------------------------------------------------------------+

156 . list _dfbeta_3 difflogINDEX diffFANG_MTBV diffFANG_EPS diffOIL diffCPI_US diffM1_US diffM2_US

     +----------------------------------------------------------------------+
  1. | _dfbeta_3 | difflog~X | diffFAN~V |  diffFANG~S | diffOIL | diffCP~S |
     |         . |  .0505795 |  .4816667 |  -.01383333 |     .11 |     .347 |
     |-----------------------------------+----------------------------------|
     |             diffM1~S              |             diffM2~S             |
     |                 40.8              |                 73.6             |
     +----------------------------------------------------------------------+

     +----------------------------------------------------------------------+
  2. | _dfbeta_3 | difflog~X | diffFAN~V |  diffFANG~S | diffOIL | diffCP~S |
     |         . |  .0442247 | -.1766667 |  -.01383333 |     5.3 |     .263 |
     |-----------------------------------+----------------------------------|
     |             diffM1~S              |             diffM2~S             |
     |                 30.1              |                 70.7             |
     +----------------------------------------------------------------------+

     +----------------------------------------------------------------------+
  3. | _dfbeta_3 | difflog~X | diffFAN~V |  diffFANG~S | diffOIL | diffCP~S |
     |         . |  .0718651 |  .4073333 |  -.19266667 |    8.36 |     .426 |
     |-----------------------------------+----------------------------------|
     |             diffM1~S              |             diffM2~S             |
     |                 41.8              |                 82.3             |
     +----------------------------------------------------------------------+
```

```
     +---------------------------------------------------------------------+
  4. | _dfbeta_3 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |         . |  .0715022 |      .156 | -.19266667 |   -4.79 |     .365 |
     |----------------------------------+----------------------------------|
     |            diffM1~S              |            diffM2~S              |
     |                32.8              |                60.8              |
     +---------------------------------------------------------------------+

     +---------------------------------------------------------------------+
  5. | _dfbeta_3 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     | -.4358326 | -.0892739 |      .083 | -.19266667 |   -5.45 |     .306 |
     |----------------------------------+----------------------------------|
     |            diffM1~S              |            diffM2~S              |
     |                 1.2              |                59.3              |
     +---------------------------------------------------------------------+

     +---------------------------------------------------------------------+
 19. | _dfbeta_3 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     | -.4779011 |  .0428643 | -2.434333 |  .60766667 |    4.66 |     .224 |
     |----------------------------------+----------------------------------|
     |            diffM1~S              |            diffM2~S              |
     |                34.2              |                50.7              |
     +---------------------------------------------------------------------+

     +---------------------------------------------------------------------+
 23. | _dfbeta_3 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |  -.267589 | -.0649729 |     -.552 |  .01555556 |    -.06 |     .471 |
     |----------------------------------+----------------------------------|
     |            diffM1~S              |            diffM2~S              |
     |                16.7              |                68.4              |
     +---------------------------------------------------------------------+

     +---------------------------------------------------------------------+
 31. | _dfbeta_3 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |  .2734603 | -.1081252 | -.4693333 |      -.963 |  -15.47 |     .273 |
     |----------------------------------+----------------------------------|
     |            diffM1~S              |            diffM2~S              |
     |                   4              |                61.4              |
     +---------------------------------------------------------------------+

     +---------------------------------------------------------------------+
 43. | _dfbeta_3 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |   .475663 |  .0296297 | -4.520333 |       .031 |   -4.38 |     .544 |
     |----------------------------------+----------------------------------|
     |            diffM1~S              |            diffM2~S              |
     |                 9.6              |               121.7              |
     +---------------------------------------------------------------------+

157 . list _dfbeta_4 difflogINDEX diffFANG_MTBV diffFANG_EPS diffOIL diffCPI_US diffM1_US diffM2_US

     +--------------------------------------------------------------------+
  1. | _dfbeta_4 | difflo~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |         . | .0505795 |  .4816667 | -.01383333 |     .11 |     .347 |
     |---------------------------------+----------------------------------|
     |            diffM1~S             |            diffM2~S              |
     |                40.8             |                73.6              |
     +--------------------------------------------------------------------+

     +--------------------------------------------------------------------+
  2. | _dfbeta_4 | difflo~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |         . | .0442247 | -.1766667 | -.01383333 |     5.3 |     .263 |
     |---------------------------------+----------------------------------|
     |            diffM1~S             |            diffM2~S              |
     |                30.1             |                70.7              |
     +--------------------------------------------------------------------+
```

```
     +-----------------------------------------------------------------+
  3. | _dfbeta_4 | difflo~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |         . | .0718651 |  .4073333 | -.19266667 |    8.36 |     .426 |
     |-----------------------------------+---------------------------------|
     |             diffM1~S              |             diffM2~S            |
     |                 41.8              |                 82.3            |
     +-----------------------------------------------------------------+

     +-----------------------------------------------------------------+
  4. | _dfbeta_4 | difflo~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |         . | .0715022 |      .156 | -.19266667 |   -4.79 |     .365 |
     |-----------------------------------+---------------------------------|
     |             diffM1~S              |             diffM2~S            |
     |                 32.8              |                 60.8            |
     +-----------------------------------------------------------------+

     +-----------------------------------------------------------------+
 19. | _dfbeta_4 | difflo~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     | -.5573393 | .0428643 | -2.434333 |  .60766667 |    4.66 |     .224 |
     |-----------------------------------+---------------------------------|
     |             diffM1~S              |             diffM2~S            |
     |                 34.2              |                 50.7            |
     +-----------------------------------------------------------------+

     +-----------------------------------------------------------------+
 43. | _dfbeta_4 | difflo~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |  .2992066 | .0296297 | -4.520333 |       .031 |   -4.38 |     .544 |
     |-----------------------------------+---------------------------------|
     |             diffM1~S              |             diffM2~S            |
     |                  9.6              |                121.7            |
     +-----------------------------------------------------------------+

     +-----------------------------------------------------------------+
 44. | _dfbeta_4 | difflo~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     | -.2705683 | -.080224 |    -1.031 |  .07366667 |   -5.78 |     .555 |
     |-----------------------------------+---------------------------------|
     |             diffM1~S              |             diffM2~S            |
     |                 27.8              |                 69.9            |
     +-----------------------------------------------------------------+

     +-----------------------------------------------------------------+
 55. | _dfbeta_4 | difflo~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |  .3759408 | .0209394 |    -2.063 |      -.148 |    2.66 |     .665 |
     |-----------------------------------+---------------------------------|
     |             diffM1~S              |             diffM2~S            |
     |                 48.8              |                 66.6            |
     +-----------------------------------------------------------------+

158 . list _dfbeta_5 difflogINDEX diffFANG_MTBV diffFANG_EPS diffOIL diffCPI_US diffM1_US diffM2_US

     +-------------------------------------------------------------------+
  1. | _dfbeta_5 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |         . |  .0505795 |  .4816667 | -.01383333 |     .11 |     .347 |
     |------------------------------------+----------------------------------|
     |              diffM1~S              |             diffM2~S             |
     |                  40.8              |                 73.6             |
     +-------------------------------------------------------------------+

     +-------------------------------------------------------------------+
  2. | _dfbeta_5 | difflog~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |         . |  .0442247 | -.1766667 | -.01383333 |     5.3 |     .263 |
     |------------------------------------+----------------------------------|
     |              diffM1~S              |             diffM2~S             |
     |                  30.1              |                 70.7             |
     +-------------------------------------------------------------------+
```

```
     +-------------------------------------------------------------------+
  3. | _dfbeta_5   difflog~X   diffFAN~V    diffFANG~S   diffOIL   diffCP~S |
     |         .    .0718651    .4073333   -.19266667      8.36       .426 |
     |-------------------------------------------------------------------|
     |              diffM1~S            |              diffM2~S           |
     |                  41.8            |                  82.3           |
     +-------------------------------------------------------------------+

     +-------------------------------------------------------------------+
  4. | _dfbeta_5   difflog~X   diffFAN~V    diffFANG~S   diffOIL   diffCP~S |
     |         .    .0715022        .156   -.19266667     -4.79       .365 |
     |-------------------------------------------------------------------|
     |              diffM1~S            |              diffM2~S           |
     |                  32.8            |                  60.8           |
     +-------------------------------------------------------------------+

     +-------------------------------------------------------------------+
 31. | _dfbeta_5   difflog~X   diffFAN~V    diffFANG~S   diffOIL   diffCP~S |
     | -.3452534   -.1081252   -.4693333         -.963    -15.47       .273 |
     |-------------------------------------------------------------------|
     |              diffM1~S            |              diffM2~S           |
     |                     4            |                  61.4           |
     +-------------------------------------------------------------------+

