Master Thesis

Is there an association between the usage of ETFs by investors and the value relevance of certain accounting numbers?

Abstract:

This thesis examines the association between the usage of ETFs (Exchange Traded Funds) by investors and the value relevance of accounting information. This means the strength of the reaction by investors to the publication of new accounting numbers. A sample of data from US ETFs is used to establish whether an association exists between the usage of ETFs by investors and five selected accounting numbers, namely EBIT, EBITDA and the amounts of net earnings, cash flow from operations and the free cash flows. The results do not indicate that there is an association between the usage of ETFs by investors and the value relevance of accounting numbers. This thesis contributes to the literature in the area of the value relevance of accounting numbers and is relevant to organisations who are interested in investor behaviour.

Keywords: ETFs, the value relevance of accounting numbers, EBIT, EBITDA, Net earnings, Cash flow from operations, Free cash flows

Name: Jeroen Jacobus de Beer

Student number: 357786

Supervisor: R. van der Wal

Second reader: C.D. Knoops

Master Accounting & Finance

Erasmus School of Economics

Erasmus Universiteit Rotterdam

Table of contents

1. Introduction

Investors are interested in keeping both costs and risks low. A popular way of doing this is using an index fund that exactly follows the exchange on which it is traded. A specific type of index fund is called an exchange-traded fund or simply ETF. Like other index funds the ETF is a fund that follows the index on which it is traded. This can for example be the Dow Jones and S&P 500 indexes in the United States or the AEX Index in the Netherlands. ETFs can also be used for commodity indexes, rather than just regular stock exchanges. The exchange-traded fund is a relatively new type of fund as it was very small in use prior to the 1990's when its use increased dramatically to containing more than currently half of the market value.

As ETFs are a relatively new phenomenon in the world of stock exchange trading there has not been much scientific research into this type of fund. With the large market share these funds currently have, an investigation of the effects of ETFs is useful to both investors and regulators. The popularity of ETFs is the highest among institutional investors like pension funds and banks that often set up their own ETFs themselves. For example Robeco has its own exchange-traded fund and this holds for the investment bank UBS as well. Because an ETF automatically follows the index the costs are relatively low and the ease of use is high, using ETFs is very useful for those institutional investors and their clients. There is a potential danger in automatically following indices as the amount of trading necessary to keep up with the indexes is relatively high. With there being a lot of movement in the market to keep up with the market it is possible that the volatility of these moves increases due to the large reaction to an initial change in stock prices. This is a point of interest for market regulators as the stability of the market as an interest to regulators. If ETFs have an impact to the stability of the market, it may be necessary for regulatory agencies like the SEC in the United States and the AFM in the Netherlands to take measures in order to secure the stability of the market. For these reasons it is necessary for the stakeholders in the financial markets that an investigation into the effects of this type of investment fund, called ETFs does take place. To control for the effects on the market a good proxy is necessary that can show these effects. In order to make investment decisions, investors need information about the market and/or specific companies in which they consider investing. A major source of information are the accounting numbers that are published by the companies themselves on at least a yearly basis. With these accounting numbers investors can form an image of how the company is doing financially and base their decision on that image. This information could possibly not be as relevant to users of ETFs than to other investors. This is because ETFs only follow the market index rather than making a conscious decision whether or not to invest

in a certain stock or commodity. It is therefore possible that the value relevance of the accounting numbers published by companies has declined due to the usage of ETFs. This would mean that the reaction of the stock market to the publication of account numbers would decline in comparison to prior reactions. For the efficiency of the financial markets it is important that the value of stocks changes accordingly to the changes in financial circumstances experienced by the companies. If not these reactions are not accurate, the financial markets would not function efficiently leading to distortions in the economy. It is for these reasons above that it is in the interest of investors and of the public in general that this subject is investigated and why making this thesis is useful for the investors and for the general public. This thesis contributes to scientific literature linking the areas of ETFs and value relevance of accounting numbers together what has not been done previously.

Accounting numbers in itself are a quite vague term as there is a large amount of numbers that is published by companies. This master thesis focuses on accounting numbers by companies from the United States, as for this country has the most available data. Accounting numbers from the United States can be split into two categories with GAAP (Generally Accepted Accounting Principles) numbers and Non-GAAP numbers. The publication of GAAP numbers is compulsory for companies, where this is not the case for Non-GAAP numbers. The reason these Non-GAAP numbers are relevant is because these numbers can also be useful for investors as they provide further information into the financial situation of a company. Non-GAAP numbers can contain important information that is not provided by the GAAP numbers. Two of the most important accounting numbers to investors are the amount of earnings and the amount of cash flows of a company. There are however several versions of these numbers and as Dechow (1994) shows there is a difference between the value relevance of these numbers. For example the amount of earnings can be specified in several ways like the net returns, EBIT and EBITDA to name but a few. These numbers are all similar, but can create very different views on the financial situation of a company. The difference between the EBIT number and the EBITDA number is the amount of depreciation and amortization. These amounts can make a difference as the amount of depreciation can be either relatively high or low in comparison with other companies within the same industry.

The same holds for the amount of cash flows by a company, here there are different versions as well with the most prominent being the cash flow from operations and the free cash flows. Making a choice between the specific versions of the accounting numbers is difficult as the version have their own merits and disadvantages over each other. Therefore the decision has been taken to use different versions of the earnings and cash flows numbers for this Master thesis as they all contain information for investors and can provide insights into the dynamics between the value relevance of

accounting numbers and the usage of ETFs. The choice for these accounting numbers will be specified and motivated further in the theoretical background part of the thesis.

The central part of this master thesis is whether there is a relation between the usage of ETFs and the value relevance of accounting numbers like the cash flow of operations. A requirement for a conclusion on this subject is that a statistically significant relation is found, without this a relation between the two variables cannot be established. This is about the general relation, so the research question will not specify exactly what accounting number is meant. The term accounting numbers here will refer to all accounting numbers used for this master thesis. Therefore the research question of this Master thesis will be the following:

Is there an association between the usage of ETFs by investors and the value relevance of certain accounting numbers?

These 'certain accounting numbers' will be specified further and receive their own sub questions. The first accounting number that will be investigated is the Earnings Before Interest and Tax, better known as EBIT. The reason that EBIT has been chosen as an accounting number for this research is that EBIT is an important indication of the profit margin of a company. If EBIT is relatively high in comparison to the amount of sales a company has, that company has relatively high profit margins which indicate good financial health. This is why EBIT as an accounting number is relevant to investors and should therefore be included in this research. The first sub question of this Master thesis will be as follows:

Is there an association between the usage of ETFs by investors and the value relevance of EBIT?

The second accounting number that will be investigated is the Earnings Before Interest, Tax, Depreciation and Amortization number, which is also known as EBITDA. In comparison to EBIT the difference is for EBITDA that the amount of depreciation and amortization are included. This is useful for investors as the amount of depreciation and amortization can be very different between companies. The reason for this difference can be that for example airline companies may lease their aeroplanes rather than buying them, keeping these aeroplanes of the balance sheet of the airline company. These insights are useful to investors and therefore has EBITDA been included in this research with regard to the value relevance. The second sub question will be the following:

Is there an association between the usage of ETFs by investors and the value relevance of EBITDA?

The third accounting number that is included in this investigation is the amount of net earnings a company has. The amount of net earnings is the most prominent of the earnings numbers, as by publications of accounting information this number is always mentioned prominently. This is also means that the reaction of investors to this number is particularly strong. Also is the amount of net earnings is GAAP number, where EBIT and EBITDA numbers are non-GAAP numbers and for the completeness of this research a GAAP number should be included. The third sub question will be defined as follows:

Is there an association between the usage of ETFs by investors and the value relevance of the amount of net earnings?

The fourth accounting number that is included in this research is the cash flows from operations. This accounting number is relevant to investors as it indicates how much a company brings in terms of money from the regular operations. Companies can get income from irregular sources like the sale of certain assets, but these are not part of the long term operations of a company and therefore do not give a good view of the financial situation of a company to investors. For these reasons the cash flows from operations have been included in this research. The fourth sub question is as follows:

Is there an association between the usage of ETFs by investors and the value relevance of the cash flows from operations?

The fifth and last accounting number that will be included in this research is the free cash flows of a company. The free cash flows are relevant to investors, because free cash flows are necessary for a company to be able to do investments. This ability to do investments is very important to potential investors as it indicates the amount of value a company can create in the future. Companies with relatively high free cash flows are worth more to investors than their competitors would be. This is also the reason why the free cash flows has been included as an accounting number for this research. The fifth and final sub question will be as follows:

Is there an association between the usage of ETFs by investors and the value relevance of the free cash flows?

How the research question and the sub questions are intended to be answered is the following: at the basis of the Master thesis will be a sample. In order to get this sample and find an answer the research question and the sub question data has to be acquired. For the ETFs data will be acquired from the Bloomberg database. This sample will have 953 ETFs from the United States with a broad market focus. This is because with broad market focus these funds are representative for the ETFs in general. The sample period is from 2010 to 2016. For the value relevance of accounting numbers, the data will come from the published accounting data on the Compustat database on the Wharton Research Database Services. To check for the robustness of the tests conducted in this Master thesis there will be several control variables. One of these control variables will be the amount of restatements a company has made, as it indicates the accuracy and therefore the value relevance of the published data by this company. The data for this control variable will come from the EDGAR database which belongs to the American regulatory agency SEC. Another control variable will be the amount of intangible assets as Schipper and Vincent (1999) show this is a good proxy for the value relevance of the accounting information. This is because intangible assets are hard to measure and therefore the accuracy of the information is hard to establish. The third control variable will be the SIC industry code of a company. This is needed to control for industry effects.

The structure of this Master thesis will be starting with the introduction which you have just read. The next part of this Master thesis will go into the background of ETFs and the value relevance of accounting numbers. With this theoretical background the hypotheses of this Master thesis can be build up. After this theoretical background the research design with the methodology and the sample used for this thesis will be discussed in further detail. Then in the following part the results of the tests will be presented. This presentation will be followed by the interpretation of the results and it will be established whether or not an association can be established. The final part of this Master thesis will be the conclusion in which the research question and the sub questions will be answered. A summary of the results is that there is not an statistically significant association between the usage of ETFs and the value relevance of certain accounting numbers. There was not any statistically significant evidence that there is an association, therefore it must be concluded that there is not an association. The regressions also showed a lack of normal distribution and heteroscedasticity, which was caused by the data used being of a financial nature and therefore subject to fluctuation over a period of time.

2. Theoretical background

2.1 Introduction chapter two

In this chapter the theoretical basis of this master thesis will be discussed. First there will be an introduction and in-depth discussion of ETFs in paragraph 2.2, the type of fund on which this thesis will be based. This will be followed by a discussion of the concept of value relevance of accounting numbers in paragraph 2.3 and the theoretical background behind this concept. After that the five different accounting numbers that will be used in this master thesis will be investigated in separate paragraphs 2.4 up to 2.8. These paragraphs will also feature the hypotheses that this thesis has in combination with the corresponding accounting number. This chapter will finish with a recap of the previous paragraphs and conclusion in the last paragraph 2.9.

2.2 Introduction ETF

In order to investigate the relation between the ETF and the value relevance of accounting numbers it is important to have a good theoretical view of these concepts. A theoretical background to ETF (Kosev & Williams, 2011) is given by a bulletin of Kosev and Williams to the Reserve Bank of Australia and other sources.

ETFs are a relatively new phenomenon being very small in size until the mid-nineties. Since the mid-nineties the amount of ETF has risen dramatically from about 1 billion US dollars in 2001 to a market value of over 1200 billion US dollars in 2010. As this popularity of ETF is relatively new, investigating the results of the product is scientifically relevant which also holds for regulators as current regulations do not account for the large existence of ETF. The book Mutual Funds and Exchange-Traded Funds: Building Blocks to Wealth by (Baker, Filbeck, & Kiymaz, 2016) describes the rise of ETF as follows: 'The distinction of offering the first exchange-traded fund (ETF) belongs to Canada with the Toronto Index Participation Fund (TIP 35) introduced in 1989. In 1993, the first U.S. ETF began trading publicly tracking the S&P 500 index (ticker symbol SPY). Over the following two decades, the industry has experienced explosive growth with more than 1,600 ETFs trading worldwide and over 1,200 in the United States as of 2013 (Investment Company Institute 2014). With the large number of securities in a typical index, most investors cannot replicate an index. An ETF is a vehicle or "wrapper" that allows investors to gain exposure to the index indirectly with one security. Attempting to replicate an index an all its underlying holding can be cumbersome and cost-prohibitive for investors. The ETF fund structure allows this broad stock representation to be established in one simple vehicle'. Another quote from this book is about the basics of ETFs. The typical ETF is a fund that tracks such as the S&P 500, NASDAQ, or Dow Jones Industrial Average

(DJIA). An ETF trades similarily to an exchange-traded stock and has no investment minimum as opposed to mutual funds that have minimum investments typically in the $1,000 to $2,500 range. An index is a portfolio of securities with pre-defined weights (e.g., price-weighted, value-weighted and float-adjusted) representing a particular market, sector, or asset class. An ETF sponsor is a finanicial institution that issues and manages the ETF and its assets. ETF sponsors introduce new ETFs when market demand is high and close ETFs when market demand is low. At the time of this writing, the five largest U.S. ETF sponsors in descending order of assets are iShares, State Street, Vanguard, Power Shares and WisdomTree. Although dozens of other ETF sponsors exist, the aformentioned list represents some of the most recognizable and influential ETF sponsors. The book by Baker et al. (2016) also states that the AUM (assets under management)of ETFs has increased from 102 billion US dollars in 2002 to almost 1.7 billion US dollars in 2013, an increase which is more than fifteen-fold. Over the same period the amount of ETFs has risen similarily from 113 to 1,294, which is more than ten-fold in terms of increase. According to the bulletin of Kosev and Williams: 'The original appeal to investors of these products as their simplicity, low-cost diversification benefits and ability to trade intraday.' Kostovestsky (2003) describes ETFs as follows: 'an ETF does not need to pay to obtain shares of constituent stocks, operating instead through a process known as creation/redemption in-kind. This means that large investors can purchase a sizable number of shares of ETFs only by supplying a stock portfolio that matches the target index in weights and that has the same value as those shares.' These advantages are particularly relevant for institutional investors as the long term focus of institutional investors is helped by the low-cost diversification ETF provide. With ETF institutional investors can remain diversified without having to constantly update their position via intraday trading. Another advantage of ETF towards other mutual funds (Poterba & Shoven, 2002) would be both after-tax and pre-tax returns of ETF are very similar to the other funds. This means that with the relatively low effort required by companies to use ETF, the use of ETF over other mutual funds is beneficial to both institutional investors and regular investors. A comparison (Agapova, 2011) between conventional mutual funds and ETF suggests that investor preference between the two funds is mainly to the preferences of the investors, as the performance of conventional mutual funds and ETF is very similar. Another reason for the use of ETF over other index mutual funds, according to Kostovetsky (2003) is that active fund manager do not have superior selectivity skills over inactive funds. These active managers would instead incur extra costs that penalize active fund shareholders. A disadvantage of ETF (Ergungor, 2012) was shown by the so-called Flash Crash of 2010. On May 6, 2010 the Dow Jones Index first dropped by 998,5 points and then recovered again by 600 points. This all occurred within the space of 20 minutes with the cause being a trade algorithm having a major glitch on the e-mini futures exchange in Chicago. What happened was merely the result of a technical glitch, but the financial effects were large and it

showed that ETF were vulnerable to technical problems. This vulnerability of ETF is a clear disadvantage for those who would be interested in using it.

