HOW EFFICIENT IS THE PRICING OF ETFs? COMPARING THE PRICING OF ETFs WITH THE PRICING OF THE UNDERLYING STOCKS AND THE INDEX

Lappeenranta–Lahti University of Technology LUT

Master's Programme in Strategic Finance and Business Analytics, Master's thesis

2022

Mika Jaakkola

Examiner(s): Professor, Mikael Collan

Jan Stoklasa, D.Sc. (Tech.)

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT School of Business and Management

Business Administration

Mika Jaakkola

How Efficient is the Pricing of ETFs? Comparing the Pricing of ETFs With the Pricing of the Underlying Stocks and the Index

Master's thesis

2022

63 pages, 6 figures, 11 tables and 8 appendices

Examiner(s): Professor Mikael Collan and Jan Stoklasa, D.Sc. (Tech)

Keywords: Exchange Traded Fund, Net Asset Value, Pricing Efficiency, Mispricing

Exchange Traded Funds (ETF) have become a vital part of the financial markets during the years, in Europe as well. This master's thesis studies the deviations between the exchange traded fund's price and the respective net asset value (NAV). Also, the return difference of the exchange traded fund and its respective index was examined. Nine European exchange traded funds was selected for a time period of four-years from February 2015 to February 2019. Average mispricing was examined with t-test and regression analysis, and the persistency of mispricing was analyzed with autoregressive model. The return difference between the ETF price and the index price was examined with three tracking error methods which were absolute return difference, standard deviation of absolute return difference and standard error of a regression.

The obtained results suggest that statistically significant mispricing is present among the selected ETFs, although the average price of the ETFs follow closely to the respective NAV. Autoregression suggest that the mispricing is persistent for at least two days which exposes the ETFs for arbitrage opportunities, but the magnitude of mispricing is a barrier for arbitrage, because the bid-ask spread widens. For this reason, the mispricing is not economically significant. Regression analysis shows strong and close relationship between the ETF price and the respective NAV. In relation to previous research, tracking error calculations show significant tracking error, meaning that the ETFs do not replicate their respective index perfectly. With the obtained results it can be said that ETFs are efficiently priced, but ETFs are not replicating their respective indexes perfectly and exhibit significant tracking errors.

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT-kauppakorkeakoulu

Kauppatieteet

Mika Jaakkola

Kuinka tehokasta pörssinoteerattujen rahastojen hinnoittelu on? Pörssinoteerattujen rahastojen hintojen vertailu sisältävien osakkeiden hintaan ja osakeindeksiin

Kauppatieteiden pro gradu -tutkielma

2022

63 sivua, 6 kuvaa, 10 taulukkoa ja 8 liitettä

Tarkastaja(t): Professori Mikael Collan ja D.Sc. Jan Stoklasa

Avainsanat: Pörssinoteerattu rahasto, nettoarvo, hinnoitteluvirhe, hinnoittelun tehokkuus

Pörssinoteeratut rahastot ovat vuosien saatossa nousseet tärkeäksi osaksi rahoitusmaailmaa, myös Euroopassa. Tämä pro gradu -tutkielma keskittyy tarkastelemaan pörssinoteerattujen rahastojen hintojen sekä niiden hallussa olevien osakkeiden arvojen eroa eli hinnoitteluvirhettä. Tämän lisäksi rahaston tuottoa verrataan osakeindeksiin, jota rahasto seuraa. Tarkasteluun valikoitui yhteensä yhdeksän eurooppalaista pörssinoteerattua rahastoa ja tarkastelu ajanjakso sijoittuu neljän vuoden aikavälille helmikuu 2015 ja helmikuu 2019. Keskimääräistä hinnoitteluvirhettä mitataan t-testillä sekä regressio analyysillä ja hinnoitteluvirheiden kestoa testataan autoregressiivisella mallilla. Rahaston ja osakeindeksin tuottojen eroa mitataan kolmella seurantavirhe-menetelmällä, jotka ovat tuottojen itseisarvojen keskimääräinen ero, näiden erojen keskihajonta sekä regressiomallin tuottama keskivirhe.