     +-------------------------------------------------------------------+
 39. | _dfbeta_5   difflog~X   diffFAN~V    diffFANG~S   diffOIL   diffCP~S |
     |  .4941571   -.0652614   -.8266667    .19344444      -.31       .518 |
     |-------------------------------------------------------------------|
     |              diffM1~S            |              diffM2~S           |
     |                  15.6            |                  54.3           |
     +-------------------------------------------------------------------+

     +-------------------------------------------------------------------+
 43. | _dfbeta_5   difflog~X   diffFAN~V    diffFANG~S   diffOIL   diffCP~S |
     | -.5849183    .0296297   -4.520333          .031     -4.38       .544 |
     |-------------------------------------------------------------------|
     |              diffM1~S            |              diffM2~S           |
     |                   9.6            |                 121.7           |
     +-------------------------------------------------------------------+

     +-------------------------------------------------------------------+
 44. | _dfbeta_5   difflog~X   diffFAN~V    diffFANG~S   diffOIL   diffCP~S |
     |  .7249437    -.080224      -1.031    .07366667     -5.78       .555 |
     |-------------------------------------------------------------------|
     |              diffM1~S            |              diffM2~S           |
     |                  27.8            |                  69.9           |
     +-------------------------------------------------------------------+

159 . list _dfbeta_6 difflogINDEX diffFANG_MTBV diffFANG_EPS diffOIL diffCPI_US diffM1_US diffM2_US

     +------------------------------------------------------------------+
  1. | _dfbeta_6   difflo~X   diffFAN~V    diffFANG~S   diffOIL   diffCP~S |
     |         .   .0505795    .4816667   -.01383333       .11       .347 |
     |------------------------------------------------------------------|
     |             diffM1~S            |              diffM2~S           |
     |                 40.8            |                  73.6           |
     +------------------------------------------------------------------+

     +------------------------------------------------------------------+
  2. | _dfbeta_6   difflo~X   diffFAN~V    diffFANG~S   diffOIL   diffCP~S |
     |         .   .0442247   -.1766667   -.01383333       5.3       .263 |
     |------------------------------------------------------------------|
     |             diffM1~S            |              diffM2~S           |
     |                 30.1            |                  70.7           |
     +------------------------------------------------------------------+
```

```
     +--------------------------------------------------------------------+
  3. | _dfbeta_6 | difflo~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |         . | .0718651 |  .4073333 | -.19266667 |    8.36 |     .426 |
     |--------------------------------------------------------------------|
     |               diffM1~S           |               diffM2~S          |
     |                   41.8           |                   82.3          |
     +--------------------------------------------------------------------+

     +--------------------------------------------------------------------+
  4. | _dfbeta_6 | difflo~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |         . | .0715022 |      .156 | -.19266667 |   -4.79 |     .365 |
     |--------------------------------------------------------------------|
     |               diffM1~S           |               diffM2~S          |
     |                   32.8           |                   60.8          |
     +--------------------------------------------------------------------+

     +--------------------------------------------------------------------+
  8. | _dfbeta_6 | difflo~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |  -.441387 | .0850744 |     1.247 |       .016 |    5.73 |      .34 |
     |--------------------------------------------------------------------|
     |               diffM1~S           |               diffM2~S          |
     |                   -2.1           |                   -4.4          |
     +--------------------------------------------------------------------+

     +--------------------------------------------------------------------+
 17. | _dfbeta_6 | difflo~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     | -.5135436 | .1198797 |      .754 |  .14866667 |   -7.33 |     .438 |
     |--------------------------------------------------------------------|
     |               diffM1~S           |               diffM2~S          |
     |                    2.5           |                    8.7          |
     +--------------------------------------------------------------------+

     +--------------------------------------------------------------------+
 41. | _dfbeta_6 | difflo~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |  .3831749 | .1590457 |  1.542667 |  .19344444 |    1.17 |     .507 |
     |--------------------------------------------------------------------|
     |               diffM1~S           |               diffM2~S          |
     |                   63.8           |                   93.4          |
     +--------------------------------------------------------------------+

     +--------------------------------------------------------------------+
 43. | _dfbeta_6 | difflo~X | diffFAN~V | diffFANG~S | diffOIL | diffCP~S |
     |  1.376199 | .0296297 | -4.520333 |       .031 |   -4.38 |     .544 |
     |--------------------------------------------------------------------|
     |               diffM1~S           |               diffM2~S          |
     |                    9.6           |                  121.7          |
     +--------------------------------------------------------------------+

160 .
161 . avplots, mlabel(DATE) //Partial regression plot, can be used for identifying influential obse

162 .
163 . drop in 43/43
    (1 observation deleted)

164 .
165 . *************
166 . **NORMALITY OF RESIDUALS
167 . *************
168 .
```

```
169 . predict res, resid
    (6 missing values generated)

170 . kdensity res, normal

171 . stem res

    Stem-and-leaf plot for res (Residuals)

    res rounded to nearest multiple of .001
    plot in units of .001

      -9* | 4
      -8* | 1
      -7* | 54
      -6* | 40
      -5* |
      -4* | 955
      -3* | 220
      -2* | 661100
      -1* | 97332
      -0* | 988544432211
       0* | 2369
       1* | 003778
       2* | 22356
       3* | 3
       4* | 78
       5* | 9
       6* | 9
       7* | 2
       8* | 05
       9* |
      10* | 9

172 . histogram res
    (bin=7, start=-.0943727, width=.02911596)

173 . pnorm res

174 . qnorm res

175 . sktest res

                       Skewness/Kurtosis tests for Normality
                                                            ------ joint ------
        Variable |        Obs  Pr(Skewness)  Pr(Kurtosis) adj chi2(2)   Prob>chi2
    -------------+---------------------------------------------------------------
             res |         59     0.3455         0.2771         2.16         0.3388

176 . swilk res //Shapiro-Wilk W test for normality.

                      Shapiro-Wilk W test for normal data

        Variable |        Obs       W           V         z       Prob>z
    -------------+------------------------------------------------------
             res |         59    0.97398      1.396     0.718    0.23638

177 . *The p-value  is based on the assumption that the distribution is normal.
178 . *high p-value => indicate that we cannot reject that res is normally distributed.
179 . sfrancia res

                     Shapiro-Francia W' test for normal data

        Variable |       Obs       W'          V'        z       Prob>z
    -------------+-----------------------------------------------------
             res |        59    0.97237      1.640     0.945    0.17224
```

```
180 .
181 . *************
182 . **LINEARITY
183 . *************
184 . graph matrix INDEX diffFANG_MTBV diffFANG_EPS L2.diffOIL diffCPI_US L4.diffM1_US diffM2_US, h

185 .
186 . reg difflogINDEX diffFANG_MTBV diffFANG_EPS diffOIL diffCPI_US diffM1_US diffM2_US

      Source |       SS           df       MS      Number of obs   =        65
-------------+----------------------------------   F(6, 58)        =     15.10
       Model |  .140158386         6  .023359731   Prob > F        =    0.0000
    Residual |  .089700564        58  .001546561   R-squared       =    0.6098
-------------+----------------------------------   Adj R-squared   =    0.5694
       Total |   .22985895        64  .003591546   Root MSE        =    .03933

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
diffFANG_MTBV |   .0502068   .0059044     8.50   0.000     .0383879    .0620258
 diffFANG_EPS |   .0361474   .0170068     2.13   0.038     .0021046    .0701903
      diffOIL |   .0009472   .0009751     0.97   0.335    -.0010045     .002899
   diffCPI_US |   .0077293   .0345217     0.22   0.824    -.0613734    .0768319
    diffM1_US |   .0005759   .0002664     2.16   0.035     .0000427    .0011092
    diffM2_US |  -.0002007   .0002624    -0.76   0.448     -.000726    .0003247
        _cons |    .007385   .0190104     0.39   0.699    -.0306685    .0454386
------------------------------------------------------------------------------

187 . //model without lags as acpr plot does not allow this
188 .
189 . acprplot diffFANG_MTBV, lowess lsopts(bwidth(1))

190 . acprplot diffFANG_EPS, lowess lsopts(bwidth(1))

191 . acprplot diffOIL, lowess lsopts(bwidth(1))

192 . acprplot diffCPI_US, lowess lsopts(bwidth(1))

193 . acprplot diffM1_US, lowess lsopts(bwidth(1))

194 . acprplot diffM2_US, lowess lsopts(bwidth(1))