There are two types of ETF, namely physical and synthetic ETF with (Kosev & Williams, 2011) physical ETF being more popular in the United States and the Asia-Pacific region and synthetic ETF being more popular in Europe. The difference between the two types is physical ETF are focused on holding a benchmark and synthetic ETF use swaps to keep up with market changes. Both types are ETF, but their approach is different to keeping the fund diversified. A difference between ETF and other types of index mutual funds is that ETF (Engle & Sarkar, 2006) is ETF use the prices determined by supply and demand to determine the amount of shares traded rather than using the calculated Net asset value (NAV) for it.

2.3 The value relevance of accounting numbers

The value relevance of accounting numbers is an important concept in this research proposal as it will be the independent variable. Investors need information in order to make their investment decisions on and the accounting numbers published by firms are important pieces as they give indications of the firm performance. Accounting numbers will lead to reactions of investors, which allows the stock market to function in a good manner. The value relevance (Collins, Maydew, & Weiss, 1997) of accounting numbers is the relative impact of the release of new information which in this case is accounting numbers is on the stock performance of a company. As new information about a company is released into the market, there will be a reaction from the market to it as investors find out things about a company that they previously were unaware of. This value relevance seems to have declined, according to (Lev & Zarowin, 1999) and (Schipper & Vincent, 2003). However (Collins, Maydew, & Weiss, 1997) disagree with this notion and report that the value relevance has in fact increased slightly. As there seems to be discussion in this area, it would be scientifically relevant to investigate these changes. With the presence of ETF investors seem to have become less involved, which lead to a drop in the value relevance of accounting numbers as they will not react as strongly to it as before. There is a difference (Harris, Lang, & Moller, 1994) in the value relevance between countries, but this thesis only focuses on US firms meaning that this is not a problem. The value of accounting numbers is relevant to investors, but not (Holthausen & Watts, 2001) for standard setters as making inferences for the accounting standards is very difficult from the information contained by the accounting numbers. Recent research like (Lev & Zarowin, 1999) and (Schipper & Vincent, 2003) have however that the value to investors of these numbers is declining, as especially technology firms seem to be little relevance for their accounting numbers in order to calculate the value of these companies. It is also indicated by (Dechow, 1994) that the relevance of

different accounting numbers themselves is different, with the earnings number being considered more valuable by investors than the number of cash flows of the company. This is due to the timing and matching problems related to the cash flows that are not present in the earnings. Therefore this thesis will use several accounting numbers to measure the value relevance to investors on. Using several accounting numbers also gives more room to detect correlated omitted variables as the effect of the presence of ETF will be different on the individual accounting numbers. There are different ways to measure value relevance, the first is called the value relevance in a narrow sense which is used by (Lev & Zarowin, 1999). This method measures the association between earnings and returns in the standard model and the cash flows are also included in the extended version. Another way of measuring value relevance is using information content, which is the market reaction to earnings announcements. This market reaction is then analysed by using event study methodology with small windows. This approach is taken for example by (Collins, Li & Xie, 2009), (Landsman, Maydew & Thornock, 2012) and (Beaver, McNichols & Wang, 2018). For this thesis, which is an association study, the first approach of the value relevance in a narrow sense is used because it measures associations and therefore best suited for an association study . This thesis will use control variables as well in prevention of correlated omitted variables. These control variables will be discussed and motivated in the data and methodology part of this thesis.

2.4. EBIT

The first accounting number that will be investigated in relation with ETFs is the Earnings Before Interest and Taxes (EBIT). This is an operational number as it measures as the name would suggest the earnings before interest and taxes have been paid. This accounting number is of interest to investors because it is an indicator of the profit margin. A relatively high EBIT for a firm suggests that there is a large difference between the costs and the sales. Higher profit margins can be an indicator for higher firm profits in the future which leads to an increase in the value of the firm in general and of its stocks. For these reasons EBIT is relevant to investors and therefore should be there be a reaction by investors to the publication of a new EBIT number. In light of this, EBIT has been chosen as one of the accounting numbers in relation with ETFs. The process of buying and selling shares within ETFs is automated and therefore not dependent on actions of a fund manager trading actively. This means that ETFs do not respond as much as other investors to signals in the market spread by the publication of new accounting numbers such as EBIT. Rather than leading to changes in market value of stocks, ETFs merely follow moves that are already happening in the market. With this in mind the first alternative hypothesis with regard to the first sub question:

H1:There is a negative association between the usage of ETFs by investors and the value relevance of EBIT.

This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new EBIT numbers by firms. This means a decline in the value relevance of EBIT.

2.5 EBITDA

The second of the accounting numbers that will be investigated in relation with the usage of ETF is the Earnings Before Interest, Taxes, Depreciation and Amortization (EBITDA). This accounting number is relevant for investors for similar reasons as EBIT with a few differences. EBITDA is relevant, because it accounts for the differences in asset financing firms use. Leasing is in a way of financing assets that allows in case of operational leasing to keep certain assets off the balance sheet. This in turn decreases the amount of depreciation and amortization a firm has to incur. With information on the finance structure of a company EBITDA can be useful for the valuation of a company. In fact EBITDA is often used in the area of corporate finance to measure the value of a company. The use of EBITDA multiples, in which the amount of EBITDA is multiplied an arbitrary amount of times to calculate the value of a company, is widespread in this field. It is for these reasons that EBITDA has been chosen as one of the accounting numbers. The second sub question of this thesis has the following alternative hypothesis:

H2:There is a negative association between the usage of ETFs by investors and the value relevance of EBITDA.

This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new EBITDA numbers by firms. This means a decline in the value relevance of EBITDA.

2.6 The amount of net earnings

The third accounting number that will be investigated in relation with the usage of ETFs is the amount of net earnings. This number is about the earnings after all costs have been deducted which is not the case in the previous two accounting numbers EBIT and EBITDA. Another difference with these two accounting numbers is that the amount of net earnings is in accordance with US GAAP, where the other two are not and therefore considered to be Non-GAAP numbers. According to

(Leung & Veenman, 2016) Non-GAAP disclosures are particularly informative in firms that suffering from losses. The amount of net earnings is one of the most published accounting numbers as it gives a broad view of how a firm is performing financially. Where EBIT and EBITDA give a more nuanced view and some background information as how the profit is generated by a firm, the amount of net earnings is the most direct view for investors at the financial performance by a firm. As the amount of net earnings is so widely published investors are most likely to react to the publication, especially those investors that gather the least amount of information. With the highest probability of a reaction among investors, the reaction should therefore also be the strongest. This would also mean that a change in reaction by investors due to the usage of ETFs should be measured the most in this particular accounting number. It is for these reason that the amount of net earnings has been chosen as an accounting number for this investigation, because although it can regarded as crude the possible reaction should be the most clear. The third alternative hypothesis is related to the third sub question.

H3:There is a negative association between the usage of ETFs by investors and the value relevance of the net earnings.

This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new net earnings numbers by firms. This means a decline in the value relevance of the net earnings.

2.7 The cash flow from operations

The fourth accounting number that will be investigated in relation to the usage of ETFs is the cash flow from operations. This number focuses on the cash flow generated by the operating activities of a firm. It does not include long term investments, so it only gives an indication of firm performance on a short term/daily basis. This is still relevant for investors as problems with the cash flow from operations would indicate that a firm has current financial problems. This part of information is not included in the prior earnings-based numbers. It is for this reason that the cash flow from operations has been chosen as an accounting number for this thesis. Cash flow from operations is however not a perfect measure of financial performance, because of the timing and matching issues (Dechow, 1994) involved with cash flows. But as earnings and cash flows are both relevant to investors, cash flows from operations will be included. Prior research into the area of value relevance of accounting numbers like Francis and Schipper (1999) and Lev and Zarowin (1999) have taken a similar approach with both using earnings and cash flows in order to measure changes in the value relevance over

time and this thesis aims to be in line with prior research. The fourth alternative hypothesis will be the expected result of the fourth sub question.

H4:There is a negative association between the usage of ETFs by investors and the value relevance of the cash flow from operations.

This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new cash flow from operations numbers by firms. This means a decline in the value relevance of the cash flow from operations.

2.8 Free cash flows

The fifth and last accounting number that will be investigated in relation with the usage of ETFs by investors is the free cash flow generated by a firm. The free cash flow is calculated by the operational cash flow followed by then deducting the capital expenditures. Free cash flows are what is left for a firm to spend after maintaining or expanding its asset base. Therefore free cash flows are the basis for firms to pursue other shareholder value creating activities. For this reason free cash flows are relevant to investors as high free cash flows indicate that there are high future growth opportunities for a firm. High free cash flows can be regarded in two forms, either the free cash flows are high when compared to competitors or the free cash flows are high in comparison to the amount of equity a firm has. It is in this aspect that free cash flows are different from cash flows from operations, because free cash flows are more future oriented where cash flow from operations is more focused on current financial performance. Because of this difference both cash flow from operations and free cash flows are both included in this investigation. Free cash flows do not suffer as much from the timing and matching issues as the cash flow from operations as illustrated by Dechow (1994). Furthermore the relevance to investors of the free cash flows are high enough to merit the inclusion of free cash flows on its own accord. The fifth and final alternative hypothesis is based on the expected result of the fifth sub question.

H5:There is a negative association between the usage of ETFs by investors and the value relevance of the free cash flows.

This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new free cash flows numbers by firms. This means a decline in the value relevance of the free cash flows.

2.9 Recap and conclusion

In this chapter the theoretical basis behind this master thesis was discussed. In chapter 2.2 the concept of ETFs was introduced as the core concept behind this master thesis. The history of ETFs was mentioned and the theory behind it was discussed. In the following chapter 2.3 the same approach was taken with the other important concept the value relevance of accounting numbers. Prior literature was used to define this concept and explain the theoretical background of this concept. In the other chapters 2.4 to 2.8 the accounting numbers that will be used for this master thesis were discussed and their specific hypotheses were put on paper. These accounting numbers are EBIT (2.4), EBITDA (2.5), Net earnings (2.6), Cash flows from operations (2.7) and the free cash flows (2.8). These numbers were chosen in order to have earnings based numbers and cash flow based numbers, as well as having GAAP and Non-GAAP accounting numbers. The difference between the GAAP and Non-GAAP is not the main focus of this thesis, but it will be remarked on in the conclusion. The hypotheses behind these accounting numbers are very similar as there is expected to be a negative association between the usage of ETFs by investors and each specific accounting number.

Table Hypotheses
H1:There is a negative association between the usage of ETFs by investors and the value relevance of EBIT. This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new EBIT numbers by firms.
H2:There is a negative association between the usage of ETFs by investors and the value relevance of EBITDA. This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new EBITDA numbers by firms.
H3:There is a negative association between the usage of ETFs by investors and the value relevance of the net earnings. This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new net earnings numbers by firms.
H4:There is a negative association between the usage of ETFs by investors and the value relevance of the cash flow from operations. This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new cash flow from operations numbers by firms.
H5:There is a negative association between the usage of ETFs by investors and the value relevance of the free cash flows. This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new free cash flows numbers by firms.

3. Methodology

3.1 Introduction chapter three

In this chapter the methodology of this thesis will be discussed in paragraph 3.2. For this thesis a data sample will be used and this data will be explained in this chapter in paragraphs and 3.3 and 3.4. The sources of this data will also be mentioned and discussed in paragraph 3.4. There will also be an explanation as to how the sample that is used for this thesis is build up in the same paragraph 3.4. Furthermore the regression models that will be used for answering the sub questions and the research question will be mentioned in this chapter in paragraph 3.2. The motivation behind using these regression models will also be present in this chapter.

3.2 Methodology and the regression formulas

In order to answer the questions posed in this thesis models will be used. In these models there will be a regression between the dependent variable, the independent variable and the control variables. The dependent variable in this case will be the twelve month returns of ETFs averaged on their 10-year returns. 'Usage' is calculated by the twelve month returns of ETFs averaged on their 10-year returns compared to the general market returns. When the growth of ETFs is larger than that of the general market that indicates a growth in usage. The independent variable will be the value relevance of the accounting number specified in the last chapter. The returns that are used for this thesis will be given a separate model for the calculation using each accounting number, but the model will be similar for all. The use of this dependent variable and independent variable is the same as in the hypotheses and the research question and therefore these regression models will be able to provide answers to the hypotheses and the research question. The control variables used as control variables are the ones already mentioned in the previous paragraph, being the amount of R&D spending, the amount of intangible assets and the amount of restatements a firm has had to do. The regression models will also feature an error term, which is standard with regression formulas.