Saadut tulokset osoittavat, että tutkielmaan valituissa pörssinoteeratuissa rahastoissa esiintyy tilastollisesti merkittävää hinnoitteluvirhettä, vaikka keskimääräisesti hinnat ovat lähellä nettoarvoa. Autoregressivinen malli osoitti, että hinnoitteluvirhe kestää ainakin kaksi päivää, joka antaa mahdollisuuden arbitraasille, mutta korkea hinnoitteluvirhe on esteenä arbitraasin harjoittamiselle, koska ero tarjotun ja kysytyn hinnan välillä kasvaa. Tämän takia hinnoitteluvirhe ei ole taloudellisesti merkittävää. Regressio analyysi osoitti, että pörssinoteeratun rahaston ja nettoarvon välillä on vahva lineaarinen yhteys. Seurantavirhe osoittaa aikaisempiin tutkielmiin viitaten, että rahastot eivät seuraa täydellisesti indeksiä ja että seurantavirhe on merkittävää. Saatujen tulosten pohjalta voidaan todeta, että pörssinoteeratut rahastot ovat tehokkaasti hinnoiteltuja, mutta rahaston ja indeksin välillä esiintyy merkittävää seurantavirhettä.

ACKNOWLEDGEMENTS

As this thesis is now complete, I would like to thank and show how grateful I am towards those who helped me to reach this point.

First, I would like to thank Mikael Collan for the help and guidance during this process as well as the second examiner Jan Stoklasa. I would also like to thank all the professors who I have met during my studies.

I would also like to thank the friends I have had the pleasure to meet during my time in Lappeenranta. I will cherish the shared moments forever.

Finally, I would like to thank my friends and family who have supported me from the start to finish.

Mika Jaakkola

23.11.2022 Helsinki

Table of Content

LIST OF ABBREVIATIONS

ETF	Exchange-traded fund
NAV	Net asset value
AUM	Assets under management
EMH	Efficient market hypothesis
AP	Authorized participant
AR	Autoregression
OLS	Ordinary Least Squares

1 Introduction

Exchange-Traded Funds, commonly known by the abbreviation ETFs, have grown exponentially during the last decade and become popular alternative for traditional investments. An ETF is a publicly traded investment vehicle that typically replicate a stock index. They offer multiple benefits for investors and these benefits are the reason for the ever-increasing demand. If an ETF replicates a stock index, it holds the same securities as the respective index. In theory, the value of an ETF should approximately be equivalent to the value of the underlying assets, which is also referred as the net asset value (NAV) of an ETF. Because ETFs are traded publicly in a stock market, the price fluctuates throughout the day as the price of the underlying shares does. This can cause significant deviation between the ETF and the underlying asset also called by the phenomenon of mispricing or pricing efficiency/inefficiency. Pricing inefficiencies can expose arbitrage opportunities for investors, depending on the magnitude of the mispricing and how persistent it is. Stock index replication ties the ETF's performance to the performance of the respective index. If the ETF does not obtain returns of same magnitude as the index, it is an indication that one should not invest in the ETF as the underlying assets obtain greater returns. This phenomenon is called tracking error. The purpose of this thesis is to study the phenomenon of ETF pricing efficiency and the ETF performance. This study aims to detect whether significant mispricing or tracking error occurs by using well known methods. The focus lies on ETFs domiciled in Europe as European ETF market is one of the fastest developing markets, yet there is very little research of ETFs domiciled in Europe compared to the US domiciled for instance.

Previous research has studied the mispricing of ETFs from the early stages of the ETF industry. Therefore, mispricing is not a new phenomenon. The early studies from Ackert and Tian (2000) and Elton et al. (2002) examined the mispricing of ETFs in the US. Both studies found the ETFs to be efficiently priced on average. Arbitrage opportunities did not arise as mispricing was not found to be significantly persistent. Pervious literature also show that larger deviations occur from time to time. ETFs that replicate international indexes have a tendency to present larger price deviations other alternative asset classes (Engle and Sarkar, 2006). A more recent study of Petäjistö (2017) found that there are arbitrage opportunities among ETFs.