195 .
196 . *************
197 . **HOMOSCEDASTICITY
198 . *************
199 . reg difflogINDEX diffFANG_MTBV diffFANG_EPS L2.diffOIL diffCPI_US L4.diffM1_US diffM2_US

      Source |       SS           df       MS      Number of obs   =        59
-------------+----------------------------------   F(6, 52)        =     14.56
       Model |  .139060132         6  .023176689   Prob > F        =    0.0000
    Residual |  .082780648        52  .001591936   R-squared       =    0.6268
-------------+----------------------------------   Adj R-squared   =    0.5838
       Total |  .221840779        58  .003824841   Root MSE        =     .0399

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
diffFANG_MTBV |   .0495907   .0062278     7.96   0.000     .0370937    .0620877
 diffFANG_EPS |    .054845   .0176374     3.11   0.003     .0194529    .0902371
             |
     diffOIL |
         L2. |  -.0016314   .0009902    -1.65   0.105    -.0036184    .0003556
             |
  diffCPI_US |   .0060121   .0349435     0.17   0.864    -.0641071    .0761314
             |
   diffM1_US |
         L4. |  -.0002651   .0002406    -1.10   0.276    -.0007478    .0002177
             |
   diffM2_US |   .0000317    .000223     0.14   0.887    -.0004157    .0004791
       _cons |   .0069373   .0203818     0.34   0.735    -.0339618    .0478363
------------------------------------------------------------------------------
```

```
200 . rvfplot, yline(0)

201 . estat imtest

    Cameron & Trivedi's decomposition of IM-test

    ---------------------------------------------------
                  Source |       chi2     df      p
    ---------------------+-----------------------------
      Heteroskedasticity |      42.72     27    0.0279
                Skewness |      23.04      6    0.0008
                Kurtosis |       2.97      1    0.0848
    ---------------------+-----------------------------
                   Total |      68.73     34    0.0004
    ---------------------------------------------------

202 . estat hettest

    Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
             Ho: Constant variance
             Variables: fitted values of difflogINDEX

             chi2(1)      =    11.54
             Prob > chi2  =   0.0007

203 .
204 . *************
205 . **MULTICOLLINIARITY
206 . *************
207 . estat vif //if high vif (above 10) try to omit some of the variables that are suffering from

        Variable |       VIF       1/VIF
    -------------+----------------------
    diffFANG_M~V |      1.11    0.903402
       diffM1_US |
             L4. |      1.09    0.918577
         diffOIL |
             L2. |      1.08    0.926616
    diffFANG_EPS |      1.08    0.929291
      diffCPI_US |      1.04    0.961706
       diffM2_US |      1.03    0.975185
    -------------+----------------------
        Mean VIF |      1.07

208 .
209 . *************
210 . **ISSUES OF INDEPENDENCE
211 . *************
212 . estat dwatson

    Number of gaps in sample:  3

    Durbin-Watson d-statistic(  7,    59) =  2.345548

213 . scatter res DATE_YM

214 .
215 . *************
216 . **INITIAL MODEL FOR EVALUATION
```

```
217 . *************
218 . reg difflogINDEX diffFANG_MTBV diffFANG_EPS L2.diffOIL diffCPI_US L4.diffM1_US diffM2_US

      Source |       SS           df       MS      Number of obs   =        59
-------------+----------------------------------   F(6, 52)        =     14.56
       Model |  .139060132         6  .023176689   Prob > F        =    0.0000
    Residual |  .082780648        52  .001591936   R-squared       =    0.6268
-------------+----------------------------------   Adj R-squared   =    0.5838
       Total |  .221840779        58  .003824841   Root MSE        =     .0399

------------------------------------------------------------------------------
 difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
diffFANG_MTBV |   .0495907   .0062278     7.96   0.000     .0370937    .0620877
 diffFANG_EPS |    .054845   .0176374     3.11   0.003     .0194529    .0902371
              |
      diffOIL |
          L2. |  -.0016314   .0009902    -1.65   0.105    -.0036184    .0003556
              |
   diffCPI_US |   .0060121   .0349435     0.17   0.864    -.0641071    .0761314
              |
    diffM1_US |
          L4. |  -.0002651   .0002406    -1.10   0.276    -.0007478    .0002177
              |
    diffM2_US |   .0000317    .000223     0.14   0.887    -.0004157    .0004791
        _cons |   .0069373   .0203818     0.34   0.735    -.0339618    .0478363
------------------------------------------------------------------------------

219 . reg difflogINDEX diffFANG_MTBV diffFANG_EPS L2.diffOIL diffCPI_US L4.diffM1_US

      Source |       SS           df       MS      Number of obs   =        59
-------------+----------------------------------   F(5, 53)        =     17.80
       Model |  .139027947         5  .027805589   Prob > F        =    0.0000
    Residual |  .082812832        53  .001562506   R-squared       =    0.6267
-------------+----------------------------------   Adj R-squared   =    0.5915
       Total |  .221840779        58  .003824841   Root MSE        =    .03953

------------------------------------------------------------------------------
 difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
diffFANG_MTBV |   .0495886     .00617     8.04   0.000     .0372133     .061964
 diffFANG_EPS |   .0547511   .0174614     3.14   0.003      .019728    .0897743
              |
      diffOIL |
          L2. |  -.0016262   .0009804    -1.66   0.103    -.0035925    .0003402
              |
   diffCPI_US |   .0057645    .034576     0.17   0.868    -.0635861    .0751151
              |
    diffM1_US |
          L4. |  -.0002696   .0002363    -1.14   0.259    -.0007435    .0002044
              |
        _cons |   .0089106   .0147881     0.60   0.549    -.0207506    .0385718
------------------------------------------------------------------------------

220 . reg difflogINDEX diffFANG_MTBV diffFANG_EPS L2.diffOIL

      Source |       SS           df       MS      Number of obs   =        62
-------------+----------------------------------   F(3, 58)        =     27.73
       Model |    .1338401         3  .044613367   Prob > F        =    0.0000
    Residual |  .093327385        58  .001609093   R-squared       =    0.5892
-------------+----------------------------------   Adj R-squared   =    0.5679
       Total |  .227167485        61  .003724057   Root MSE        =    .04011
```

```
------------------------------------------------------------------------------
 difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
diffFANG_MTBV |   .0513121   .0060214     8.52   0.000     .0392589    .0633653
 diffFANG_EPS |   .0461792   .0172469     2.68   0.010     .0116557    .0807027
              |
      diffOIL |
          L2. |  -.0012946   .0009717    -1.33   0.188    -.0032398    .0006505
              |
        _cons |   .0082333   .0053637     1.54   0.130    -.0025032    .0189698
------------------------------------------------------------------------------

221 . reg difflogINDEX diffFANG_MTBV diffFANG_EPS

      Source |       SS           df       MS      Number of obs   =        65
-------------+----------------------------------   F(2, 62)        =     40.51
       Model |  .130218672         2  .065109336   Prob > F        =    0.0000
    Residual |  .099640278        62  .001607101   R-squared       =    0.5665
-------------+----------------------------------   Adj R-squared   =    0.5525
       Total |   .22985895        64  .003591546   Root MSE        =    .04009

------------------------------------------------------------------------------
 difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
diffFANG_MTBV |   .0501983   .0059475     8.44   0.000     .0383093    .0620872
 diffFANG_EPS |   .0402229   .0168916     2.38   0.020      .006457    .0739887
        _cons |    .009837    .005181     1.90   0.062    -.0005196    .0201937
------------------------------------------------------------------------------

222 .
223 . *********************************************************
224 . ****************SPECIFYING FINAL MODEL*****************
225 . *********************************************************
226 . clear all

227 . cd "C:\Users\rinem\Dropbox\IB\master thesis\STATA"
    C:\Users\rinem\Dropbox\IB\master thesis\STATA

228 .
229 . //import excel data
230 . import exc STATA_FANG.xlsx, firstrow clear

231 .
232 . //Creates date variable with YYYY/MM
233 . generate int DATE_YM = mofd(DATE)

234 . format DATE_YM %tmCY-N

235 . drop DATE

236 . tsset DATE_YM
            time variable:  DATE_YM, 2012-06 to 2017-12
                    delta:  1 month

237 . gen logINDEX = log(INDEX)

238 . gen difflogINDEX = D.logINDEX
    (1 missing value generated)

239 . gen diffFANG_MTBV = D.FANG_MTBV
    (1 missing value generated)
```

```
240 . drop in 1/1
    (1 observation deleted)