The regression models will be as follows:

ETF: = $\beta_0 + \beta_1 \times EBIT + \beta_2 \times R\&D + \beta_3 \times IA + \beta_4 \times Restatement + \varepsilon_\tau$

ETF: = $\beta_0 + \beta_1 \times EBITDA + \beta_2 \times R\&D + \beta_3 \times IA + \beta_4 \times Restatement + \varepsilon_\tau$

ETF: = $\beta_0 + \beta_1 \times Net\ earnings + \beta_2 \times R\&D + \beta_3 \times IA + \beta_4 \times Restatement + \varepsilon_\tau$

ETF: = $\beta_0 + \beta_1 \times Cash\ flow\ from\ operations + \beta_2 \times R\&D + \beta_3 \times IA + \beta_4 \times Restatement + \varepsilon_\tau$

ETF: = $\beta_0 + \beta_1 \times Free\ cash\ flow + \beta_2 \times R\&D + \beta_3 \times IA + \beta_4 \times Restatement + \varepsilon_\tau$

This thesis will look into several control variables, the first of which is the amount a firm spends on Research and Development (R&D). According to Lev and Zarowin (1999), Intangible investments, R&D in particular, are generally considered as the major driver of business change, creating new products, franchises, and improved production processes. The amount of R&D spending can indicate the value relevance as higher spending would mean that a firm is more innovation focused. It is uncertain how much the spending on R&D will provide in the future in terms of the value will create for a firm and therefore it becomes harder to calculate what a firm is or will be worth. With this higher uncertainty the value relevance will decline, so it is for this reason that the amount of R&D spending will be taken into account as an control variable. Another control variable is similar to the previous which is the amount of intangible assets. According to Lev and Zarowin (1999) the amount of intangible assets is relevant to the value relevance, because the intangible assets are often a large influence to the future growth opportunities of a firm. The value relevance depends whether the future value of a firm can be calculated, which is significantly harder with intangible assets who are hard to valuate by their nature. This is in combination with the future growth opportunities that are uncertain in value. So this means that higher amount of intangible assets leads to an increase in the uncertainty about the value of a firm. This in turn decreases the value relevance and for this reason it is the amount of intangible assets is used as a control variable in this thesis. In order to measure the intangible assets the natural logarithm will be used. The third control variable for this thesis will the amount of restatements a firm has made over the measuring period. If a firm has to make a restatement it has a damaging effect on the reputation of a firm and on its publicized numbers. Defond and Jiambalvo (1991) have found negative market reactions from earnings restatements because of accounting errors. When the numbers a firm publishes are less credible their value relevance for investors declines, as investors are uncertain of the truthfulness they will rely on them less in order to make their decisions. Firms with more restatements are judged to have a lower value relevance and therefore this variable is used as a control variable. This influence on the value relevance should be taken into account. Data on this is acquired or can be calculated through the EDGAR database of the SEC as all data filed to the SEC is published there.

Variable Description	
Dependent variable:	
ETF	Weighted average of growth of market value of ETFs used (based on 10 year return)
Independent variables:	
EBIT	Earnings Before Interest and Taxes for each year of available data
EBITDA	Earnings Before Interest, Taxes, Depreciation and Amortization for each year of available data
Net earnings	The amount of net earnings as reported by the company for each year of available data
Cash flow from operations	The amount of cash flow from operations for each year of available data
Free cash flows	The amount of free cash flows for each year of available data
Control variables:	
R&D	The amount of spending on Research and Development for each year of available data
IA	The amount of intangible assets of a company (natural logarithm) for each year of available data
Restatement	The amount of restatements of accounting information for each year of available data

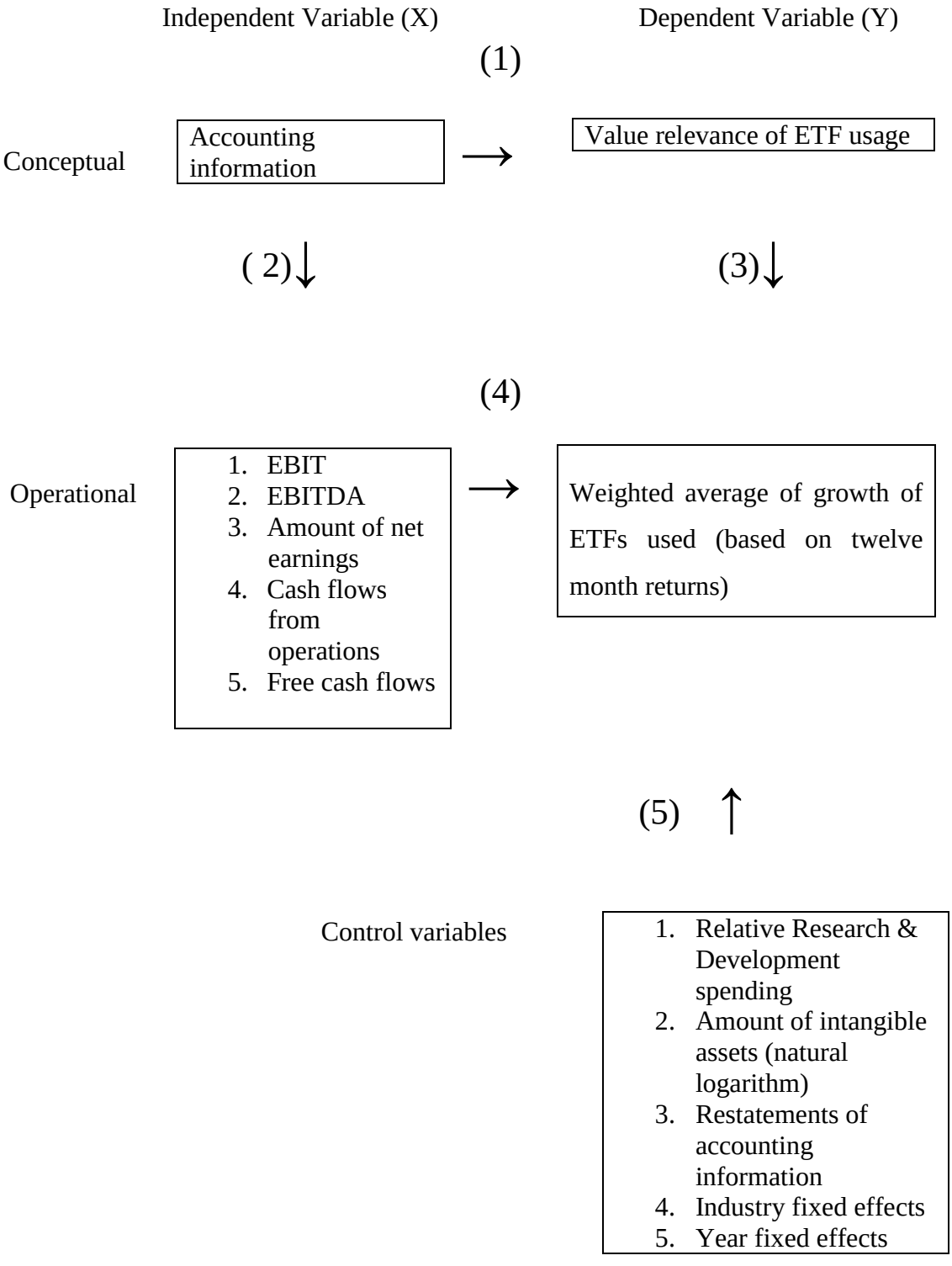
3.3 Data

For this thesis data has been obtained in order to answer the sub questions followed by the research question. The data for this sample has come from several sources. The first source from which data has been acquired is the database of Bloomberg. The reason of this decision to use Bloomberg is that Bloomberg contains information on ETFs, in particular their financial performance and their characteristics. These characteristics are for example in what specific industry or industries those ETFs invest in and in what particular percentage those ETFs do. It is specified in Bloomberg what strategies the ETFs follow, whether this ETF is value focused, growth focused or is blended in its focus. In order to capture both value focused and growth focused ETFs for this thesis data is used from ETFs that are blended in focus. The amount of ETFs used should be large enough to be balanced in whether the focus of the ETFs on value or growth. In order to capture a wide variety of the economy the ETFs chosen for this thesis are not specifically focussed on one industry, but operate in

many different industries. This approach is quite general, therefore industry effects will be used when investigating this data in order to keep the internal validity of this thesis high. Further exploration into this subject is however beyond the scope of this thesis and those who wish to investigate this subject more in the future are certainly welcome to do so.

Another source of data for this thesis is the Wharton Research Database Services or simply WRDS. To be specific the Compustat database is used in order to acquire the financial data of firms. This data is necessary to calculate the value relevance of the accounting numbers mentioned in the previous chapter. These numbers are all present in Compustat, so therefore the choice is made for this specific database out of the many databases WRDS provides. If there is any missing data on a firm, which is possible especially for the Non-GAAP numbers as publication of them is not required. The EDGAR database of the SEC will be used to get the missing information. This information should be available in the 10-K filings that firms do to the SEC, which then publicizes the filings on their EDGAR database. The amount of information a firm has to give to the SEC is more than it has to give to the general public. The variables taken from these databases will be further specified in the table in the appendix.

Predictive Validity Framework / Libby boxes



3.4 Validity

Another important part of this thesis is the prevention of endogeneity issues. There are two major endogeneity issues possible in this thesis. These issues are with external validity and internal validity. The goal of this thesis is establish a relation and the factors that affect it have to be accounted for. A problem here could be reverse causality, where the causal relation might not be between the independent variable and the dependent variable, but the other way around. For the omitted variables control variables will be used to counter problems with it. A visual presentation of this is presented in the Predictive Validity Framework on the previous page. This is also known as the Libby boxes based on (Libby, 2002). A correlation index between all the variables will also be provided in the Appendix. The sample that has been chosen for this thesis is with the external validity in mind, so that the sample is as representative as possible. In order to preserve the internal validity of this thesis, it is required to look into control variables as it is possible that they significantly influence the results. This influence should be prevented in order to keep the internal validity. Data availability is an issue when it comes to the ETF data, but it should not be a problem for the outcome of this thesis as all the necessary calculations in order to answer the research question and the sub questions can still be made. This means that the results required to answer the questions can be achieved. A visual presentation of this can be found in the Predictive Validity Framework.

3.5 Sample and data collection

The sample period will be from 2007 to 2015 for the Compustat database, because earlier data is not available or too early when compared to the ETF data. The source for the ETF data is as mentioned earlier the Bloomberg database where data over the period of the last 10 years can be found. This means that for the ETF data the sample period will be from 2007 to 2017. A disadvantage for this approach is there is a financial crisis falling early in the period, which may influence the results. For this reason year fixed effects will be used and in order to prevent industries influencing the results industry fixed effects will be used as well. Financial crises are however a part of the financial system and given the losses in that period have been made up by 2017 this should not be a problem. There also may be a survivorship bias as only ETFs that have existed throughout the measurement period are featured in this thesis, but as the ETFs in this thesis are wide in nature this bias should not play a large role. The sample period for the control variables will be from 2007 to 2015 like for the dependent variables as well, given that the sources for them are the same Compustat and EDGAR databases from WRDS and the SEC respectively. The similar sample period also ensures that there can be a good control, as there are not any observations on the dependent variables that are not matched by observations on the controlling variables. The sample period for the independent

variable is longer, which is not perfect for the comparison to the dependent and controlling variables. All observations do occur within the ten-year period used for the ETF data. These observations have an effect on the ten-year returns of the ETFs with the most recent data missing due to lack of availability. To account for this the last year return will be deducted from the overall 10 year return. As the data on 2015 became available in the first months of 2016 the changes in value of the ETFs should be accounted for. Also there will be the use of firm year fixed effects and industry fixed effects in order to increase the internal validity of this thesis. The sample itself exists of 254 observations and will be specified further in the next table. The sample started out with 1,168 observations from the weighted average of the growth of ETFs, but with missing observations this dropped to 254.

Table 2 Number of Observations			
EBIT	12,653	EBITDA	12,653
Net earnings	12,653	Cash flow from operations	12,653
Free cash flow	12,653	Weighted average of the growth of ETFs	1,168
Relative R&D	12,494	Natural logarithm Intangible assets	12,609
Amount of restatements	25,866		
Weighted average of the growth of ETFs	1,168	Combined after missing observations	254

3.6 Recap chapter three

In chapter three the data and methodology of this thesis were discussed. It was explained what the sources for the data were, namely three databases from Bloomberg, Compustat of WRDS and the EDGAR database from the SEC. There was also a motivation provided why these databases were chosen. For the ETF side of the data it was explained why certain ETFs were chosen for this thesis, namely a wide range of ETFs in terms of portfolio with a blended strategy. This was done in order to capture both value oriented and growth oriented funds in the data. For this thesis it is important to cast a wide net over the rather diverse world of ETFs, therefore both major types of ETFs are used in this sample. The sources for the data of the dependent variables and control variables were also provided. This being Compustat and EDGAR databases that have been previously mentioned with the note that the Compustat database is the most important of the two with the EDGAR database being used for data that is missing in Compustat. For the control variables it was specified which ones are going to be used, namely the amount of R&D spending, the amount of intangible assets and the

amount of restatements a firm has had to do. The motivation for the choice of these control variables was also provided. A sample was created out of the data and this sample was specified in terms of the amount of observations used and from which time period the data has come from. Some issues with the data or the time period of the data were discussed and discussed as well, because the internal validity of this thesis should be well guarded. Furthermore in this chapter the methodology of this thesis was explained. In order to calculate the answers to the research question and the sub questions a regression between a dependent variable (the value relevance of the specific accounting number) and independent variable (the aggregate growth in the value of the ETFs) is used. For each of the five accounting numbers used there is a separate formula, but the structure and the variables that are used are similar. There are also control variables and an error term in these regressions. The exact formulas that are used can be found in chapter 3.3 were they are specified per accounting number. To finish off the recap of this chapter a conclusion can be reached. This conclusion that the data was gathered from three different databases, namely the Bloomberg database, the Compustat database from WRDS and the EDGAR database from the SEC. The sample used in this thesis was shown and it was explained which choices had been made for the data used in it. Part of this was a discussion of the flaws in the data and how these flaws were addressed. This was followed by the methodology part of this chapter in which the regression used for this thesis was explained and the exact formulas were provided per accounting number.

4. Results

4.1 Introduction chapter four

In this chapter the descriptive statistics and the results (In the tables have been adjusted to letter size 10) from the regressions are discussed. The descriptive statistics will be first in paragraph 4.2. As there are five regressions in this thesis, there will be five paragraphs dedicated to the discussion of the results. The first regression to be discussed is involving EBIT. Its results will be discussed in paragraph 4.3. The second regression that will be discussed is involving EBITDA. The discussion of the results of this regression will be in paragraph 4.4. The third regression that will be discussed is involving Net earnings. The results of this regression will be discussed in paragraph 4.5. The fourth regression that will be discussed is involving the cash flow from operations. It will be in paragraph 4.6 that the results of this regression will be discussed. The fifth and final regression that will be discussed is involving the amount of free cash flows. The discussion of the results of this regression will be in paragraph 4.7. This chapter will finish with a recap and conclusion in paragraph 4.8, with this paragraph being based on the results obtained in the preceding paragraphs.