Whereas mispricing is not a new phenomenon, tracking error has been examined by the academic research since Pope and Yadav (1994) as similar methods have been used for mutual funds. Studies such as Shin and Soydemir (2010) and Blitz and Huij (2012) show significant tracking error for ETFs in the US, and on average they seem to underperform their respective index. Almelu and Goya (2022) found significant tracking error in the more volatile markets of India and the ETFs also overperformed their respective indexes. As can be seen, tracking error is not market related and it can be present in any market.

The exponential growth of ETFs has attracted the academic research. As the amount of academic research on ETF pricing efficiency has grown, still most of the research focuses on the US domiciled ETFs. As ETF markets in Europe are growing in a similar space as the global ETF markets, it is necessary to extend the already existing literature of ETFs domiciled in Europe. The methods used in this thesis are similar as previous studies have used to study European domiciled ETFs. Most of the previous literature focuses on one market as this study uses ETFs from different European markets such as Norway, France and the UK. Also, many studies focus on either to study the performance of the selected ETFs compared to the respective indexes or purely on the pricing efficiency of the ETFs. This thesis conducts both methods.

1.1 Objectives

According to efficient market hypothesis (EMH) asset prices should fully reflect all available information and this applies to ETFs as well. Previous studies have shown that there is mispricing to be found among ETFs. Previous studies have also shown that ETFs exhibit significant tracking error, and that it is not market related. The purpose of this study is to discover the state of mispricing and its magnitude in European markets. The answer to the overall research question can be found by studying the following objectives:

1. To investigate the presence of mispricing by observing the difference between closing price and NAV in terms of premiums and discounts.

Previous academic research has found significant deviations between ETF trading price and its NAV in several different markets. Since ETFs are tracking the performance of another index or asset, the value of an ETF should be close to its respective NAV. Because both ETFs and the underlying assets are publicly traded, ETFs should have a trading price that is close to NAV. Despite this reasoning there has been evidence of both economically and statistically significant price deviations. Evidence also suggests that premiums/discounts to can be persistent for number of days which is an indication of consistent pricing inefficiency. The purpose is to examine whether ETFs traded in Europe are subject to this kind of phenomenon.

2. To investigate if there is evidence of persistent mispricing

ETF mispricing can present opportunities to obtain excess profits. Previous research has obtained persistent mispricing which exposes the ETF to arbitrage opportunities. Arbitrage profits can be obtained with relatively simple trading strategies based on exploiting the mispricing of the ETF compared to its NAV. If ETFs in European markets were efficiently priced, any attempt to profit from mispricing should not yield any excess returns.

3. To measure how strong is the relationship between ETF price and the respective NAV

This research objective is important as this study seeks to find answers to how different the ETF price is to its respective NAV. It is important to understand the relationship between the ETF price and its respective NAV and how the changes in NAV effects to the price of the ETF. Previous studies have found the NAV to be part-reason for mispricing.

4. To investigate whether ETFs exhibit significant tracking error.

For an investor, tracking error is one of the most vital sources of information when choosing to invest into an ETF. With tracking error, one can determine how well the ETF is performing compared to the respective index it replicates. Even portfolios, that are perfectly indexed against an index, behave differently than the respective index, even though this difference may be slight. The purpose is to determine how well the selected ETFs are tracking their respective index.

1.2 Structure

This thesis is structured as follows. First, general information about ETFs and background of pricing efficiency and ETF pricing mechanism are discussed in section two, which also includes literature review of previous studies on ETF mispricing and ETF performance. The next section is a description for the data sample used in this thesis. In the fourth section the used models are introduced in detail and the reasoning for their usage. The fifth section presents empirical results. The final section contains summary & conclusion of the results and suggestions for further research.

2 Theoretical framework

2.1 General information about ETFs

ETFs are publicly traded investment vehicles or different pools of investments which intend to replicate some specific index, bonds, commodities or a basket of assets like a mutual fund. (Kostovetsky, 2003) ETFs can be structured to even track a specific investment strategy. ETFs combines the characteristics of close-end funds and traditional open-end mutual funds. Unlike mutual funds, which can be traded (acquired or redeemed) only once a day, ETFs can be traded continuously throughout the day as any other security in the stock market, and therefore the price of an ETF can change throughout the day. (Deville, 2008)