241 .
242 . *****
243 . reg difflogINDEX diffFANG_MTBV diffFANG_EPS

          Source |       SS            df       MS      Number of obs   =        66
    -------------+----------------------------------   F(2, 63)        =     20.34
           Model |  .090211468         2  .045105734   Prob > F        =    0.0000
        Residual |  .139704478        63  .002217531   R-squared       =    0.3924
    -------------+----------------------------------   Adj R-squared   =    0.3731
           Total |  .229915946        65  .003537168   Root MSE        =    .04709

    ------------------------------------------------------------------------------
     difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
    -------------+----------------------------------------------------------------
    diffFANG_MTBV |   .0332764   .0057411     5.80   0.000     .0218038     .044749
     diffFANG_EPS |   .0436438   .0198256     2.20   0.031     .0040255    .0832622
            _cons |   .0154755   .0059396     2.61   0.011     .0036062    .0273448
    ------------------------------------------------------------------------------

244 . *****
245 .
246 . *************
247 . **UNUSUAL AND INFLUENTIAL DATA
248 . *************
249 . predict r, rstudent //generate studentized residuals (used for finding outliers)

250 . list r DATE difflogINDEX diffFANG_MTBV diffFANG_EPS  if abs(r) > 2

         +------------------------------------------------------------+
         |         r   DATE_YM   difflog~X   diffFAN~V    diffFANG~S |
         |------------------------------------------------------------|
      5. | -2.196368   2012-11   -.0892739        .083   -.19266667 |
     38. |  2.486108   2015-08    .1916103        2.05   -.01111111 |
     43. |  4.992943   2016-01    .0296297   -4.520333         .031 |
         +------------------------------------------------------------+

251 . predict lev, leverage

252 . list lev DATE difflogINDEX diffFANG_MTBV diffFANG_EPS if lev > .21

         +--------------------------------------------------------+
         |      lev   DATE_YM   difflo~X   diffFAN~V   diffFA~S |
         |--------------------------------------------------------|
     43. | .3350333   2016-01   .0296297   -4.520333       .031 |
         +--------------------------------------------------------+

253 . lvr2plot, mlabel(DATE)

254 . predict d, cooksd //Conventional cutoff point = 4/n

255 . list d DATE if d > 4/66

         +---------------------+
         |        d   DATE_YM |
         |---------------------|
     19. | .2643256   2014-01 |
     31. | .2044458   2015-01 |
     38. | .1479522   2015-08 |
     43. | 3.034265   2016-01 |
     49. | .0891057   2016-07 |
         |---------------------|
     55. | .1063152   2017-01 |
         +---------------------+
```

```
256 . predict dfit, dfits  //Conventional cutoff point = 2*sqrt(k/n)

257 . list dfit DATE if dfit > 2*sqrt(6/66)

         +--------------------+
         |     dfit   DATE_YM |
         |--------------------|
    19.  | .9103484   2014-01 |
    38.  | .6930777   2015-08 |
    43.  | 3.544057   2016-01 |
         +--------------------+

258 . dfbeta
                       _dfbeta_1: dfbeta(diffFANG_MTBV)
                       _dfbeta_2: dfbeta(diffFANG_EPS)

259 . scatter _dfbeta_1 _dfbeta_2 , ylabel(-.75(.25).5) yline(.246 -.246) mlabel(DATE DATE DATE)

260 . list _dfbeta_1 difflogINDEX diffFANG_MTBV diffFANG_EPS  if abs(_dfbeta_1) > 2/sqrt(66)

         +-----------------------------------------------+
         | _dfbeta_1   difflo~X   diffFAN~V   diffFANG~S |
         |-----------------------------------------------|
    19.  | -.7133223   .0428643   -2.434333    .60766667 |
    38.  |  .6115019   .1916103        2.05   -.01111111 |
    41.  |  .3203335   .1590457    1.542667    .19344444 |
    43.  |  -3.46233   .0296297   -4.520333         .031 |
    55.  | -.4979572   .0209394      -2.063        -.148 |
         +-----------------------------------------------+

261 . list _dfbeta_2 difflogINDEX diffFANG_MTBV diffFANG_EPS  if abs(_dfbeta_2) > 2/sqrt(66)

         +-----------------------------------------------+
         | _dfbeta_2   difflog~X   diffFAN~V   diffFAN~S |
         |-----------------------------------------------|
    19.  |  .5532886    .0428643   -2.434333   .60766667 |
    31.  |  .7510691   -.1081252   -.4693333       -.963 |
    49.  | -.4826459   -.0192909   -.1556667        .791 |
         +-----------------------------------------------+

262 . avplots, mlabel(DATE)

263 .
264 . *************
265 . **STATIONARITY (ONLY COINTEGRATION)
266 . *************
267 . vecrank logINDEX FANG_MTBV FANG_EPS //finds no cointegration equations

                       Johansen tests for cointegration
    Trend: constant                                    Number of obs =      64
    Sample:  2012-09 - 2017-12                                  Lags =       2
    -------------------------------------------------------------------------------
                                                                 5%
    maximum                                          trace    critical
      rank    parms       LL         eigenvalue   statistic     value
        0      12     -11.798827          .        28.6051*_   29.68
        1      17     -1.9449626      0.26504       8.8974     15.41
        2      20      2.3002353      0.12424       0.4070      3.76
        3      21       2.503727      0.00634
    -------------------------------------------------------------------------------
```

```
268 . //above is carried out as the original Johansen test found one equation which was cointegrate
269 .
270 . drop in 55/55
    (1 observation deleted)

271 . drop in 43/43 //ref earlier section we drop observation 2016-01 after testing for cointegrati
    (1 observation deleted)

272 . drop in 19/19
    (1 observation deleted)

273 .
274 . *************
275 . **NORMALITY OF RESIDUALS
276 . *************
277 . reg difflogINDEX diffFANG_MTBV diffFANG_EPS

          Source |       SS           df       MS      Number of obs   =        63
    -------------+----------------------------------   F(2, 60)        =     99.54
           Model |  .176288707         2  .088144353   Prob > F        =    0.0000
        Residual |  .053128484        60  .000885475   R-squared       =    0.7684
    -------------+----------------------------------   Adj R-squared   =    0.7607
           Total |  .229417191        62  .003700277   Root MSE        =    .02976

    ------------------------------------------------------------------------------
     difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
    -------------+----------------------------------------------------------------
    diffFANG_MTBV |   .0705838   .0052374    13.48   0.000     .0601075    .0810601
     diffFANG_EPS |    .021891   .0131068     1.67   0.100    -.0043266    .0481086
            _cons |    .001217   .0040286     0.30   0.764    -.0068413    .0092753
    ------------------------------------------------------------------------------

278 . predict res, resid

279 . kdensity res, normal

280 . histogram res
    (bin=7, start=-.09213172, width=.02426623)

281 . pnorm res

282 . qnorm res

283 . sktest res

                        Skewness/Kurtosis tests for Normality
                                                              ------ joint ------
        Variable |        Obs  Pr(Skewness)  Pr(Kurtosis) adj chi2(2)    Prob>chi2
    -------------+---------------------------------------------------------------
             res |         63     0.9009        0.0335        4.58         0.1014

284 . swilk res

                       Shapiro-Wilk W test for normal data

        Variable |        Obs       W           V         z       Prob>z
    -------------+------------------------------------------------------
             res |         63    0.95507      2.540     2.015    0.02197

285 . sfrancia res

                      Shapiro-Francia W' test for normal data

        Variable |       Obs       W'          V'        z       Prob>z
    -------------+-----------------------------------------------------
             res |        63    0.94749      3.284     2.281    0.01127
```

```
286 .
287 . *************
288 . **LINEARITY
289 . *************
290 . graph matrix res diffFANG_MTBV diffFANG_EPS, half

291 . acprplot diffFANG_MTBV, lowess lsopts(bwidth(1)) mlabel(DATE)

292 . acprplot diffFANG_EPS, lowess lsopts(bwidth(1)) mlabel (DATE)

293 .
294 . *************
295 . **HOMOSCEDASTICITY
296 . *************
297 . rvfplot, yline(0)

298 . estat imtest

    Cameron & Trivedi's decomposition of IM-test

    ---------------------------------------------------
                  Source |       chi2     df      p
    ---------------------+-----------------------------
      Heteroskedasticity |       7.48      5    0.1873
                Skewness |       2.96      2    0.2272
                Kurtosis |       3.31      1    0.0687
    ---------------------+-----------------------------
                   Total |      13.76      8    0.0883
    ---------------------------------------------------