4.2 Descriptive statistics

Name variable	Observations	Mean	Standard Deviation	Min	Max	25%	75%	95% Confidence Interval	
GrowthWeightETF	1,168	0.0032017	0.016609	2.15e-08	0.309832	0.0000197	0.000876	0.0022482	0.0041552
EBIT	12,653	790.8941	3266.562	-12193	71230	-0.7125	261.352	733.9716	847.8166
EBITDA	12,653	1186.384	4587.504	-4431	81730	1.7045	397.264	1106.403	1266.853
Ni	12,653	496.8395	2751.539	-38732	104821	-4.2195	141.132	448.8917	544.7873
Oancf	12,653	938.0021	3821.109	-17332	81266	1.143	278.026	871.4161	1004.588
Fcf	12,653	475.3861	2463.534	-135191.3	57321	-0.89	210.4065	432.4571	518.3152
A	25,866	1.393911	0.8729359	1	12	1	1	1.383235	1.404587
Ln_xrd	12,609	3.068995	2.361815	-6.097755	9.54933	1.481718	4.566211	3.027766	3.110233
Rel_rdinten	12,494	36.55752	1154.696	-15.92729	85982,76	0.0002538	0.0523698	16.30865	56.80699
SicN	12,653	4225.361	1729.349	100	9997	3312	3861	4195.225	4255.496
Fyear	12,653	2010.843	2.564515	2007	2015	2009	2013	2010.799	2010.888

4.3 Results first regression (EBIT)

This paragraph is about the results from the first regression which based on the independent variable EBIT. The results from this regression will show correct whether the first hypothesis is correct , meaning that there is a negative association between the usage of ETFs by investors and the value relevance of EBIT. Fixed effects have been used in this regression with SicN representing the industry effects and fyear the year effects. This regression has 254 observations and its results are presented below along with a correlation table of all the variables used in this regression.

Correlation Table	* = 90 % significant ** = 95% significant ***= 99% significant						
	GrowthWeightETF	EBIT	A	Ln_xrd	Rel_rdinten	fyear	SicN
GrowthWeightETF	1.0000						
EBIT	-0.0321	1.0000					
A	0.0263	0.1226	1.0000				
Ln_xrd	-0.0426	0.4711	0.0053	1.0000			
Rel_rdinten	-0.0143	-0.0232	0.0330	-0.0112	1.0000		
Fyear	0.0357	0.1032	-0.0761	0.2178	-0.0003	1.0000	
SicN	0.0387	-0.0052	0.1243	-0.0805	-0.0568	-0.0391	1.0000

None of the correlations shown in this table is statistically significant. There are some that are worth mentioning. The largest correlation in this table is the correlation between the natural logarithm of the intangible assets (Ln_xrd) and EBIT which is 0.4711. The logic behind this is that firms with more EBIT can invest more and therefore have more intangible assets. A smaller correlation, 0.2178, exists between the firm year (Fyear) and the natural logarithm of the intangible assets (Ln_xrd). A reason for this correlation could be that intangible assets build up over time. The Shapiro-Wilk test with a graph of the observations from the independent and dependent variables in the Appendix shows that the variables are not normally distributed. An explanation for this is that the variables are financial data that fluctuate over the period of years. According to the Breusch-Pagan test in the Appendix the variables in this regression are also heteroskedastic with the F-statistic having a probability of 0.000. This is more evidence that the variables are not normally distributed. Also in the Appendix is the Breusch-Godfrey test for autocorrelation which shows that there is not autocorrelation as the null hypothesis of no serial correlation is not rejected.

Regression model: ETF: = $\beta_0 + \beta_1 \times EBIT + \beta_2 \times R\&D + \beta_3 \times IA + \beta_4 \times Restatement + \varepsilon_\tau$

Number of observations	254
F (4,154)	0.54
Prob > F	0.7058
R-squared	0.0822
Adjusted R-squared	-0.5078
Root MSE	0.0262

GrowthWeightETF	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
EBIT	3.88e-07	9.97e-07	0.39	0.697	-1.58e-06	2.36e-06
A	-0.0002853	0.0022013	-0.13	0.897	-0.0046339	0.0040633
Ln_xrd	-0.0016914	0.0011779	-1.44	0.153	-0.0040183	0.0006356
Rel_rdinten	-2.44e-07	1.42e-06	-0.17	0.864	-3.06e-06	2.57e-06
_cons	0.0105858	0.01068937	0.99	0.223	-0.0105198	0.0316913
SicN	F (95, 154) =		0.140	1.000		

The results of this regression are not statistically significant. The F-score is 0.54 which a probability of 0.7058 which is too low to be statistically significant. The adjusted R-squared is negative at -0.5078, this can be explained by the R-squared only being 0.0822. The adjustment than leads to the adjusted R-squared becoming negative. None of the T-scores are statistically significant either, although the probability of the natural logarithm (Ln_xrd) at 0.153 does come close to being statistically significant when industry fixed effects (SicN) are used to correct.

Number of observations	254
F (4 ,241)	0.22
Prob > F	0.9296
R-squared	0.0303
Adjusted R-squared	-0.0180
Root MSE	0.0215

GrowthWeightETF	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
EBIT	-8.92e-08	6.96e-07	-0.13	0.898	-1.46e-06	1.28e-06
A	0.0009812	0.001329	0.74	0.461	-0.0016367	0.003599
Ln_xrd	-0.0003195	0.007278	-0.44	0.661	-0.0017532	0.0011143
Rel_rdinten	-1.81e-07	1.10e-06	-0.17	0.869	-2.34e-06	1.98e-06
_cons	0.0009429	0.0063962	0.15	0.883	-0.0116568	0.0135425
Fyear	F(8,241) =		0.845	0.564		

The results of this regression when corrected for year fixed effects (Fyear) are not statistically significant. The F-score is 0.22 with the probability of 0.9296 which is too low to be statistically significant. The adjusted R-squared is negative at -0.0180 that can be explained by the low R-squared of 0.0303 which then after adjustment becomes negative. There are not any T-scores that are statistically significant with the most important being the T-score of EBIT. As this T-score is not statistically significant the first hypothesis has to be rejected.

4.4 Results second regression (EBITDA)

This paragraph includes the results of the second regression which is based on the independent variable of EBITDA. This regression tests whether the second hypothesis is correct, meaning that there is a negative association between the increase in usage of ETFs by investors and the value relevance of EBITDA. In this regression there are 254 observations and industry fixed effects are applied along with year fixed effects. For this SicN (industry fixed effects) and fyear (year fixed effects) are used. A correlation table for all the variables used in this regression is also provided.

Correlation Table	* = 90 % significant ** = 95% significant ***= 99% significant						
	GrowthWeightETF	EBITDA	A	Ln_xrd	Rel_rdinten	Fyear	SicN
GrowthWeightETF	1.0000						
EBITDA	-0.0307	1.0000					
A	0.0263	0.1143	1.0000				
Ln_xrd	-0.0426	0.4942	0.0053	1.0000			
Rel_rdinten	-0.0143	-0.0244	0.0330	-0.0112	1.0000		
Fyear	0.0357	0.1690	-0.0761	0.2178	-0.0003	1.0000	
SicN	0.0387	-0.0394	0.1243	-0.0805	-0.0568	-0.0391	1.000

In this correlation table there are not any statistically significant correlations. The largest correlation is the correlation between EBITDA and the natural logarithm of the intangible assets (Ln_xrd). The explanation for this could be that firms with more EBITDA have more room for investments which leads to a higher amount of intangible assets. In the Shapiro-Wilk test in the Appendix with a plot of the dependent and independent variables it is shown that the variables are not normally distributed. This can be explained by the fact that the variables are financial data that can fluctuate over a period of time. In the Breusch-Pagan test in the Appendix the variables of this regression are, as shown by the F-statistic with a probability of 0.000, heteroskedastic which is more evidence that the variables are not normally distributed. Testing for autocorrelation with the Breusch-Godfrey test in the Appendix shows that there is not autocorrelation as the null hypothesis for serial correlation is not rejected.

Regression model: ETF: = $\beta_0 + \beta_1 \times EBITDA + \beta_2 \times R\&D + \beta_3 \times IA + \beta_4 \times Restatement + \varepsilon_\tau$

Number of observations	254
F (4 ,154)	0.55
Prob > F	0.6994
R-squared	0.0824
Adjusted R-squared	-0.5074
Root MSE	0.0262

GrowthWeightETF	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
EBITDA	3.42e-07	7.93e-07	0.43	0.666	-1.22e-06	1.91e-06
A	-0.0003108	0.002204	-0.14	0.888	-0.0046648	0.0040432
Ln_xrd	-0.0017398	0.001206	-1.44	0.151	-0.0041223	0.0006426
Rel_rdinten	-2.46e-07	1.42e-06	-0.17	0.863	-3.06e-06	2.56e-06
_cons	0.0107375	0.0107132	1.00	0.318	-0.0104263	0.0319012
SicN	F (95, 154) =		0.140	1.000		

The results of this regression are not statistically significant with the control for industry fixed effects (SicN). The F-statistic is 0.55 with a probability of 0.6994 and that is not enough to be statistically significant. The adjusted R-squared is negative at -0.5074, but this can be explained by the adjustment that is made for the adjusted R-squared. When the regular R-squared has a low value,

0.0824 in this case the adjusted R-squared can be negative. None of the T-statistics is statistically significant either, although the T-statistic of the natural logarithm of the intangible assets does come close with a probability of 0.151.

Number of observations	254
F (4 ,241)	0.21
Prob > F	0.9319
R-squared	0.0302
Adjusted R-squared	-0.0181
Root MSE	0.0215

GrowthWeightETF	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
EBITDA	-7.36e-09	4.69e-07	-0.02	0.988	-9.31e-07	9.16e-07
A	0.0009609	0.0013285	0.72	0.470	-0.001656	0.0035778
Ln_xrd	-0.0003571	0.0007353	0.49	0.628	-0.0018055	0.0010914
Rel_rdinten	-1.79e-07	1.10e-06	-0.16	0.871	-2.24e-06	1.98e-06
_cons	0.0011001	0.0063934	0.17	0.864	-0.0114941	0.0136942
Fyear	F(8,241)=		0.846	0.563		

When controlled for year fixed effects (Fyear) the results of this regression are still not statistically significant. The F-score is 0.21 with a probability of 0.9319 which is not close to being statistically significant. The adjusted R-squared is -0.0181 which is negative and can be explained by the R-squared being only 0.0302. The T-scores of the variables are also not statistically significant with the most important being that one of EBITDA. Without statistically significant results for EBITDA the second hypothesis has to be rejected.

4.5 Results third regression (Net earnings)

The third regression which involves the amount of net earnings (ni) and its results will be discussed in this paragraph. The results of this regression determine whether the third hypothesis is correct or has to be rejected. This means that either is or is not a negative association between the increase in usage of ETFs by investors and the value relevance of the amount of net earnings. The regression has 254 observations and uses industry fixed effects and year fixed effects. These fixed effects are

represented by SicN for industry and fyear for year. For all the variables in this regression there is also a correlation table provided in this paragraph.

Correlation Table	* = 90 % significant ** = 95% significant ***= 99% significant						
	GrowthWeightETF	Ni	A	Ln_xrd	Rel_rdinten	Fyear	SicN
GrowthWeightETF	1.0000						
Ni	-0.424	1.0000					
A	0.0263	-0.0966	1.0000				
Ln_xrd	-0.0426	0.2567	0.0053	1.0000			
Rel_rdinten	-0.0143	-0.0128	0.0330	-0.0112	1.0000		
Fyear	0.0357	0.0278	-0.0761	0.2178	-0.0003	1.0000	
SicN	0.0387	-0.0500	0.1243	-0.0805	-0.0568	-0.0391	1.0000

There are not any statistically significant correlations in this correlation table. The correlation between the net earnings (ni) and the natural logarithm of the intangible assets is the largest with 0.2567. The reason for this correlation is that firms with higher earnings can invest more which leads to a build-up of intangible assets. The Shapiro-Wilk test in the Appendix with a plot of the dependent and independent variables shows that the variables are not normally distributed which can be explained by the fact that the variables are financial data that can fluctuate across a period of time. In the Appendix there is the Breusch-Pagan test that shows by the F-statistic with a probability of 0.0000 that the variables are heteroskedastic. This provides more evidence for the fact that the variables are not normally distributed. There is not autocorrelation in this model according to the Breusch-Godfrey test in the Appendix. The null hypothesis of no serial correlation was not rejected.

Regression model:

ETF: = $\beta_0 + \beta_1 \times Net\ earnings + \beta_2 \times R\&D + \beta_3 \times IA + \beta_4 \times Restatement + \varepsilon_\tau$

Number of observations	254
F (4 ,154)	0.53
Prob > F	0.7128
R-squared	0.0820
Adjusted R-squared	-0.5081
Root MSE	0.0262

GrowthWeightETF	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
Ni	3.99e-07	1.18e-06	0.34	0.736	-1.93e-06	2.73e-06
A	-0.0002202	0.0021994	-0.10	0.920	-0.0045262	0.0041247
Ln_xrd	-0.0016329	0.0011395	-1.43	0.154	-0.003884	0.0006182
Rel_rdinten	-2.52e-07	1.42e-06	-0.18	0.860	-3.06e-06	2.56e-06
_cons	0.0102604	0.0106036	0.97	0.335	-0.0106869	0.0312077
SicN	F(95,154)=		0.138	1.00		

There are not statistically significant results in this regression. The F-score is 0.53 and has a probability of 0.7128 which is not enough to be statistically significant. The adjusted R-squared is also negative at -0.5081. This is possible, because of the adjustment made to the R-squared can cause the adjusted R-squared if the regular R-squared is small. This is the case with the R-squared only being 0.0820. The coefficients of this regression with the control for industry fixed effects (SicN) are not statistically significant as well. Although the coefficient of the natural logarithm of the intangible assets (Ln_xrd) comes close with at T-score of -1.43 and a probability of 0.154.

Number of observations	254
F (4 ,241)	0.28
Prob > F	0.886
R-squared	0.0314
Adjusted R-squared	-0.0169
Root MSE	0.0215

GrowthWeightETF	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
Ni	-4.17e-07	7.79e-07	-0.54	0.593	-1.95e-06	1.12e-06
A	0.0008903	0.0013221	0.67	0.501	-0.001714	0.0034947
Ln_xrd	-0.0002684	0.0006682	-0.40	0.688	-0.0015848	0.0010479
Rel_rdinten	-1.82e-07	1.10e-06	-0.17	0.868	-2.34e-06	1.98e-06
_cons	0.0012377	0.006242	0.20	0.843	-0.0110581	0.0135336
Fyear	F (8,241) =		0.861	0.550		

When the regression is controlled for year fixed effects (Fyear) there are also not any statistically significant results. The F-score is 0.28 with a probability of 0.886 which is too high to be statistically significant. The adjusted R-squared is a little bit negative at -0.0169. The explanation for this is that the adjustment causes the already low R-squared at 0.0314 to become negative. The variables do not have statistically significant results either with the most important being the amount of net earnings that has a T-score of -0.54 and a probability of 0.593. Without statistically significant results the third hypothesis has to be rejected.