ETFs offer a variety of benefits that are attractive for investors. One of these benefits is lower trading costs. As ETFs intend to replicate a specific index or a basket of assets, less active management is required, because ETFs tend to passively replicate their respective asset class chosen, whereas mutual funds require active management as they try to overperform the respective index, and therefore are more expensive for investors to invest in. The fact that ETFs replicate a benchmark index introduces us the next benefit for investors which is tax efficiency. As ETFs passively replicates their respective benchmark, they tend to also to have a lower portfolio turnover than mutual funds which are actively trying to beat the benchmark. Regarding other benefits that ETFs offer for investors, by investing into an ETF one can obtain exposure to well diversified indexes. Instead of buying multiple stocks, investing into even one ETF could gain exposure to markets that were not possible for an average investor to invest in. Some foreign markets have a variety of requirements for investors to even obtain the ability to invest in that market, such as a local bank account or a local custodian. From this perspective the possibility for diversification has a strong link to lower trading cost. Lastly, compared to mutual funds, ETFs are more flexible and transparent. The flexibility benefit is obtained by the fact that ETFs can be traded in stock exchange throughout the day like any other traditional stock at market-determined prices. Most ETFs reveal their holdings daily, which makes it possible for investors to compute the NAV, if not provided by the distributor. (Foucher & Gray, 2014; Hill et al. 2015)

The first ETFs were introduced to the world in the early 90s in the U.S. stock market as American Stock Exchange listed the so called "spider" to track the S&P 500 index. The first European ETF was listed in the Deutsche Börse in 2000 which replicated the Stoxx Europe 50 and Euro Stoxx 50 indexes. (Poterba & Shoven, 2002; Ben-David et al., 2012). Ever since the first ETF was introduced, the above-mentioned benefits that the ETFs provide compared to mutual funds or traditional stocks, have been the reason for the rapid growth of ETFs in the financial markets. These benefits do not only attract new investors to the market, but new funds as well. The growth of ETFs has been exponential for past decade as can be seen from figure 1 which illustrates the assets under management (AUM) of global ETFs from 2003 to 2021. The AUM have almost tripled from 2016 to 2021 as global ETFs held over AUM worth of over 10 trillion USD.