299 . estat hettest

    Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
             Ho: Constant variance
             Variables: fitted values of difflogINDEX

             chi2(1)      =     3.12
             Prob > chi2  =   0.0772

300 .
301 . *************
302 . **MULTICOLLINIARITY
303 . *************
304 . vif

        Variable |       VIF       1/VIF
    -------------+----------------------
    diffFANG_EPS |      1.03    0.967237
    diffFANG_M~V |      1.03    0.967237
    -------------+----------------------
        Mean VIF |      1.03

305 .
306 . *************
307 . **ISSUES OF INDEPENDENCE
308 . *************
309 . dwstat

    Number of gaps in sample:  3

    Durbin-Watson d-statistic(  3,    63) =  2.132781
```

```
310 . scatter res DATE_YM

311 . graph matrix res difflogINDEX diffFANG_MTBV diffFANG_EPS, half

312 .
313 . *************
314 . **OMITTED VARIABLES
315 . *************
316 . linktest

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(2, 60)        =    110.32
       Model |  .180368016         2  .090184008   Prob > F        =    0.0000
    Residual |  .049049175        60  .000817486   R-squared       =    0.7862
-------------+----------------------------------   Adj R-squared   =    0.7791
       Total |  .229417191        62  .003700277   Root MSE        =    .02859

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |   .9606945   .0703335    13.66   0.000     .8200066    1.101382
      _hatsq |   1.792261    .802321     2.23   0.029     .1873799    3.397142
       _cons |  -.0050065   .0044927    -1.11   0.270    -.0139933    .0039802
------------------------------------------------------------------------------

317 . ovtest

    Ramsey RESET test using powers of the fitted values of difflogINDEX
           Ho:  model has no omitted variables
                      F(3, 57) =      5.21
                      Prob > F =      0.0030

318 .
319 . *************
320 . **FINAL MODEL
321 . *************
322 .
323 . reg difflogINDEX diffFANG_MTBV diffFANG_EPS //final model

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(2, 60)        =     99.54
       Model |  .176288707         2  .088144353   Prob > F        =    0.0000
    Residual |  .053128484        60  .000885475   R-squared       =    0.7684
-------------+----------------------------------   Adj R-squared   =    0.7607
       Total |  .229417191        62  .003700277   Root MSE        =    .02976

-------------------------------------------------------------------------------
 difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
diffFANG_MTBV |   .0705838   .0052374    13.48   0.000     .0601075    .0810601
 diffFANG_EPS |    .021891   .0131068     1.67   0.100    -.0043266    .0481086
        _cons |    .001217   .0040286     0.30   0.764    -.0068413    .0092753
-------------------------------------------------------------------------------

324 .
325 . *************
326 . **STANDARDIZED COEFFICIENTS
327 . *************
328 . reg difflogINDEX diffFANG_MTBV diffFANG_EPS, beta

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(2, 60)        =     99.54
       Model |  .176288707         2  .088144353   Prob > F        =    0.0000
    Residual |  .053128484        60  .000885475   R-squared       =    0.7684
-------------+----------------------------------   Adj R-squared   =    0.7607
       Total |  .229417191        62  .003700277   Root MSE        =    .02976
```

```
------------------------------------------------------------------------------
 difflogINDEX |      Coef.   Std. Err.      t    P>|t|                     Beta
--------------+---------------------------------------------------------------
diffFANG_MTBV |   .0705838   .0052374    13.48   0.000                 .8513357
 diffFANG_EPS |    .021891   .0131068     1.67   0.100                 .1055057
        _cons |    .001217   .0040286     0.30   0.764                        .
------------------------------------------------------------------------------

329 .
330 . log close
      name:  <unnamed>
       log:  C:\Users\rinem\Dropbox\IB\master thesis\STATA\FANG_log.smcl
  log type:  smcl
 closed on:  27 Apr 2018, 12:55:56
-------------------------------------------------------------------------------------------
```

```
      name:  <unnamed>
       log:  C:\Users\rinem\Dropbox\IB\master thesis\STATA\US_log.smcl
  log type:  smcl
 opened on:  27 Apr 2018, 12:54:00

 1 . //import excel data
 2 . import exc STATA_US.xlsx, firstrow clear

 3 .
 4 . //Creates date variable with YYYY/MM
 5 . generate int DATE_YM = mofd(DATE)

 6 . format DATE_YM %tmCY-N

 7 . drop DATE

 8 . drop F

 9 . tsset DATE_YM
           time variable:  DATE_YM, 1994-06 to 1999-12
                   delta:  1 month

10 . gen logINDEX = log(INDEX)

11 . gen difflogINDEX = D.logINDEX
   (1 missing value generated)

12 . gen diffUS_MTBV = D.US_MTBV
   (1 missing value generated)

13 . drop in 1/1
   (1 observation deleted)

14 .
15 . *****
16 . reg difflogINDEX diffUS_MTBV diffUS_EPS

      Source |       SS           df       MS      Number of obs   =        66
-------------+----------------------------------   F(2, 63)        =     38.75
       Model |  .290107691         2  .145053845   Prob > F        =    0.0000
    Residual |  .235812638        63  .003743058   R-squared       =    0.5516
-------------+----------------------------------   Adj R-squared   =    0.5374
       Total |  .525920328        65  .008091082   Root MSE        =    .06118

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 diffUS_MTBV |   .0602188   .0076693     7.85   0.000     .0448929    .0755447
  diffUS_EPS |   2.171603   .5885529     3.69   0.000     .9954736    3.347732
       _cons |   .0193176   .0089371     2.16   0.034     .0014583    .0371769
------------------------------------------------------------------------------

17 . *****
18 .
19 . *************
20 . **UNUSUAL AND INFLUENTIAL DATA
21 . *************
22 . predict r, rstudent //generate studentized residuals (used for finding outliers)
```

```
23 . list r DATE difflogINDEX diffUS_MTBV diffUS_EPS  if abs(r) > 2

        +---------------------------------------------------------+
        |         r   DATE_YM   difflo~X    diffUS_~V   diffUS_~S |
        |---------------------------------------------------------|
    44. |  2.257961   1998-02   .1759973   -.4263333   .02366667 |
    49. |  2.275845   1998-07   .2157912    .9736667   .00233333 |
    56. |  3.031522   1999-02   .2426953   -.2376667        .032 |
    60. | -2.374253   1999-06   -.105989   -1.297667   .03833333 |
    62. |  2.468184   1999-08   .0003285   -3.246667        .022 |
        +---------------------------------------------------------+

24 . predict lev, leverage

25 . list lev DATE difflogINDEX diffUS_MTBV diffUS_EPS if lev > .21

        +-------------------------------------------------------+
        |      lev   DATE_YM   difflo~X   diffUS_~V   diffUS~S |
        |-------------------------------------------------------|
    62. | .2143267   1999-08   .0003285   -3.246667       .022 |
        +-------------------------------------------------------+

26 . lvr2plot, mlabel(DATE)

27 . predict d, cooksd //Conventional cutoff point = 4/n

28 . list d DATE if d > 4/66

        +---------------------+
        |        d   DATE_YM  |
        |---------------------|
    43. |  .069319   1998-01  |
    44. | .0714397   1998-02  |
    51. | .1033448   1998-09  |
    56. | .2061147   1999-02  |
    60. | .2724449   1999-06  |
        |---------------------|
    62. | .5125224   1999-08  |
        +---------------------+

29 . predict dfit, dfits  //Conventional cutoff point = 2*sqrt(k/n)

30 . list dfit DATE if dfit > 2*sqrt(6/66)

        +---------------------+
        |     dfit   DATE_YM  |
        |---------------------|
    56. | .8359007   1999-02  |
    62. | 1.289124   1999-08  |
        +---------------------+

31 . dfbeta
                              _dfbeta_1: dfbeta(diffUS_MTBV)
                              _dfbeta_2: dfbeta(diffUS_EPS)

32 . scatter _dfbeta_1 _dfbeta_2 , ylabel(-.75(.25).5) yline(.246 -.246) mlabel(DATE DATE DATE)

33 . list _dfbeta_1 DATE difflogINDEX diffUS_MTBV diffUS_EPS  if abs(_dfbeta_1) > 2/sqrt(66)

        +----------------------------------------------------------+
        | _dfbeta_1   DATE_YM   difflog~X   diffUS_~V   diffUS_~S |
        |----------------------------------------------------------|
    37. | -.2477747   1997-07   -.0172977   -2.078333   .01816667 |
    49. |  .2546706   1998-07    .2157912    .9736667   .00233333 |
    60. |  .4780252   1999-06    -.105989   -1.297667   .03833333 |
    62. | -1.185718   1999-08    .0003285   -3.246667        .022 |
        +----------------------------------------------------------+
```