4.6 Results fourth regression (cash flow from operations)

It is the fourth regression that will be discussed in this paragraph. This regression involves the fourth independent variable the amount of cash flow from operations and has 254 observations. The results from this regression will show whether the fourth hypothesis is correct or has to be rejected. This means that there either is or is not a negative association between the usage of ETFs by investors and the value relevance of the cash flow from operations. In this regression there is also a test for the fixed effects of industry and year. These fixed effects are represented by SicN for industry and fyear for year in this regression. Provided is also a correlation table for all the variable used in this regression.

Correlation Table	* = 90 % significant ** = 95% significant ***= 99% significant						
	GrowthWeightETF	Oancf	A	Ln_xrd	Rel_rdinten	fyear	SicN
GrowthWeightETF	1.0000						
Oancf	-0.0195	1.0000					
A	0.0263	0.1313	1.0000				
Ln_xrd	-0.0426	0.4633	0.0053	1.0000			
Rel_rdinten	-0.0143	-0.0224	0.0330	-0.0112	1.0000		
Fyear	0.0357	0.1806	-0.0761	0.2178	-0.0003	1.0000	
SicN	0.0387	0.0053	0.1243	-0.0805	-0.0568	-0.0391	1.0000

In this correlation table none of the correlations is statistically significant. The largest correlation is 0.4633 and is between the amount of cash flow from operations (Oancf) and the natural logarithm of the intangible assets. An explanation for this correlation is that firms with more cash flow from operations can invest more which is associated with the amount of intangible assets. The variables are not normally distributed as shown by the Shapiro-Wilk test in the Appendix together with a plot of the dependent and independent variables. The fact that the data are financial data can explain this, as financial data fluctuates over time. The F-statistic of the Breusch-Pagan test in the Appendix has a probability of 0.0000, which shows that the variables are heteroskedastic and that means more evidence for the variables not being normally distributed. The Breusch-Godfrey test in the Appendix is for auto correlation. This is not found as the null hypothesis of no serial correlation is not rejected.

Regression model ETF: = $\beta_0 + \beta_1 \times Cash\ flow\ from\ operations + \beta_2 \times R\&D + \beta_3 \times IA + \beta_4 \times Restatement + \varepsilon_\tau$

Number of observations	254
F (4,154)	0.55
Prob > F	0.6994
R-squared	0.0824
Adjusted R-squared	-0.5074
Root MSE	0.0262

GrowthWeightETF	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
Oancf	4.55e-07	1.05e-06	0.43	0.667	-1.63e-06	2.54e-06
A	-0.0003113	0.0022041	-0.14	0.888	-0.0046655	0.004043
Ln_xrd	-0.0017475	0.0012151	-1.44	0.152	-0.0041479	0.0006529
Rel_rdinten	-2.47e-07	1.42e-06	-0.17	0.862	-3.06e-06	2.56e-06
_cons	0.0107277	0.0107097	1.00	0.318	-0.0104292	0.0318847
SicN	F(95,154) =		0.141	1.000		

This regression does not have any statistically significant results when controlled for industry fixed effects (SicN). The F-score is 0.55 and has a probability of 0.6994 which is not statistically significant. The adjusted R-squared is negative at -0.5074 which is caused by the adjustment made to the R-squared. If the R-squared is low which is the case here with 0.0824, this can happen. The T-scores of the tested variables are not showing statistically significant results either. The closest to being statistically significant is the natural logarithm of the intangible assets (Ln_xrd) with a T-score of -1.44 and a probability of 0.152.

Number of observations	254
F (4 ,241)	0.22
Prob > F	0.9286
R-squared	0.0303
Adjusted R-squared	-0.0180
Root MSE	0.0215

GrowthWeightETF	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
Oancf	8.03e-08	5.19e-07	0.15	0.877	-9.43e-07	1.10e-06
A	0.0009268	0.0013323	0.70	0.487	-0.0016976	0.0035512
Ln_xrd	-0.0004125	0.0007212	-0.57	0.568	-0.0018331	0.0010081
Rel_rdinten	-1.77e-07	1.10e-06	-0.16	0.872	-2.34e-06	1.98e-06
_cons	0.0013462	0.006408	0.21	0.834	-0.0112768	1.98e-06
Fyear	F(8,241) =		0.855	0.555		

The regression does not have any statistically significant results when controlled for year fixed effects (Fyear) as was the case when controlled for industry fixed effects (SicN). The F-score is 0.22 and has a probability of 0.9286 which is not enough to be statistically significant. The adjusted R-squared is negative at -0.0180. This score being a little bit negative can be explained by the adjustment made to the R-squared which can cause the adjusted R-squared to become negative if the R-squared is sufficiently low. The T-scores of the variables are also not statistically significant. The most important is the cash flow from operations (Oancf) which has a T-score of 0.15 and a probability of 0.877. With the results of this variable not being statistically significant, the fourth hypothesis has to be rejected.

4.7 Results of the fifth regression (free cash flows)

The discussion of the fifth regression and its results is what this paragraph consists of. The regression is based on the fifth independent variable the amount of free cash flows and has 254 observations. On the basis of the results of this regression it will be concluded whether the fifth hypothesis is correct or must be rejected. The conclusion would be that either there is or there is not a negative association between the usage of ETFs by investors and the value relevance of the free cash flows. As a control measure fixed effects are used in this regression in the form of SicN for industry and fyear for year. Included in this paragraph is also a correlation table for the all the variables used in this regression.

Correlation Table	* = 90 % significant ** = 95% significant ***= 99% significant						
	GrowthWeightETF	FCF	A	Ln_xrd	Rel_rdinten	Fyear	SicN
GrowthWeightETF	1.0000						
FCF	-0.0298	1.0000					
A	0.0263	0.1210	1.0000				
Ln_xrd	-0.0426	0.3903	0.0053	1.0000			
Rel_rdinten	-0.0143	-0.0191	0.0330	-0.0112	1.0000		
Fyear	0.0357	0.1241	-0.0761	0.2178	-0.0003	1.0000	
SicN	0.0387	-0.0106	0.1243	-0.0805	-0.0568	-0.0391	1.0000

All of the correlations presented in this correlation table are not statistically significant. The largest of the correlations is the correlation between the free cash flows (Fcf) and the natural logarithm of the intangible assets (Ln_xrd) at 0.3903. This correlation can be explained by the fact that firms that have more free cash flows can invest more and more investments leads to more intangible assets. It is

shown by the Shapiro-Wilk test along with a plot of the observations of the dependent and independent variables in the Appendix that the variables are not normally distributed. As the variables are financial data they are subject to fluctuation over time which can explain these results. The Breusch-Pagan test also shows that the variables are not normally distributed with the F-statistic having a probability of 0.0000. This indicates heteroscedasticity in the variables. Autocorrelation is not indicated in the variables, because the null hypothesis of no serial correlation in the Breusch-Godfrey Test in the Appendix is not rejected.

Regression model: ETF: = $\beta_0 + \beta_1 \times Free\ cash\ flow + \beta_2 \times R\&D + \beta_3 \times IA + \beta_4 \times Restatement + \varepsilon_\tau$

Number of observations	254
F (4 ,154)	0.53
Prob > F	0.7160
R-squared	0.0819
Adjusted R-squared	-0.5083
Root MSE	0.0262

GrowthWeightETF	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
Fcf	3.31e-07	1.06e-06	0.31	0.756	-1.77e-06	2.43e-06
A	-0.0003012	0.0022072	-0.14	0.892	-0.0046616	0.0040592
Ln_xrd	-0.0016233	0.0011425	-1.42	0.157	-0.0038803	0.0006336
Rel_rdinten	-2.47e-07	1.42e-06	-0.17	0.863	-3.06e-06	2.57e-06
_cons	0.0105434	0.0107259	0.98	0.327	-0.0106454	0.0317322
SicN	F(95,154) =		0.139	1.000		

This regression has a control for industry fixed effects (SicN) and with this control in this place it does not have any statistically significant results. The F-score is 0.53 and has a probability of 0.7160 which is not sufficient to be statistically significant. The adjusted R-squared is -0.5083 which is negative and can be explained by the adjustment made to get to the adjusted R-squared. When the R-squared is low, which is the case here with 0.0819, the adjusted R-squared can become negative. The tested variables do not have statistically significant results either. The variable that comes closest to being

statistically significant is the natural logarithm of the intangible assets (Ln_xrd) with a T-score of -1.42 and a probability of 0.157.

Number of observations	254
F (4 ,241)	0.22
Prob > F	0.9264
R-squared	0.0304
Adjusted R-squared	-0.0179
Root MSE	0.0215

GrowthWeightETF	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
Fcf	-1.56e-07	7.85e-07	-0.20	0.843	-1.70e-06	1.39e-06
A	0.000992	0.0013277	0.75	0.456	-0.0016234	0.0036074
Ln_xrd	-0.0003095	0.0006984	-0.44	0.658	-0.0016852	0.0010663
Rel_rdinten	-1.81e-07	1.10e-06	-0.17	0.869	-2.24e-06	1.98e-06
_cons	0.0008763	0.0063628	0.14	0.891	-0.0116576	0.0134101
Fyear	F(8,241) =		0.847	0.562		

When the regression is controlled for year fixed effects (Fyear) there are not any statistically significant results as was the case with the regression was controlled for industry fixed effects (SicN). The F-score is 0.22 and has a probability of 0.9264 which is not statistically significant. The adjusted R-squared is negative at -0.0179, this can be explained by the adjustment made to the R-squared which can become negative if the R-squared is sufficiently low. The R-squared is 0.0304, therefore this is the case. The variables tested in this regression do not show statistically significant results either. The most important variable is the free cash flows (Fcf) which has a T-score of -0.20 and a probability of 0.843. This is not statistically significant and with this result of this variable that means that the fifth hypothesis has to be rejected.

4.8 Recap

In this fourth chapter the results from the regressions were shown and discussed. For each regression the correlation table and results from the regression were provided in the text, descriptive statistics and graphs can be found in the appendix. The discussion on the research question will follow in the conclusion which is in the next chapter. The descriptive statistics were shown and discussed in paragraph 4.2.

The value relevance of EBIT was the first independent variable to be tested in paragraph 4.3 This was in relation with the first sub question. Linked to this sub question was the first hypothesis which expected on the basis of research like Lev & Zarowin (1999) and Schipper & Vincent (2003) that there would be a negative association between the dependent and independent variable. The results of the regressions were not statistically significant, therefore the first hypothesis had to be rejected. This means that there is not an association between the dependent and independent variables, since there is not any statistically significant results that would suggest that there is an association. Now the question is why this is not the case. Where based on Lev & Zarowin (1999) there is a relation between the control variables and the dependent variable, this does not hold for the independent variable. The value relevance of an accounting number in this case EBIT, depends on the strength of reaction of investors to the publication of new EBIT numbers and the perceived value of this new information. The intuition here is that because of investors using ETFs their reaction to the publication of new EBIT numbers would be less, as they are less involved in the market and the ETFs are reacting for them. The results from the regressions show that there does not appear to be a change in reaction from investors due to their usage of ETFs to the publication of new EBIT numbers. Also the perceived value of new EBIT numbers for investors does not appear to have changed as the control variable of the amount of restatements by a company has not produced statistically significant results. These two things that influence the value relevance of EBIT have not changed, therefore the conclusion on the first sub question must be that there is not an association between the usage of ETFs by investors and the value relevance of EBIT.

The second independent variable tested in paragraph 4.4 was the value relevance of EBITDA. This was in relation to the second hypothesis which is based on the second sub question. The hypothesis is that on the basis of Lev & Zarowin (1999) and Schipper & Vincent (2003) there is a negative association between the dependent and independent variables. The results of these regressions are not statistically significant and therefore the conclusion must be that the hypothesis had to be rejected. EBIT and EBITDA are very similar and with the similar results it seems that the cause of the

results is also similar. The conclusion to the second sub question is that there is not an association between the usage of ETFs and the value relevance of EBITDA.

Paragraph 4.5 was where the amount of net earnings, the third independent variable, was tested. This was in relation to the third sub question on which the third hypothesis is based. The hypothesis is that on the basis of Lev & Zarowin (1999) and Schipper & Vincent (2003) there is a negative association between the dependent and independent variables. The results of these regressions are not statistically significant and therefore the conclusion must be that the hypothesis had to be rejected. The amount of net earnings is like EBIT and EBITDA, but with the important difference that the amount of net earnings conforms to US GAAP. This is not the case for EBIT and EBITDA, this is why the amount of net earnings was important to research. On the basis of the results the strength of reaction of investors and the perceived value of new net earnings numbers has not changed. The conclusion to the third sub question is that there is not an association between the usage of ETFs and the value relevance of the amount of net earnings.

The fourth regression and its results were discussed in paragraph 4.6. The fourth independent variable that was tested is the value relevance of the cash flow from operations. This was in relation to the fourth sub question which forms the basis for the fourth hypothesis. The hypothesis is that on the basis of Lev & Zarowin (1999) and Schipper & Vincent (2003) there is a negative association between the dependent and independent variables. The results of these regressions are not statistically significant and therefore the conclusion must be that the hypothesis had to be rejected. According to Francis & Schipper (1999) and Lev & Zarowin the cash flow from operations is in terms of value relevance similar to EBIT and EBITDA. Because the results are similar, it seems that the cause of the results is also similar. The conclusion to the fourth sub question is that there is not an association between the usage of ETFs and the value relevance of the cash flow from operations.

The fifth and last independent variable tested in this thesis was the value relevance of the free cash flows. This was in paragraph 4.7 and in relation to the fifth hypothesis which is based on the fifth sub question. The hypothesis is that on the basis of Lev & Zarowin (1999) and Schipper & Vincent (2003) there is a negative association between the dependent and independent variables. The results of these regressions are not statistically significant and therefore the conclusion must be that the hypothesis had to be rejected. The free cash flows are similar to the cash flow from operations, but does not suffer as much from the timing and matching issues as shown by Dechow (1994) that the cash flow from operations has. Because the similarity between the two forms of cash flows and the similar results of the regressions, a similar cause of the results seems to be the case. The conclusion

to the second sub question is that there is not an association between the usage of ETFs and the value relevance of free cash flows.

In the next chapter a conclusion for these results will be discussed and complete overview of this thesis and all its aspects will be provided.