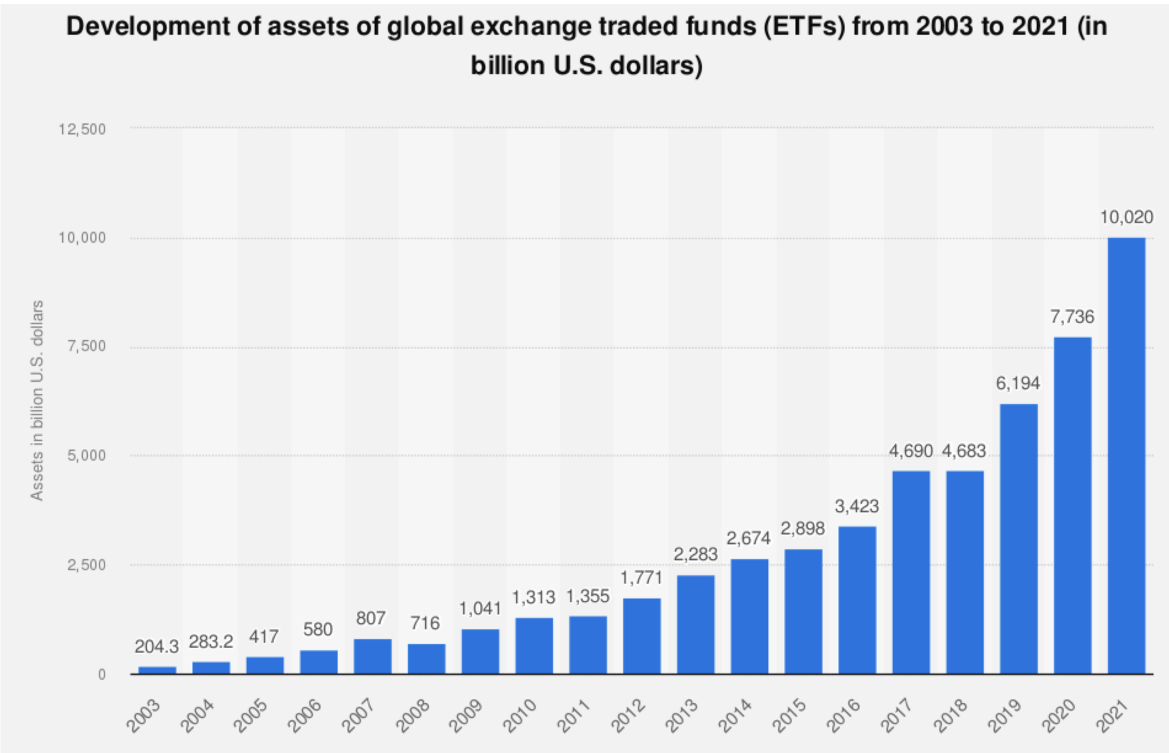

Figure 1. Development of AUM of global ETFs from 20023 to 2021 (Statista, 2022)

The ETF markets in Europe have been developing in a similar pace in terms of AUM as global ETFs as can be seen from table 2. The total AUM almost tripled in European markets as well as ETFs listed in Europe had over 500 billion USD of total AUM in 2016 and in 2021 the value was over 1,5 trillion USD. According to a survey conducted by Pricewater-houseCooper (2022) European markets are one of the fastest growing and the prediction is that European markets reaches at least 3 trillion USD AUM until 2026. Similarly, it is believed that global ETF AUM will reach 18 trillion until 2026.

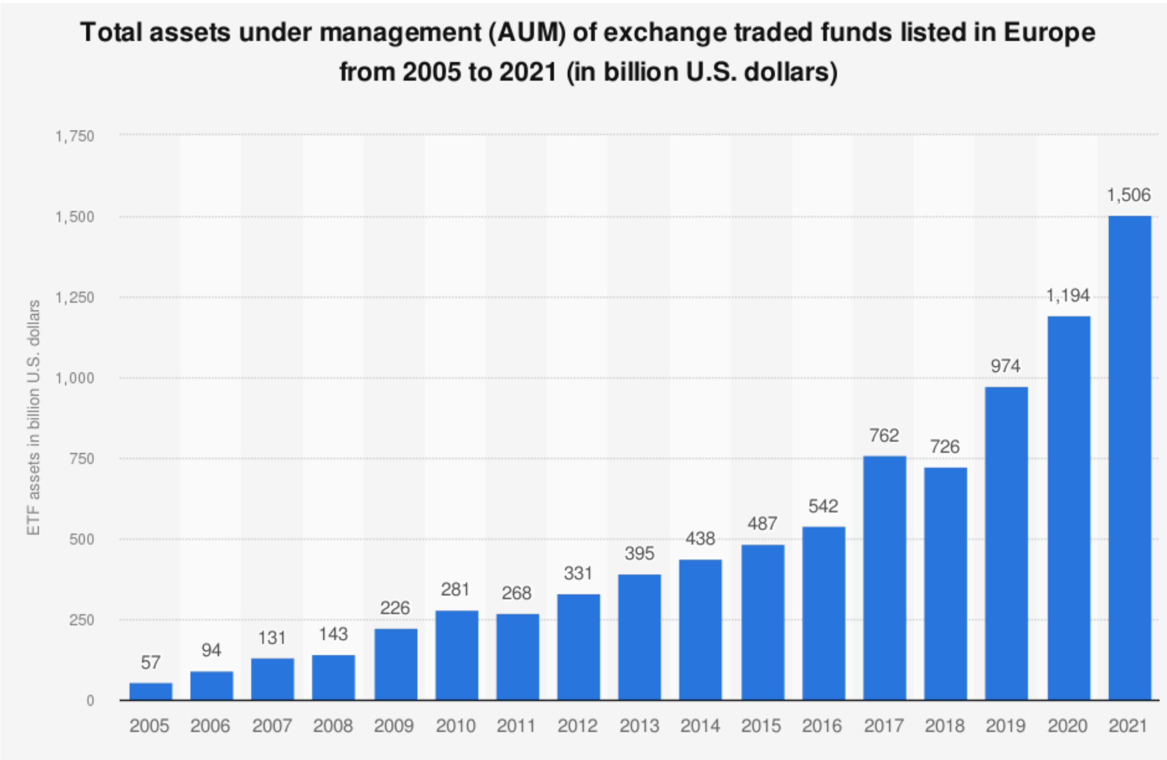

Figure 2. Development of AUM of ETFs in Europe from 2005 to 2021 (Statista, 2022)