```
34 . list _dfbeta_2 DATE difflogINDEX diffUS_MTBV diffUS_EPS  if abs(_dfbeta_2) > 2/sqrt(66)

       ┌─────────────────────────────────────────────────────────┐
       │  _dfbeta_2   DATE_YM   difflog~X   diffUS_~V   diffUS_EPS │
       ├─────────────────────────────────────────────────────────┤
   43. │   .4146431   1998-01   -.1596131       -.903   -.03266667 │
   44. │   .3513183   1998-02    .1759973   -.4263333    .02366667 │
   51. │  -.4814856   1998-09   -.0576286   -.5346667         .032 │
   56. │   .7280378   1999-02    .2426953   -.2376667         .032 │
   60. │  -.7589083   1999-06    -.105989   -1.297667    .03833333 │
       ├─────────────────────────────────────────────────────────┤
   62. │   .4154735   1999-08    .0003285   -3.246667         .022 │
       └─────────────────────────────────────────────────────────┘

35 . avplots, mlabel(DATE)

36 .
37 . drop in 62/62 //observation with date 1999-08 is dropped
   (1 observation deleted)

38 . drop in 60/60
   (1 observation deleted)

39 . drop in 56/56
   (1 observation deleted)

40 .
41 . *************
42 . **Stationarity
43 . *************
44 .
45 . //tests differenced variables
46 . /*dfuller H0:random walk with a possible drift, Ha: stationarity around a linear time trend
   > p-value >0,05 => fail to reject H0*/
47 . dfuller difflogINDEX, trend

   Dickey-Fuller test for unit root                   Number of obs   =        59

                               ────────── Interpolated Dickey-Fuller ─────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(t)           -10.957            -4.130            -3.491            -3.175
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000

48 . dfuller diffUS_MTBV, trend

   Dickey-Fuller test for unit root                   Number of obs   =        59

                               ────────── Interpolated Dickey-Fuller ─────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(t)           -11.200            -4.130            -3.491            -3.175
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000

49 . dfuller diffUS_EPS, trend

   Dickey-Fuller test for unit root                   Number of obs   =        59

                               ────────── Interpolated Dickey-Fuller ─────────
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(t)            -7.043            -4.130            -3.491            -3.175
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000
```

```
50 .
51 . pperron difflogINDEX, trend

   Phillips-Perron test for unit root                 Number of obs   =        59
                                                      Newey-West lags =         3

                                  ---------- Interpolated Dickey-Fuller ---------
                     Test         1% Critical       5% Critical      10% Critical
                  Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)          -72.496           -26.006           -19.962           -16.926
    Z(t)            -11.375            -4.130            -3.491            -3.175
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.0000

52 . pperron diffUS_MTBV, trend

   Phillips-Perron test for unit root                 Number of obs   =        59
                                                      Newey-West lags =         3

                                  ---------- Interpolated Dickey-Fuller ---------
                     Test         1% Critical       5% Critical      10% Critical
                  Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)          -68.375           -26.006           -19.962           -16.926
    Z(t)            -12.482            -4.130            -3.491            -3.175
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.0000

53 . pperron diffUS_EPS, trend

   Phillips-Perron test for unit root                 Number of obs   =        59
                                                      Newey-West lags =         3

                                  ---------- Interpolated Dickey-Fuller ---------
                     Test         1% Critical       5% Critical      10% Critical
                  Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(rho)          -59.059           -26.006           -19.962           -16.926
    Z(t)             -7.073            -4.130            -3.491            -3.175
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.0000

54 .
55 . *************
56 . **NORMALITY OF RESIDUALS
57 . *************
58 . reg difflogINDEX diffUS_MTBV diffUS_EPS

         Source |       SS           df       MS      Number of obs   =        63
   -------------+----------------------------------   F(2, 60)        =     51.31
          Model |  .291560791         2  .145780395   Prob > F        =    0.0000
       Residual |   .17047057        60  .002841176   R-squared       =    0.6310
   -------------+----------------------------------   Adj R-squared   =    0.6187
          Total |  .462031361        62  .007452119   Root MSE        =     .0533

   ------------------------------------------------------------------------------
   difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
   -------------+----------------------------------------------------------------
    diffUS_MTBV |   .0675358   .0076969     8.77   0.000     .0521397    .0829319
     diffUS_EPS |   1.838068   .5711137     3.22   0.002     .6956709    2.980466
          _cons |   .0176101   .0078108     2.25   0.028     .0019862    .0332339
   ------------------------------------------------------------------------------
```

```
59 . predict res, resid

60 . kdensity res, normal

61 . histogram res
   (bin=7, start=-.09794776, width=.03451811)

62 . pnorm res

63 . qnorm res

64 . sktest res

                       Skewness/Kurtosis tests for Normality
                                                             ─────── joint ───────
       Variable │        Obs  Pr(Skewness)  Pr(Kurtosis) adj chi2(2)   Prob>chi2
   ─────────────┼───────────────────────────────────────────────────────────────
            res │         63     0.1109        0.9196        2.66         0.2639

65 . swilk res

                      Shapiro-Wilk W test for normal data

       Variable │        Obs       W           V         z       Prob>z
   ─────────────┼──────────────────────────────────────────────────────
            res │         63    0.97412      1.463     0.822    0.20553

66 . sfrancia res

                     Shapiro-Francia W' test for normal data

       Variable │       Obs       W'          V'        z       Prob>z
   ─────────────┼─────────────────────────────────────────────────────
            res │        63    0.97552      1.531     0.817    0.20695

67 .
68 . *************
69 . **LINEARITY
70 . *************
71 . graph matrix res diffUS_MTBV diffUS_EPS, half

72 . acprplot diffUS_MTBV, lowess lsopts(bwidth(1)) mlabel(DATE)

73 . acprplot diffUS_EPS, lowess lsopts(bwidth(1)) mlabel (DATE)

74 .
75 . *************
76 . **HOMOSCEDASTICITY
77 . *************
78 . rvfplot, yline(0)

79 . estat imtest

   Cameron & Trivedi's decomposition of IM-test

   ───────────────────────────────────────────────────
                 Source │       chi2     df      p
   ─────────────────────┼─────────────────────────────
     Heteroskedasticity │      10.29      5    0.0675
               Skewness │       2.58      2    0.2759
               Kurtosis │       0.11      1    0.7349
   ─────────────────────┼─────────────────────────────
                  Total │      12.98      8    0.1126
   ───────────────────────────────────────────────────
```

```
80 . estat hettest

   Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
            Ho: Constant variance
            Variables: fitted values of difflogINDEX

            chi2(1)      =     0.19
            Prob > chi2  =   0.6618

81 .
82 . *************
83 . **MULTICOLLINIARITY
84 . *************
85 . vif

       Variable |       VIF       1/VIF
   -------------+----------------------
      diffUS_EPS |      1.04    0.960530
     diffUS_MTBV |      1.04    0.960530
   -------------+----------------------
       Mean VIF |      1.04

86 .
87 . *************
88 . **ISSUES OF INDEPENDENCE
89 . *************
90 . dwstat

   Number of gaps in sample:  3

   Durbin-Watson d-statistic(  3,    63) =  1.827757

91 . scatter res DATE_YM

92 . graph matrix res difflogINDEX diffUS_MTBV diffUS_EPS, half

93 .
94 . *************
95 . **OMITTED VARIABLES
96 . *************
97 . linktest

         Source |       SS           df       MS      Number of obs   =        63
   -------------+----------------------------------   F(2, 60)        =     51.42
          Model |  .291782249         2  .145891124   Prob > F        =    0.0000
       Residual |  .170249112        60  .002837485   R-squared       =    0.6315
   -------------+----------------------------------   Adj R-squared   =    0.6192
          Total |  .462031361        62  .007452119   Root MSE        =    .05327

   ------------------------------------------------------------------------------
   difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
   -------------+----------------------------------------------------------------
           _hat |   1.019351   .1205397     8.46   0.000     .7782354    1.260466
         _hatsq |  -.2763441   .9891723    -0.28   0.781    -2.254983    1.702295
          _cons |     .00096   .0086751     0.11   0.912    -.0163927    .0183127
   ------------------------------------------------------------------------------

98 . ovtest

   Ramsey RESET test using powers of the fitted values of difflogINDEX
          Ho:  model has no omitted variables
                    F(3, 57) =      0.37
                    Prob > F =      0.7726
```

```
. 
. *************
. **FINAL MODEL
. *************
. 
. reg difflogINDEX diffUS_MTBV diffUS_EPS //final model