5 Conclusion

This chapter will be the conclusion of this thesis. To start off with the conclusion it is necessary to have a good framework and therefore this chapter will begin with a recap of the previous chapters. Chapter 1 was the introduction of this thesis in which the core concepts, ETFs and the value relevance of accounting numbers were introduced. Some background information on these concepts was provided and the motivation, both scientifically and practically, behind this thesis was also explained. After that the research question and the five sub questions were formulated. These are the following:

Research question	Is there an association between the usage of ETFs by investors and the value relevance of certain accounting numbers?
Sub question 1	Is there an association between the usage of ETFs by investors and the value relevance of EBIT?
Sub question 2	Is there an association between the usage of ETFs by investors and the value relevance of EBITDA?
Sub question 3	Is there an association between the usage of ETFs by investors and the value relevance of the net earnings?
Sub question 4	Is there an association between the usage of ETFs by investors and the value relevance of the cash flow from operations?
Sub question 5	Is there an association between the usage of ETFs by investors and the value relevance of the free cash flows?

In Chapter 2 the theoretical background behind this thesis was introduced. The core concepts ETFs and the value relevance of accounting numbers were discussed in depth at the hand of scientific sources. For the value relevance of accounting five proxies were found, namely EBIT, EBITDA, the net earnings, the cash flow from operations and the free cash flows. Each of these proxies was discussed in a separate paragraph and based on the theoretical background hypotheses were formed based on each of the sub questions. These hypotheses were the following:

Table Hypotheses
H1:There is a negative association between the usage of ETFs by investors and the value relevance of EBIT. This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new EBIT numbers by firms.
H2:There is a negative association between the usage of ETFs by investors and the value relevance of EBITDA. This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new EBITDA numbers by firms.
H3:There is a negative association between the usage of ETFs by investors and the value relevance of the net earnings. This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new net earnings numbers by firms.
H4:There is a negative association between the usage of ETFs by investors and the value relevance of the cash flow from operations. This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new cash flow from operations numbers by firms.
H5:There is a negative association between the usage of ETFs by investors and the value relevance of the free cash flows. This association is expected to be negative, as the decreased market participation which is caused by the usage of ETFs leads to a decrease in market reaction to the publication of new free cash flows numbers by firms.

In chapter 3 the methodology of this thesis was explained. In order to answer the research question and following sub questions it was necessary to obtain data. First the data was gathered, an explanation for the choice of this specific data and the sources of this data can be found in this chapter. Out of this data the sample was created that would be used to answer the questions. How this sample was made is explained in paragraph 3.5. In order to calculate the results on which the sub questions can be answered five regression formulas were used and these can be found in paragraph 3.2. The methodology used for this thesis is not perfect, which means that there are some validity issues. These issues on the internal and external validity and the measures taken to prevent these issues can be found in paragraph 3.4.

In Chapter 4 the results of the regressions are presented in the following order: EBIT in paragraph 4.2, EBITDA in paragraph 4.3, the amount of net earnings in paragraph 4.4, the cash flow from operations in paragraph 4.5 and the free cash flows in paragraph 4.6. These results are the basis on which the research question will be answered and this thesis will be concluded.

Now all the sub questions have been answered, it is time to answer the research question on which this entire thesis is based. The answer to the research question is that there is not an association between the usage of ETFs by investors and the value relevance of certain accounting numbers, at least the accounting numbers that were covered in this thesis. The explanation for this lack of association would be that investors have not changed their reaction to the publication of these new accounting numbers. The usage of ETFs by investors has been increasing over the period examined in this thesis, however this does according to the results not have an effect on the accounting numbers. The assumption was the increase in usage of ETFs by investors would lead to the investors becoming less active in the market. This decrease in market activity would then lead to a decrease in reaction to the publication of new accounting numbers. The decrease in reaction would then cause a drop in the value relevance of these accounting numbers, in the case of this thesis EBIT, EBITDA, the net earnings, the cash flow from operations and the free cash flows. The results indicate that this assumption was false and without it can only be concluded that there is not an association between the usage of ETFs and the value relevance of certain numbers. As there is not a difference between the results of the GAAP and Non-GAAP numbers, because they were all not statistically significant, there does not appear to be a difference between the GAAP and Non-GAAP numbers in terms of the association between the usage of ETFs and the value relevance.

To close this thesis off, it is important to say what the limitations of this thesis are. Possibilities for further research in the future should also be mentioned. One of the limitations of this thesis is that there is the focus on five different accounting numbers, namely EBIT, EBITDA, the net earnings, the

cash flow from operations and the free cash flows. The conclusions from these accounting numbers is not necessarily applicable for other accounting numbers. Doing this thesis again with other accounting numbers is a possibility for future research.

Another limitation is that the results are based on firms from the US, a country which does not follow IFRS. It could be the case that for IFRS countries the results would be different and therefore conclusions from this thesis should not be drawn for them. This is also an area where future research is possible and would be valuable as a large amount of countries do follow IFRS. A third limitation of this thesis is that the time period that is investigated contains a large financial crisis. Year fixed effects have been used to deal with this limitation, although it is still possible that this financial crisis might have an effect. It is an area that deserves future research as investigating other years would show this effect if it is still there.

6. Appendix:

Bibliografie

Agapova, A. (2011). Conventional mutual index funds versus exchange-traded funds. *Journal of Financial Markets, 14*(2), 323-343.

Baker, H. K., Filbeck, G., & Kiymaz, H. (2016). *Mutual Funds and Exchange-Traded Funds Building Blocks To Wealth.* New York: Oxford University Press.

Beaver, W. H., McNichols, M. F., & Wang, Z. Z. (2018). The information content of earnings announcements: new insights from intertemporal and cross-sectional behavior. *Review of Accounting Studies, 23*(1), 95-135.

Collins, D. W., Li, O. Z., & Xie, H. (2009). What drives the increased informativeness of earnings announcements over time? *Review of Accounting Studies, 14*(1), 1-30.

Collins, D., Maydew, E., & Weiss, I. (1997). Changes in the value-relevance of earnings and book values over the past 40 years. *Journal of Accounting and Economics, 24*(1), 39-67.

Dechow, P. (1994). Accounting earnings and cash flows as measures of firm performance The role of accounting accruals. *Journal of Accounting & Economics, 18*(1), 3-42.

Engle, R., & Sarkar, D. (2006). Premiums-Discounts and Exchange Traded Funds. *The Journal of Derivatives, 13*(4), 27-45.

Ergungor, O. E. (2012). *Exchange-Traded Funds.* Cleveland, Ohio USA: Federal Bank of Cleveland.

Francis, J., & Schipper, K. (1999). Have Financial Statements Lost Their Relevance? *Journal of Accounting Research, 37*(2), 319-352.

Harris, T. H., Lang, M., & Moller, H. P. (1994). The Value Relevance of German Accounting Measures: An Empirical Analysis. *Journal of Accounting Research, 32*(2), 187-209.

Holthausen, R. W., & Watts, R. L. (2001). The relevance of the value-relevance literature for financial accounting standard setting. *Journal of Accounting and Economics, 31*(1-3), 3-75.

Kosev, M., & Williams, T. (2011). *Exchange-traded Funds.* Sydney: Reserve Bank of Australia.

Kostovetsky, L. (2003). Index Mutual Funds and Exchange-Traded Funds. *The Journal of Portfolio Management, 29*(4), 80-92.

Landsman, W. R., Maydew, E. L., & Thornock, J. R. (2012). The information content of annual earnings announcements and mandatory adoption of IFRS. *Journal of Accounting and Economics, 53*(1-2), 34-54.

Leung, E., & Veenman, D. (2016). *Non-GAAP Earnings Disclosure in Loss Firms.* Rotterdam: Erasmus Universiteit Rotterdam.

Lev, B., & Zarowin, P. (1999). The Boundaries of Financial Reporting and How to Extend Them. *Journal of Accounting Research, 37*(2), 353-385.

Libby, R., Bloomfield, R., & Nelson, M. W. (2002). Experimental research in financial accounting. *Accounting, Organizations and Society, 27*(8), 775-810.

Poterba, J., & Shoven, J. (2002). *EXCHANGE TRADED FUNDS: A NEW INVESTMENT OPTION FOR TAXABLE INVESTORS.* Cambridge, Massachusetts USA: Massachusetts Institute of Technology Department of Economics.

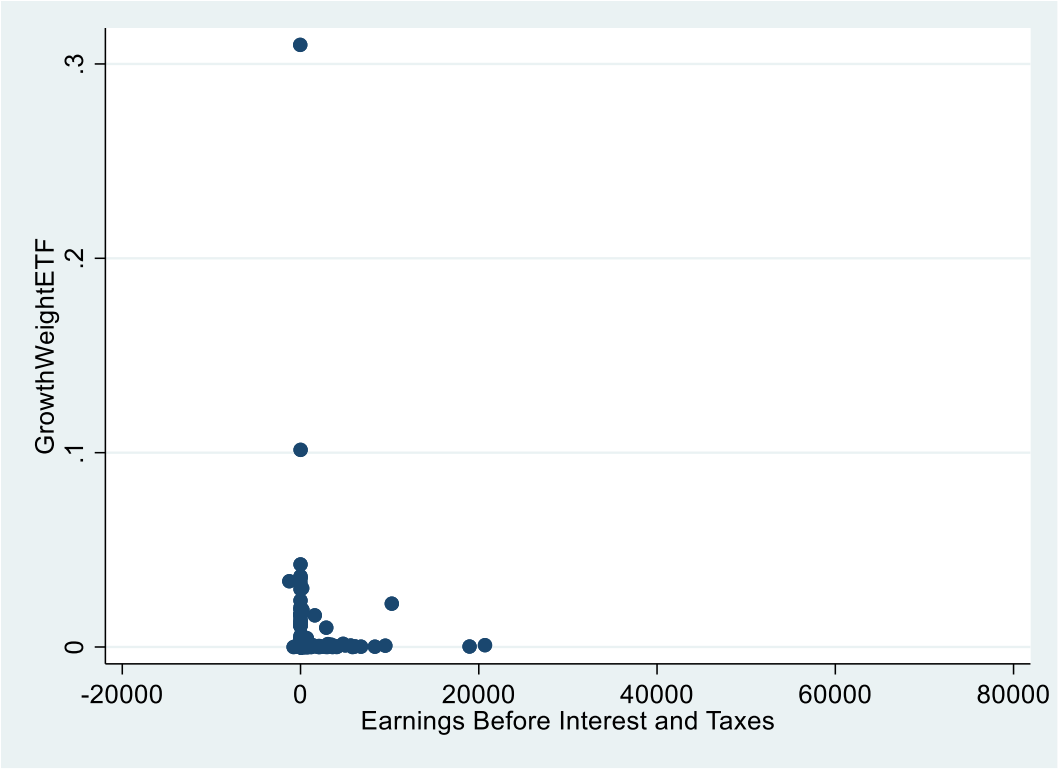
Schipper, K., & Vincent, L. (2003). Earnings quality. *Accounting Horizons, 17*, 97-110.

Table 1 List of variables

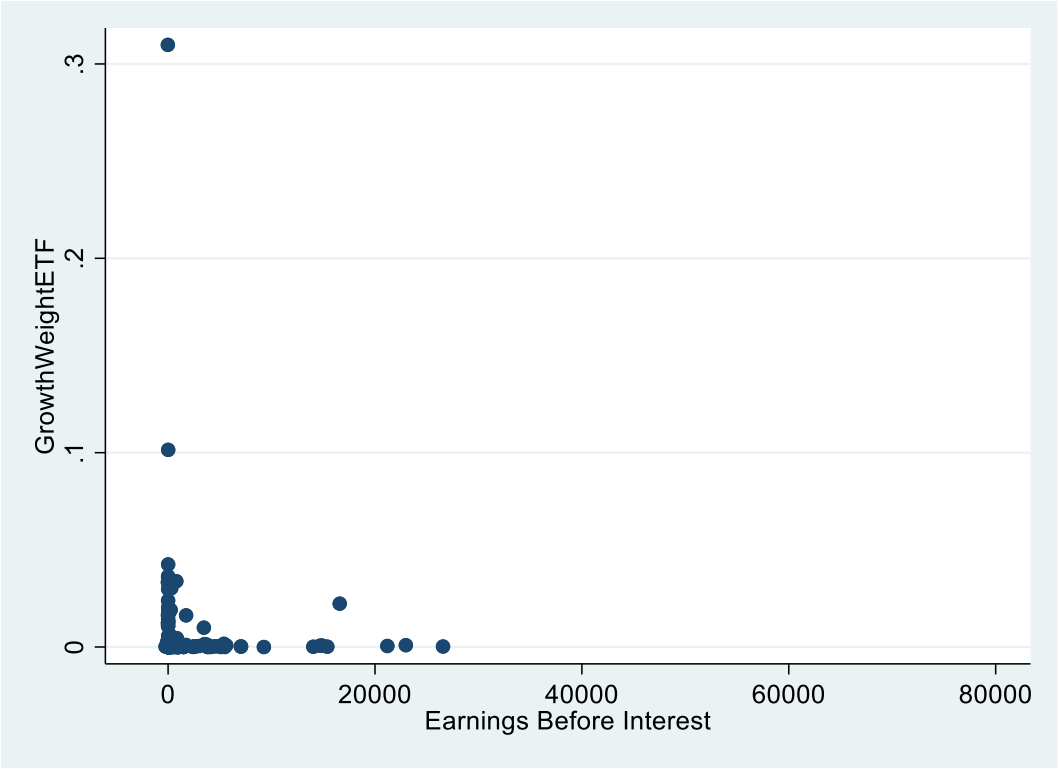
EBIT	Earnings before Interest and Taxes
EBITDA	Earnings before Interest, Taxes, Depreciation and Amortization
Ni	Net earnings
Oancf	Amount of cash flow from operations
Fcf	Amount of free cash flows
SicN	Industry represented by SIC code
GrowthWeightETF	Weighted average of the growth of ETFs
Ln_xrd	Natural logarithm of the amount of intangible assets
Rel_rdintern	Relative amount of R&D investment
A	Amount of restatements
Fyear	Year
e2	e^2, with 'e' being the residuals (used for heteroscedasticity testing)

Table 2 Shapiro-Wilk test for normality + plots with independent and dependent variables