2.2 ETF Pricing and Creation/Redemption Process

ETFs can be divided into two groups depending on how they replicate the benchmark index, physical ETFs and synthetic ETFs. Physical ETF replicates the benchmark index as closely as possible by mimicking the weights of the underlying assets. Synthetic ETF tracks the benchmark index with derivative contracts such as swaps. Even though both physical and synthetic ETFs require collateral from their counterparty, they are still prone to different sources of counterparty risk. As physical ETFs engage in security lending, it exposes the fund to risk of default by the borrower. On the other hand, synthetic ETFs are exposed to default of the counterparty in the derivative contract. (Naumenko and Chystiakova, 2015; Ben-David et al. 2016)

The ETF market can be divided in two markets, primary market and secondary market. Primary market is the market where ETF shares are created and redeemed. This process involves large institutions also called as authorized participants (AP). Such institutions in Europe are for example Morgan Stanley, Citi Bank, UBS and Deutsche Bank. The process of creation and redemption is illustrated step by step in figure 3. The creation process starts with an AP and an ETF issuer. The flow of the process does not change whether the issuer wants to create new shares to meet demand or whether the issuer wants to create a whole new ETF. The role of an AP is to typically act as a market maker and liquidity provider. If the issuer wants to create new shares to meet the demand, the issuer publishes a creation basket which works a list of the securities it is willing to accept for creating new shares of the ETF. This is repeated daily. After the creation basket is published, an AP is the one to acquire the needed securities from the market. An AP that is willing to acquire and deliver the needed securities is in exchange given a bundle of new shares of the ETF. This bundle is also referred as creation unit which consists of a large quantity of ETF shares, ranging from 25 000 to 200 000 shares. This process is also called as in-kind process. If a security or a creation unit is hard for the AP to acquire, the issuer may also permit cash for the AP as a substitute. On the contrary, AP could also be charged a fee as an adjustment to offset possible transaction costs. The redemption process works similarly as the creation process but the other way around. Whereas securities were acquired from the market and put into the ETF, now the ETF is unwrapped back to securities. (Deville, 2008; Blackrock, 2017; Antoniewicz and Heinrichs, 2014)