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(2, 60)        =     51.31
       Model |  .291560791         2  .145780395   Prob > F        =    0.0000
    Residual |   .17047057        60  .002841176   R-squared       =    0.6310
-------------+----------------------------------   Adj R-squared   =    0.6187
       Total |  .462031361        62  .007452119   Root MSE        =     .0533

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 diffUS_MTBV |   .0675358   .0076969     8.77   0.000     .0521397    .0829319
  diffUS_EPS |   1.838068   .5711137     3.22   0.002     .6956709    2.980466
       _cons |   .0176101   .0078108     2.25   0.028     .0019862    .0332339
------------------------------------------------------------------------------

. 
. *************
. **STANDARDIZED COEFFICIENTS
. *************
. reg difflogINDEX diffUS_MTBV diffUS_EPS, beta

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(2, 60)        =     51.31
       Model |  .291560791         2  .145780395   Prob > F        =    0.0000
    Residual |   .17047057        60  .002841176   R-squared       =    0.6310
-------------+----------------------------------   Adj R-squared   =    0.6187
       Total |  .462031361        62  .007452119   Root MSE        =     .0533

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
 diffUS_MTBV |   .0675358   .0076969     8.77   0.000                 .7020638
  diffUS_EPS |   1.838068   .5711137     3.22   0.002                 .2575117
       _cons |   .0176101   .0078108     2.25   0.028                        .
------------------------------------------------------------------------------

. 
. log close
      name:  <unnamed>
       log:  C:\Users\rinem\Dropbox\IB\master thesis\STATA\US_log.smcl
  log type:  smcl
 closed on:  27 Apr 2018, 12:54:08
-------------------------------------------------------------------------------------------
```

```
      name:  <unnamed>
       log:  C:\Users\rinem\Dropbox\IB\master thesis\STATA\CH_log.smcl
  log type:  smcl
 opened on:  27 Apr 2018, 12:51:52

. 
. //import excel data
. import exc STATA_CH.xlsx, firstrow clear

. 
. //Creates date variable with YYYY/MM
. generate int DATE_YM = mofd(DATE)

. format DATE_YM %tmCY-N

. drop DATE

. drop F

. drop G

. drop H

. drop I

. tsset DATE_YM
        time variable:  DATE_YM, 2012-06 to 2017-12
                delta:  1 month

. gen logINDEX = log(INDEX)

. gen difflogINDEX = D.logINDEX
(1 missing value generated)

. gen diffCH_MTBV = D.CH_MTBV
(1 missing value generated)

. drop in 1/1
(1 observation deleted)

. 
. *****
. reg difflogINDEX diffCH_MTBV diffCH_EPS

      Source |       SS           df       MS      Number of obs   =        66
-------------+----------------------------------   F(2, 63)        =     35.64
       Model |  .219082239         2   .10954112   Prob > F        =    0.0000
    Residual |  .193658251        63   .00307394   R-squared       =    0.5308
-------------+----------------------------------   Adj R-squared   =    0.5159
       Total |   .41274049        65  .006349854   Root MSE        =    .05544

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 diffCH_MTBV |   .0899674    .010944     8.22   0.000     .0680977    .1118372
  diffCH_EPS |   .0424263   .0164976     2.57   0.012     .0094585    .0753941
       _cons |    .015137   .0069687     2.17   0.034     .0012111    .0290629
------------------------------------------------------------------------------
```

```
21 . *****
22 .
23 . *************
24 . **UNUSUAL AND INFLUENTIAL DATA
25 . *************
26 . predict r, rstudent //generate studentized residuals (used for finding outliers)

27 . list r DATE difflogINDEX diffCH_MTBV diffCH_EPS  if abs(r) > 2

     +-----------------------------------------------------------+
     |        r   DATE_YM   difflog~X   diffCH_~V      diffCH_~S |
     |-----------------------------------------------------------|
  7. | 2.735125   2013-01   -.0442491   -2.166333           .176 |
 14. |  3.35107   2013-08    .2945204    1.247667      .01466667 |
 19. | 3.853364   2014-01    .0989795      -1.363      .57933333 |
 24. | 2.157061   2014-06    .1762681    .5033333      .01966667 |
 43. | 2.360222   2016-01    .0110188      -1.572      .46236667 |
     +-----------------------------------------------------------+

28 . predict lev, leverage

29 . list lev DATE difflogINDEX diffCH_MTBV diffCH_EPS if lev > .21

     +------------------------------------------------------------+
     |      lev   DATE_YM   difflog~X   diffCH_~V      diffCH_EPS |
     |------------------------------------------------------------|
  7. | .2136614   2013-01   -.0442491   -2.166333            .176 |
 31. | .3554977   2015-01   -.0921288    .0066667      -1.9043333 |
 46. | .3204918   2016-04    .0795579    .2196667       1.8993333 |
 58. | .2710867   2017-04     .055222        .511         -1.6495 |
     +------------------------------------------------------------+

30 . lvr2plot, mlabel(DATE)

31 . predict d, cooksd //Conventional cutoff point = 4/n

32 . list d DATE if d > 4/66

     +--------------------+
     |        d   DATE_YM |
     |--------------------|
  7. | .6143625   2013-01 |
 14. | .2289341   2013-08 |
 19. | .5315937   2014-01 |
 31. | .0680149   2015-01 |
 41. | .0751285   2015-11 |
     |--------------------|
 43. | .2641744   2016-01 |
 46. | .0971339   2016-04 |
 58. | .2272857   2017-04 |
     +--------------------+

33 . predict dfit, dfits  //Conventional cutoff point = 2*sqrt(k/n)

34 . list dfit DATE if dfit > 2*sqrt(6/66)

     +--------------------+
     |     dfit   DATE_YM |
     |--------------------|
  7. | 1.425724   2013-01 |
 14. | .8934889   2013-08 |
 19. | 1.394754   2014-01 |
 43. | .9219658   2016-01 |
 58. | .8313528   2017-04 |
     +--------------------+
```

```
35 . dfbeta
                       _dfbeta_1: dfbeta(diffCH_MTBV)
                       _dfbeta_2: dfbeta(diffCH_EPS)

36 . scatter _dfbeta_1 _dfbeta_2 , ylabel(-.75(.25).5) yline(.246 -.246) mlabel(DATE DATE DATE)

37 . list _dfbeta_1 DATE difflogINDEX diffCH_MTBV diffCH_EPS  if abs(_dfbeta_1) > 2/sqrt(66)

     +----------------------------------------------------------+
     | _dfbeta_1   DATE_YM   difflog~X   diffCH_~V   diffCH_~S  |
     |----------------------------------------------------------|
  7. | -1.369437   2013-01   -.0442491   -2.166333        .176  |
 14. |  .7840423   2013-08    .2945204    1.247667   .01466667  |
 19. | -1.130345   2014-01    .0989795      -1.363   .57933333  |
 39. |  .2920958   2015-09    -.155498      -1.342   .03666667  |
 41. |  .4243724   2015-11     .223814    1.247667   .02566667  |
     |----------------------------------------------------------|
 43. | -.8105009   2016-01    .0110188      -1.572   .46236667  |
     +----------------------------------------------------------+

38 . list _dfbeta_2 DATE difflogINDEX diffCH_MTBV diffCH_EPS  if abs(_dfbeta_2) > 2/sqrt(66)

     +----------------------------------------------------------+
     | _dfbeta_2   DATE_YM   difflog~X   diffCH~V    diffCH_EPS |
     |----------------------------------------------------------|
 19. |   .550449   2014-01    .0989795     -1.363     .57933333 |
 31. |  .4395443   2015-01   -.0921288   .0066667    -1.9043333 |
 46. | -.5247925   2016-04    .0795579   .2196667     1.8993333 |
 58. | -.7972999   2017-04     .055222       .511       -1.6495 |
     +----------------------------------------------------------+

39 . avplots, mlabel(DATE)

40 .
41 . drop in 58/58
   (1 observation deleted)

42 . drop in 19/19
   (1 observation deleted)

43 . drop in 7/7 //observation with date 2014-01 is dropped
   (1 observation deleted)

44 .
45 . *************
46 . **Stationarity
47 . *************
48 .
49 . //tests differenced variables
50 . /*dfuller H0:random walk with a possible drift, Ha: stationarity around a linear time trend
   > p-value >0,05 => fail to reject H0*/
51 . dfuller difflogINDEX, trend

   Dickey-Fuller test for unit root                   Number of obs   =        59

                                  ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
   ------------------------------------------------------------------------------
    Z(t)            -6.952            -4.130            -3.491            -3.175
   ------------------------------------------------------------------------------
   MacKinnon approximate p-value for Z(t) = 0.0000
```