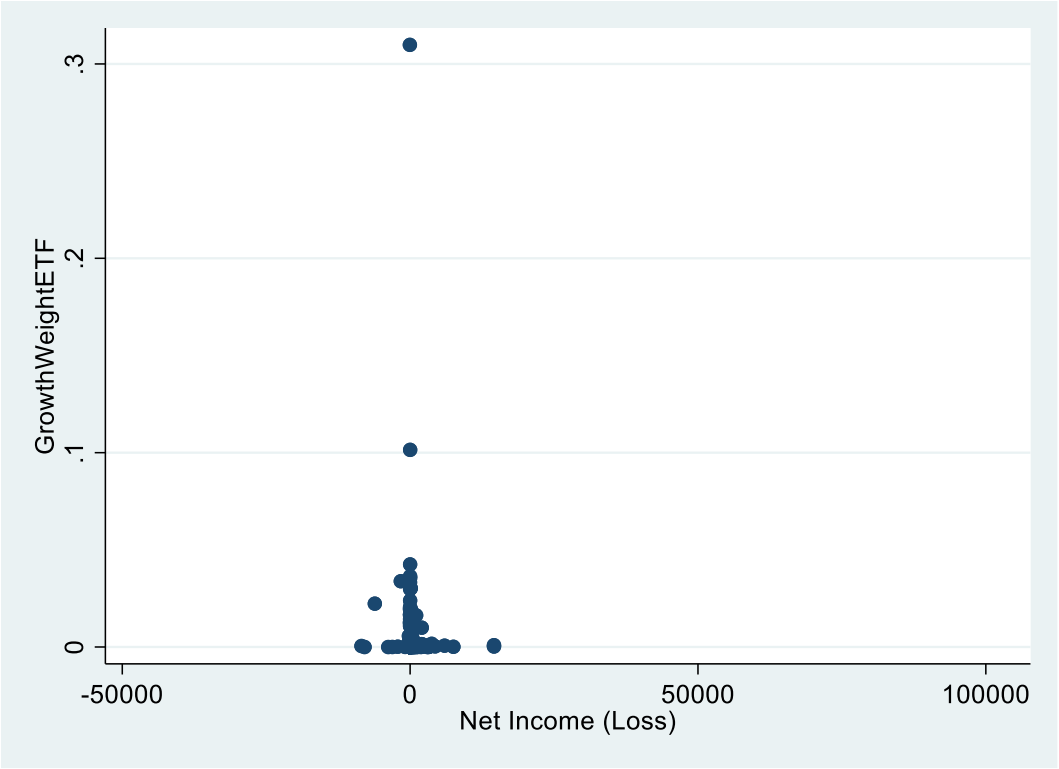
Shapiro-Wilk test for normal data					
Variable	Observations	W	V	z	Probability>z
GrowthWeightETF	1,168	0.16722	605.060	15.962	0.00000
EBIT	12,653	0.25917	4513.239	22.666	0.00000
A	25,686	0.90275	1066.760	19.093	0.00000
Ln_xrd	12,609	0.99695	18.524	7.862	0.00000
Rel_rdinten	12,494	0.01153	5957.324	23.406	0.00000
Fyear	12,653	0.78449	1312.887	19.340	0.00000
SicN	12,653	0.98414	96.595	12.311	0.00000
Note: The normal approximation to the sampling distribution of W' is valid for 4<=n<=2000.					



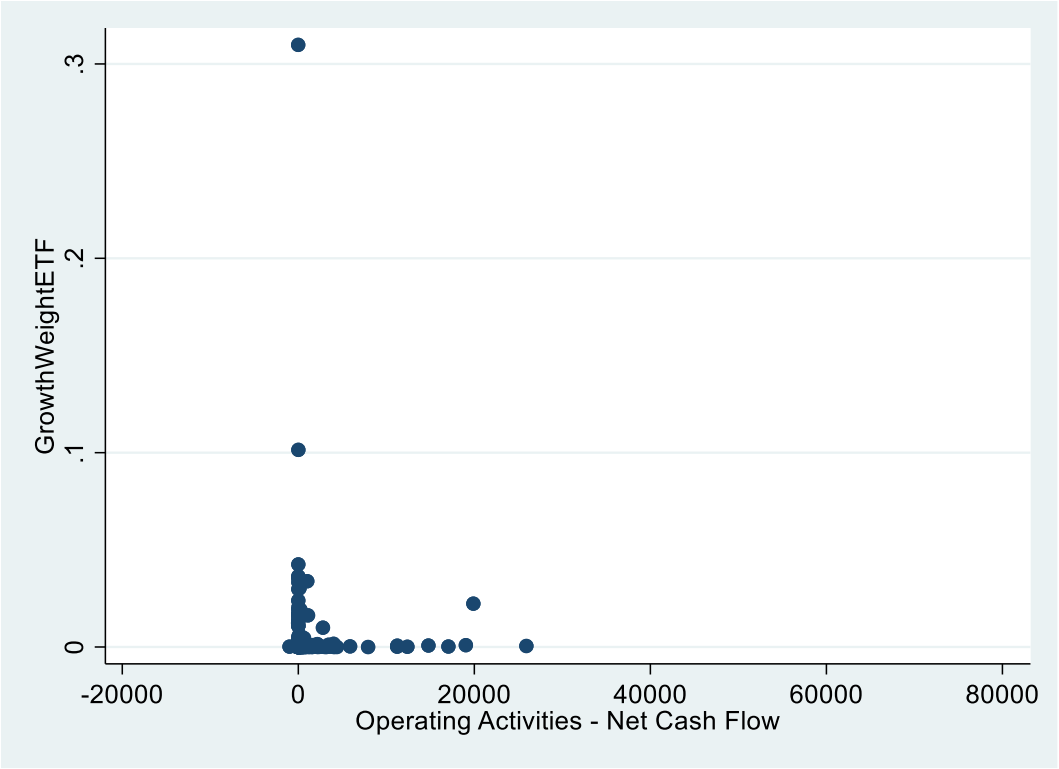
Shapiro-Wilk test for normal data					
Variable	Observations	W	V	Z	Probability>z
GrowthWeightETF	1,168	0.16722	605.060	15.962	0.00000
EBITDA	12,653	0.26825	4457.906	22.633	0.00000
A	25,686	0.90275	1066.760	19.093	0.00000
Ln_xrd	12,609	0.99695	18.524	7.862	0.00000
Rel_rdinten	12,494	0.01153	5957.324	23.406	0.00000
Fyear	12,653	0.78449	1312.887	19.340	0.00000
SicN	12,653	0.98414	96.595	12.311	0.00000
Note: The normal approximation to the sampling distribution of W' is valid for 4<=n<=2000.					



Shapiro-Wilk test for normal data					
Variable	Observations	W	V	z	Probability>z
GrowthWeightETF	1,168	0.16722	605.060	15.962	0.00000
Ni	12,653	0.23040	4688.463	22.769	0.00000
A	25,686	0.90275	1066.760	19.093	0.00000
Ln_xrd	12,609	0.99695	18.524	7.862	0.00000
Rel_rdinten	12,494	0.01153	5957.324	23.406	0.00000
Fyear	12,653	0.78449	1312.887	19.340	0.00000
SicN	12,653	0.98414	96.595	12.311	0.00000
Note: The normal approximation to the sampling distribution of W' is valid for 4<=n<=2000.					



Shapiro-Wilk test for normal data					
Variable	Observations	W	V	z	Probability>z
GrowthWeightETF	1,168	0.16722	605.060	15.962	0.00000
Oancf	12,653	0.25752	4523.288	22.672	0.00000
A	25,686	0.90275	1066.760	19.093	0.00000
Ln_xrd	12,609	0.99695	18.524	7.862	0.00000
Rel_rdinten	12,494	0.01153	5957.324	23.406	0.00000
Fyear	12,653	0.78449	1312.887	19.340	0.00000
SicN	12,653	0.98414	96.595	12.311	0.00000
Note: The normal approximation to the sampling distribution of W' is valid for 4<=n<=2000.					



Shapiro-Wilk test for normal data					
Variable	Observations	W	V	z	Probability>z
GrowthWeightETF	1,168	0.16722	605.060	15.962	0.00000
Fcf	12,653	0.24374	4607.199	22.721	0.00000
A	25,686	0.90275	1066.760	19.093	0.00000
Ln_xrd	12,609	0.99695	18.524	7.862	0.00000
Rel_rdinten	12,494	0.01153	5957.324	23.406	0.00000
Fyear	12,653	0.78449	1312.887	19.340	0.00000
SicN	12,653	0.98414	96.595	12.311	0.00000
Note: The normal approximation to the sampling distribution of W' is valid for 4<=n<=2000.					

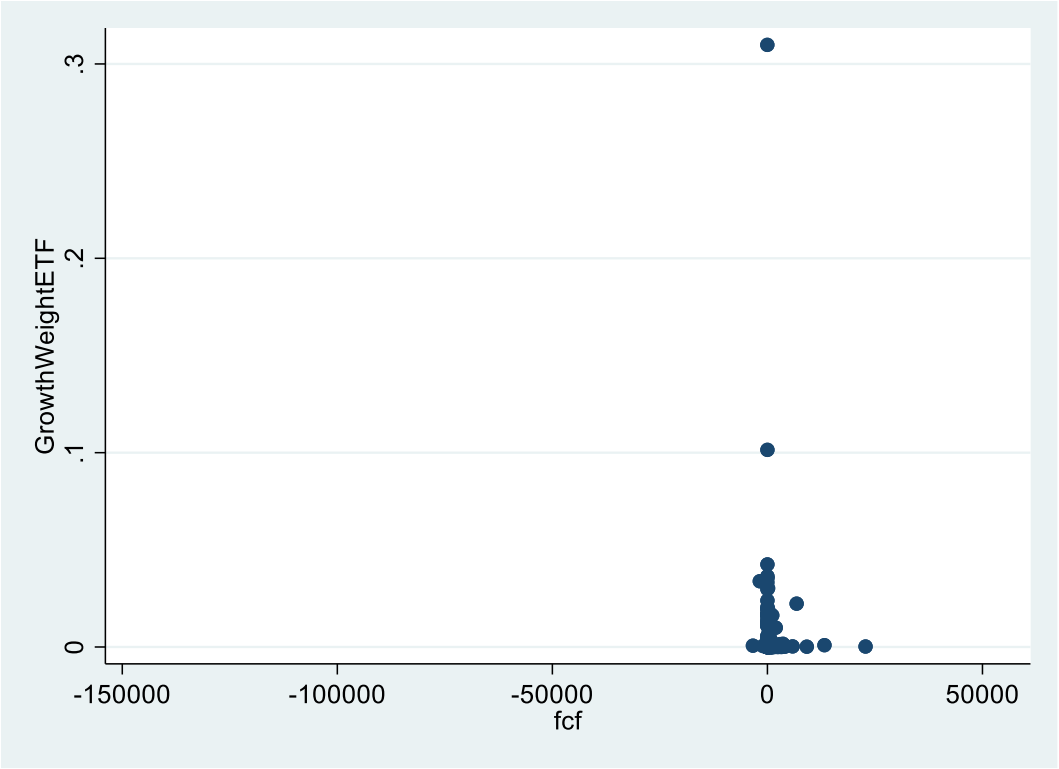
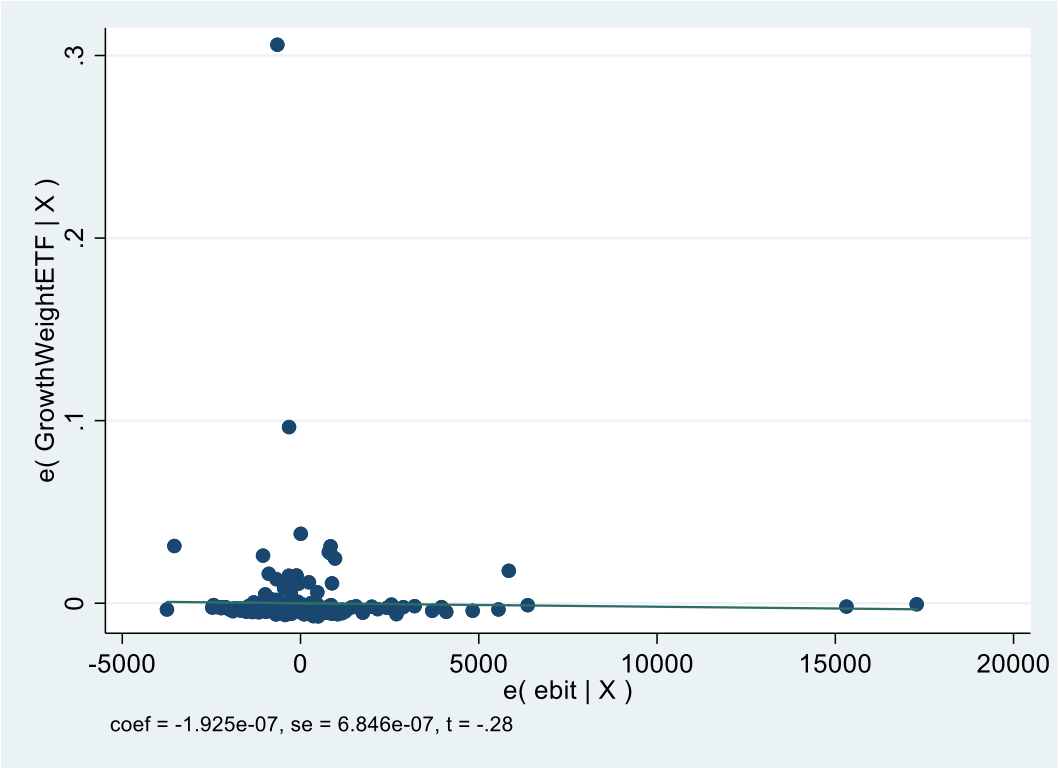


Table 3 Heteroscedasticity (Breusch-Pagan) tests with residual plots

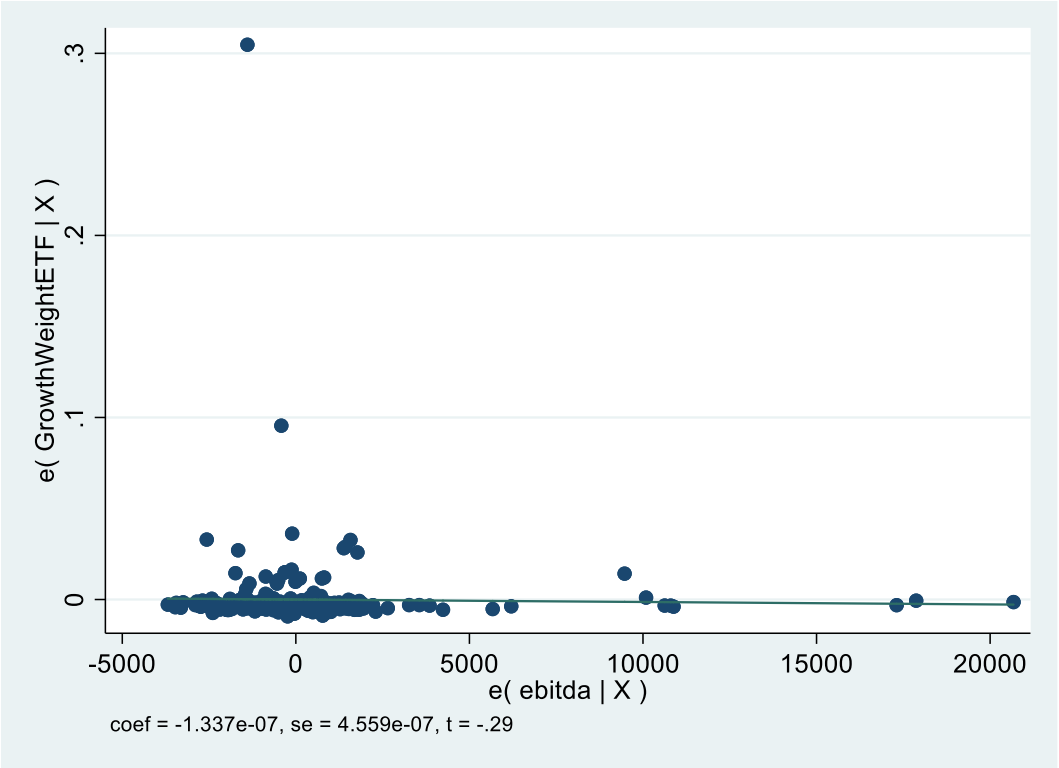
Number of observations	254
F (7 ,246)	253.11
Prob > F	0.0000
R-squared	0.8781
Adjusted R-squared	0.8746
Root MSE	0.00209

e2	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
GrowthWeightETF	0.2589824	0.0061792	41.91	0.000	0.2468115	0.2711534
EBIT	-1.47e-08	6.67e-08	-0.22	0.826	-1.46e-07	1.17e-07
A	-0.0001892	0.001284	-1.47	0.142	-0.0004422	0.0000638
Ln_xrd	0.0000742	0.0000702	1.06	0.291	-0.000064	0.0002125
Rel_rdinten	3.24e-08	1.05e-07	0.31	0.758	-1.74e-07	2.39e-07
SicN	-1.24e-07	7.29e-08	-1.71	0.089	-2.68e-07	1.91e-08
Fyear	0.0000829	0.0000519	1.60	0.112	-0.0000194	0.0001852
_cons	-0.1661317	0.1044407	-1.59	0.113	-0.3718438	0.0395804



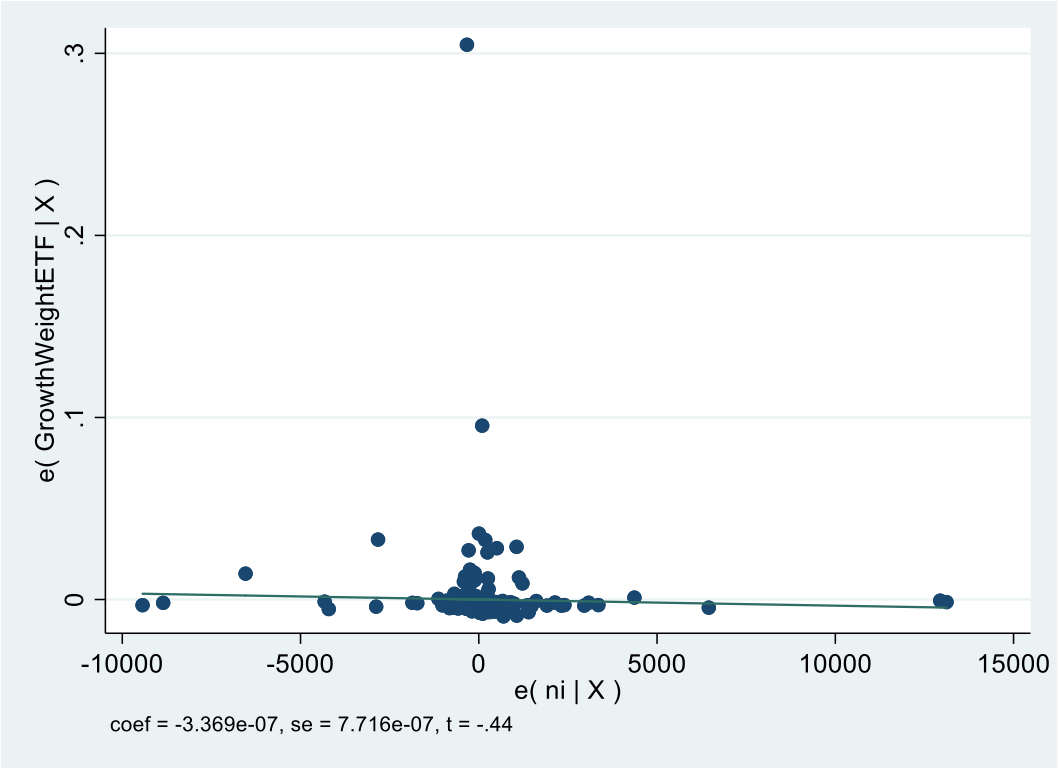
Number of observations	254
F (7,246)	253.42
Prob > F	0.0000
R-squared	0.8782
Adjusted R-squared	0.8747
Root MSE	0.00209

e2	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
GrowthWeightETF	0.2589442	0.0061759	41.93	0.000	0.2467798	0.2711086
EBITDA	-2.46e-08	4.43e-08	-0.56	0.578	-1.12e-07	6.25e-08
A	-0.0001832	0.001284	-1.43	0.155	-0.0004362	0.0000697
Ln_xrd	0.0000857	0.0000706	1.21	0.226	-0.0000535	0.0002248
Rel_rdinten	3.13e-08	1.05e-07	0.30	0.765	-1.75e-07	2.38e-07
SicN	-1.25e-07	7.28e-08	-1.72	0.086	-2.69e-07	1.80e-08
Fyear	0.0000852	0.0000521	1.64	0.103	-0.0000174	0.0001878
_cons	-0.170789	0.1047584	-1.63	0.104	-0.3771268	0.0355487



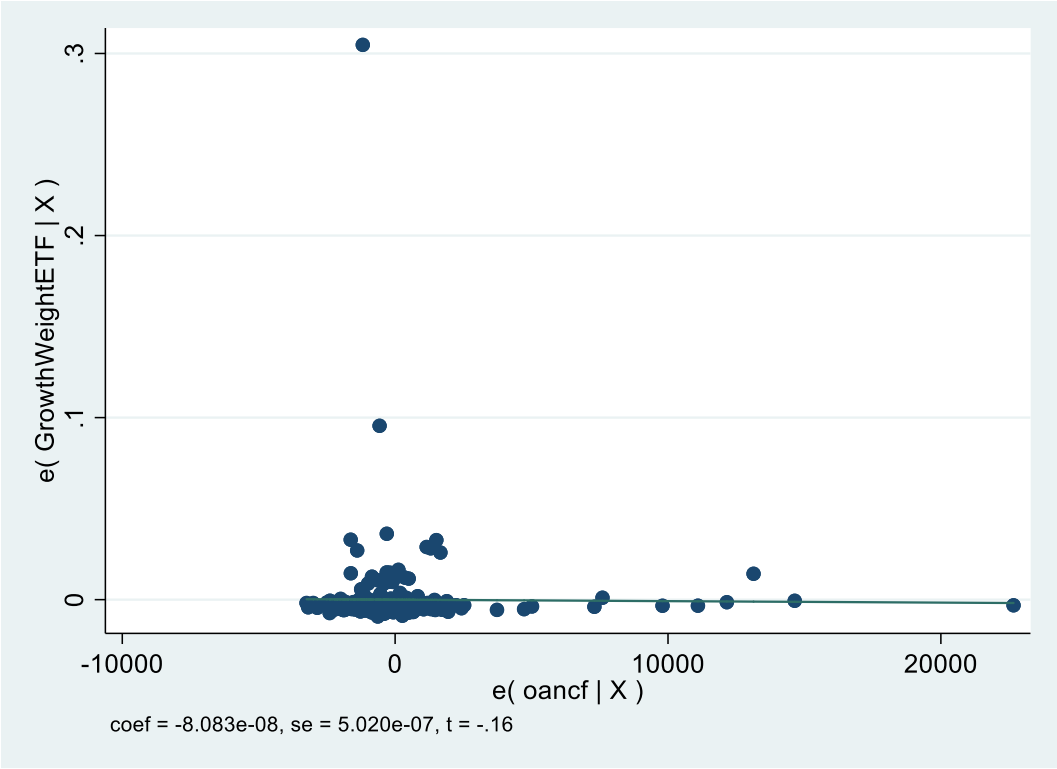
Number of observations	254
F (7 ,246)	253.46
Prob > F	0.0000
R-squared	0.8784
Adjusted R-squared	0.8749
Root MSE	0.00209

e2	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
GrowthWeightETF	0.2591099	0.0061767	41.95	0.000	0.2469439	0.2712758
Ni	4.43e-08	7.49e-08	0.59	0.555	-1.03e-07	1.92e-07
A	-0.0001815	0.0001278	-1.45	0.148	-0.0004373	0.0000663
Ln_xrd	0.0000573	0.0000644	0.89	0.374	-0.0000695	0.0001841
Rel_rdinten	3.35e-08	1.05e-07	0.32	0.750	-1.73e-07	2.40e-07
SicN	-1.24e-07	7.28e-08	-1.70	0.090	-2.67e-07	1.94e-08
Fyear	0.0000839	0.0000519	1.61	0.108	-0.0000184	0.001862
_cons	-0.1681041	0.1044386	-1.61	0.109	-0.373812	0.0376039



Number of observations	254
F (7 ,246)	253.85
Prob > F	0.0000
R-squared	0.8784
Adjusted R-squared	0.8749
Root MSE	0.00209

e2	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
GrowthWeightETF	0.2589562	0.0061705	41.97	0.000	0.2468025	0.27111
Oancf	-4.02e-08	4.87e-08	-0.83	0.410	-1.36e-07	5.57e-08
A	-0.0001771	0.001286	-1.38	0.170	-0.0004304	0.0000761
Ln_xrd	0.0000924	0.0000693	1.33	0.184	-0.0000442	0.000229
Rel_rdinten	3.09e-08	1.05e-07	0.30	0.768	-1.75e-07	2.37e-07
SicN	-1.23e-07	7.28e-08	-1.69	0.093	-2.66e-07	2.05e-08
Fyear	0.0000873	0.0000522	1.67	0.095	-0.0000154	0.0001901
_cons	-0.1751092	0.1049039	1.67	0.096	-0.3817335	0.0315152



Number of observations	254
F (7 ,246)	253.06
Prob > F	0.0000
R-squared	0.8781
Adjusted R-squared	0.8746
Root MSE	0.00209

e2	Coefficient	Standard Error	T	P>[t]	95% Confidence Interval	
GrowthWeightETF	0.2589995	0.00618	41.91	0.000	0.2468269	0.271172
Fcf	-5.07e-09	7.43e-08	-0.07	0.946	-1.51e-07	1.41e-07
A	-0.0001919	0.001284	-1.49	0.136	-0.0004448	0.0000611
Ln_xrd	0.0000688	0.0000671	1.02	0.307	-0.0000634	0.0002009
Rel_rdinten	3.28e-08	1.05e-07	0.31	0.755	-1.74e-07	2.39e-07
SicN	-1.25e-07	7.29e-08	-1.71	0.088	-2.68e-07	1.88e-08
Fyear	0.0000829	0.000052	1.59	0.112	-0.0000195	0.0001854
_cons	-0.1662071	0.1046073	-1.59	0.113	-0.3722472	0.039833

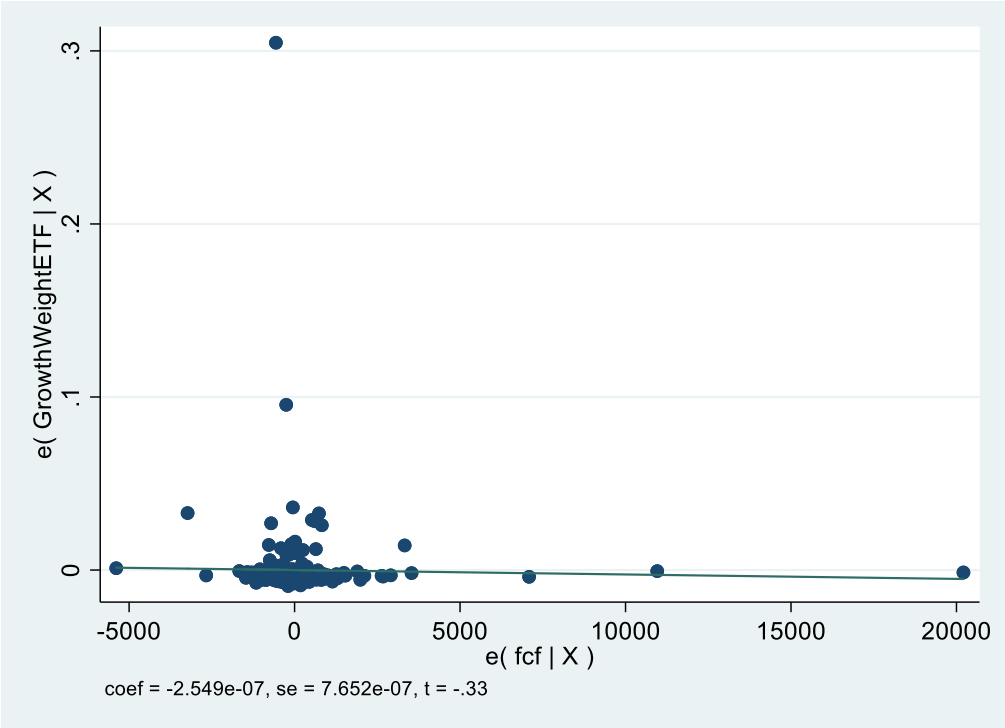


Table 4 Autocorrelation (Breusch-Godfrey) tests

Breusch-Godfrey LM test for autocorrelation (first regression with EBIT)			
Lags (p)	Chi^2	Df	Prob > chi^2
1	0.128	1	0.7208
H0: no serial correlation			

Breusch-Godfrey LM test for autocorrelation (first regression with EBITDA)			
Lags (p)	Chi^2	df	Prob > chi^2
1	0.125	1	0.7237
H0: no serial correlation			

Breusch-Godfrey LM test for autocorrelation (third regression with Net earnings)			
Lags (p)	Chi^2	df	Prob > chi^2
1	0.127	1	0.7217
H0: no serial correlation			

Breusch-Godfrey LM test for autocorrelation (fourth regression with Cash flow from operations)			
Lags (p)	Chi^2	df	Prob > chi^2
1	0.118	1	0.7312
H0: no serial correlation			

Breusch-Godfrey LM test for autocorrelation (fifth regression with Free cash flows)			
Lags (p)	Chi^2	df	Prob > chi^2
1	0.121	1	0.7281
H0: no serial correlation			