```
52 . dfuller diffCH_MTBV, trend

   Dickey-Fuller test for unit root                   Number of obs   =        59

                                  ────────── Interpolated Dickey-Fuller ─────────
                     Test         1% Critical       5% Critical      10% Critical
                  Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(t)             -5.926            -4.130            -3.491            -3.175
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000

53 . dfuller diffCH_EPS, trend

   Dickey-Fuller test for unit root                   Number of obs   =        59

                                  ────────── Interpolated Dickey-Fuller ─────────
                     Test         1% Critical       5% Critical      10% Critical
                  Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(t)             -7.242            -4.130            -3.491            -3.175
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000

54 .
55 . pperron difflogINDEX, trend

   Phillips-Perron test for unit root                 Number of obs   =        59
                                                      Newey-West lags =         3

                                  ────────── Interpolated Dickey-Fuller ─────────
                     Test         1% Critical       5% Critical      10% Critical
                  Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(rho)          -54.429           -26.006           -19.962           -16.926
    Z(t)             -6.944            -4.130            -3.491            -3.175
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000

56 . pperron diffCH_MTBV, trend

   Phillips-Perron test for unit root                 Number of obs   =        59
                                                      Newey-West lags =         3

                                  ────────── Interpolated Dickey-Fuller ─────────
                     Test         1% Critical       5% Critical      10% Critical
                  Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(rho)          -40.906           -26.006           -19.962           -16.926
    Z(t)             -5.788            -4.130            -3.491            -3.175
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000

57 . pperron diffCH_EPS, trend

   Phillips-Perron test for unit root                 Number of obs   =        59
                                                      Newey-West lags =         3

                                  ────────── Interpolated Dickey-Fuller ─────────
                     Test         1% Critical       5% Critical      10% Critical
                  Statistic           Value             Value             Value
   ──────────────────────────────────────────────────────────────────────────────
    Z(rho)          -58.834           -26.006           -19.962           -16.926
    Z(t)             -7.255            -4.130            -3.491            -3.175
   ──────────────────────────────────────────────────────────────────────────────
   MacKinnon approximate p-value for Z(t) = 0.0000
```

```
58 .
59 . *************
60 . **NORMALITY OF RESIDUALS
61 . *************
62 . reg difflogINDEX diffCH_MTBV diffCH_EPS

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(2, 60)        =     69.32
       Model |  .280305529         2  .140152765   Prob > F        =    0.0000
    Residual |   .12131064        60  .002021844   R-squared       =    0.6979
-------------+----------------------------------   Adj R-squared   =    0.6879
       Total |  .401616169        62   .00647768   Root MSE        =    .04496

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 diffCH_MTBV |   .1201042   .0105372    11.40   0.000     .0990268    .1411817
  diffCH_EPS |   .0415457   .0157429     2.64   0.011     .0100552    .0730363
       _cons |   .0047193   .0059696     0.79   0.432    -.0072216    .0166602
------------------------------------------------------------------------------

63 . predict res, resid

64 . kdensity res, normal

65 . histogram res
   (bin=7, start=-.07931688, width=.03645869)

66 . pnorm res

67 . qnorm res

68 . sktest res

                    Skewness/Kurtosis tests for Normality
                                                          ------ joint ------
    Variable |        Obs  Pr(Skewness)  Pr(Kurtosis) adj chi2(2)   Prob>chi2
-------------+---------------------------------------------------------------
         res |         63     0.0000        0.0004        24.46         0.0000

69 . swilk res

                   Shapiro-Wilk W test for normal data

    Variable |        Obs       W           V         z       Prob>z
-------------+------------------------------------------------------
         res |         63    0.86396      7.690     4.409    0.00001

70 .
71 . *************
72 . **LINEARITY
73 . *************
74 . graph matrix res diffCH_MTBV diffCH_EPS, half

75 . acprplot diffCH_MTBV, lowess lsopts(bwidth(1)) mlabel(DATE)

76 . acprplot diffCH_EPS, lowess lsopts(bwidth(1)) mlabel (DATE)

77 .
78 . *************
```

```
 79 . **HOMOSCEDASTICITY
 80 . *************
 81 . rvfplot, yline(0)

 82 . estat imtest

    Cameron & Trivedi's decomposition of IM-test

    ---------------------------------------------------
                  Source |       chi2     df      p
    ---------------------+-----------------------------
      Heteroskedasticity |      40.34      5    0.0000
                Skewness |       9.21      2    0.0100
                Kurtosis |       2.18      1    0.1399
    ---------------------+-----------------------------
                   Total |      51.73      8    0.0000
    ---------------------------------------------------

 83 . estat hettest

    Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
             Ho: Constant variance
             Variables: fitted values of difflogINDEX

             chi2(1)      =     3.73
             Prob > chi2  =   0.0534

 84 .
 85 . *************
 86 . **MULTICOLLINIARITY
 87 . *************
 88 . vif

        Variable |       VIF       1/VIF
    -------------+----------------------
      diffCH_EPS |      1.00    0.999246
     diffCH_MTBV |      1.00    0.999246
    -------------+----------------------
        Mean VIF |      1.00

 89 .
 90 . *************
 91 . **ISSUES OF INDEPENDENCE
 92 . *************
 93 . dwstat

    Number of gaps in sample:  3

    Durbin-Watson d-statistic(  3,    63) =  2.102684

 94 . scatter res DATE_YM

 95 . graph matrix res difflogINDEX diffCH_MTBV diffCH_EPS, half

 96 .
 97 . *************
 98 . **OMITTED VARIABLES
 99 . *************
100 . linktest

          Source |       SS           df       MS      Number of obs   =        63
    -------------+----------------------------------   F(2, 60)        =     87.35
           Model |  .298941152         2  .149470576   Prob > F        =    0.0000
        Residual |  .102675017        60   .00171125   R-squared       =    0.7443
    -------------+----------------------------------   Adj R-squared   =    0.7358
           Total |  .401616169        62   .00647768   Root MSE        =    .04137
```

```
------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |   .9360069   .0805046    11.63   0.000     .7749738     1.09704
      _hatsq |   2.486455   .7534691     3.30   0.002     .9792926    3.993618
       _cons |  -.0110542   .0065103    -1.70   0.095    -.0240768    .0019684
------------------------------------------------------------------------------

101 . ovtest

    Ramsey RESET test using powers of the fitted values of difflogINDEX
           Ho:  model has no omitted variables
                      F(3, 57) =      5.47
                      Prob > F =      0.0022

102 .
103 . *************
104 . **FINAL MODEL
105 . *************
106 .
107 . reg difflogINDEX diffCH_MTBV diffCH_EPS //final model

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(2, 60)        =     69.32
       Model |  .280305529         2  .140152765   Prob > F        =    0.0000
    Residual |   .12131064        60  .002021844   R-squared       =    0.6979
-------------+----------------------------------   Adj R-squared   =    0.6879
       Total |  .401616169        62   .00647768   Root MSE        =    .04496

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 diffCH_MTBV |   .1201042   .0105372    11.40   0.000     .0990268    .1411817
  diffCH_EPS |   .0415457   .0157429     2.64   0.011     .0100552    .0730363
       _cons |   .0047193   .0059696     0.79   0.432    -.0072216    .0166602
------------------------------------------------------------------------------

108 .
109 . *************
110 . **STANDARDIZED COEFFICIENTS
111 . *************
112 . reg difflogINDEX diffCH_MTBV diffCH_EPS, beta

      Source |       SS           df       MS      Number of obs   =        63
-------------+----------------------------------   F(2, 60)        =     69.32
       Model |  .280305529         2  .140152765   Prob > F        =    0.0000
    Residual |   .12131064        60  .002021844   R-squared       =    0.6979
-------------+----------------------------------   Adj R-squared   =    0.6879
       Total |  .401616169        62   .00647768   Root MSE        =    .04496

------------------------------------------------------------------------------
difflogINDEX |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
 diffCH_MTBV |   .1201042   .0105372    11.40   0.000                 .8090336
  diffCH_EPS |   .0415457   .0157429     2.64   0.011                  .187315
       _cons |   .0047193   .0059696     0.79   0.432                        .
------------------------------------------------------------------------------

113 .
114 . log close
      name:  <unnamed>
       log:  C:\Users\rinem\Dropbox\IB\master thesis\STATA\CH_log.smcl
  log type:  smcl
 closed on:  27 Apr 2018, 12:52:00
-------------------------------------------------------------------------------------------
```