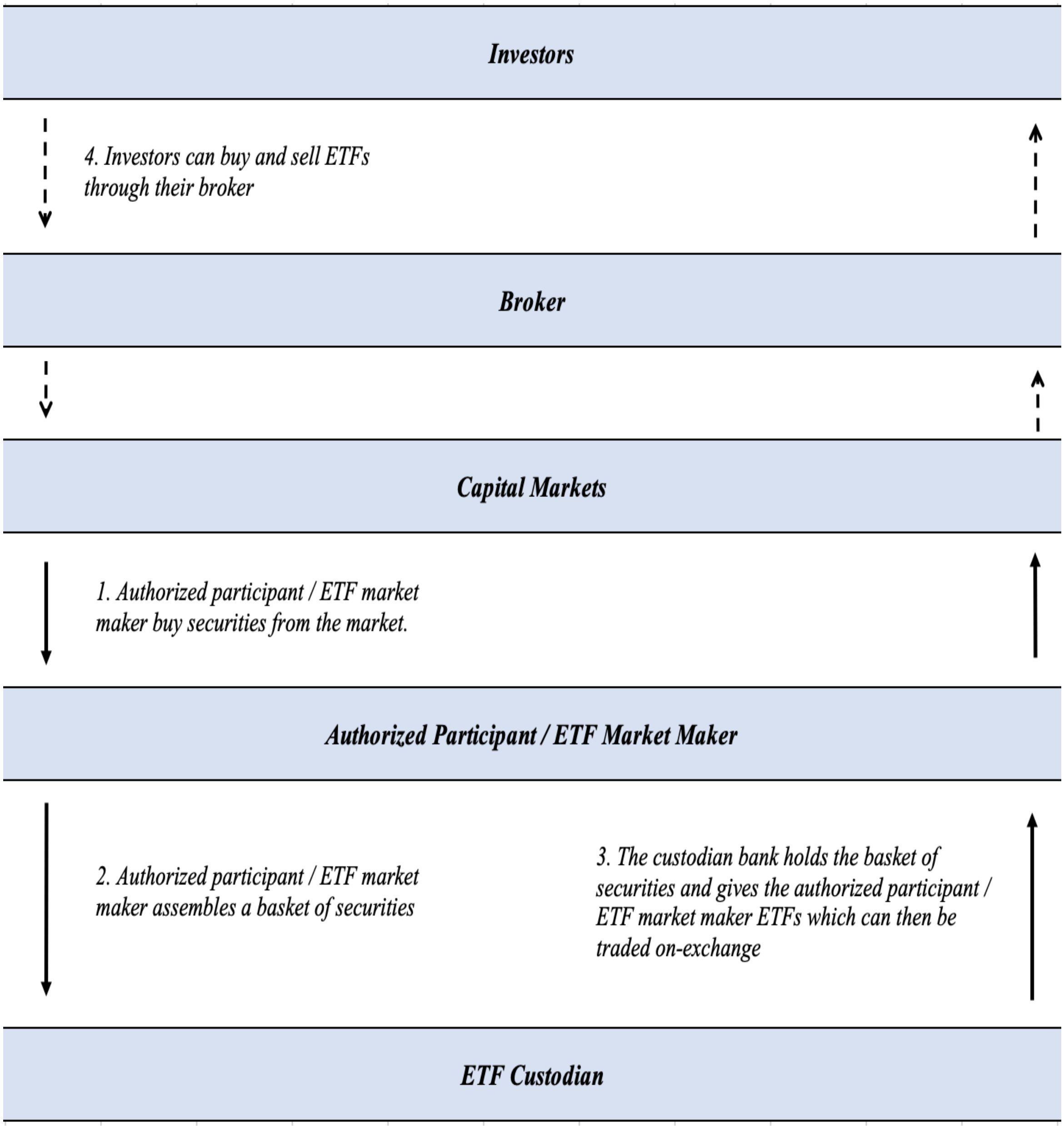

Figure 3. Creation and redemption process (London Stock Exchange, 2014)

As the creation / redemption process was introduced from a physical ETF point of view, the process is similar for synthetic ETFs as well. As the AP acquired securities in exchange for ETF shares, in the case of synthetic ETF the AP submits cash for the issuer to receive ETF shares in return. This process is called in-cash. The issuer then acquires a basket of securities from the swap counterparty with the received cash and by doing that they enter into a total-return swap with each other. The swap in this sense means that the return generated by the basket of securities is paid to the swap counterparty and the issuer gets paid the return of the benchmark index. Vanguard (2013)

The secondary market is a place where the created ETF shares are introduced by AP to be publicly traded, usually in a stock exchange. As AP is given ETF shares in exchange for acquired securities, the AP have a possibility to trade these shares in the market as any publicly traded instrument and institutional investors can then buy these shares through their broker. Because APs have the ability to be involved in both primary and secondary market, they can manage the number of shares to be traded publicly. (Blackrock, 2017)

Because an ETF replicates a pool of assets, the value of an ETF should be approximately equivalent to the value of the underlying assets. The NAV should also reflect the price at which the ETF trades on the secondary market. If deviation occurs between the price of the ETF and its NAV, the ETF is said to be either under- or overpriced against its NAV. If such a pricing inefficiency occurs, ETFs have a unique arbitrage process, which is linked to the creation and redemption process. With ETF arbitrage the prices can be set back to equilibrium. (Delcoure & Zhong, 2007) As mentioned above, APs can manage the number of shares that are publicly traded. For this reasons APs also have an important role in maintaining the price efficiency of ETFs with arbitrage, which is argued to ensure efficient pricing in the secondary market. In ETF arbitrage APs use their role to create or redeem shares of the ETF in order to obtain arbitrage profits. These actions are dependent on whether the ETF price is overpriced or underpriced compared to its NAV. If an ETF is seen to be overpriced compared to its nav, an AP can buy the underlying assets of the ETF and convert the acquired assets into new ETF shares and sell these shares to make arbitrage profit. Similarly, if an ETF is seen to trade below the price of its underlying assets, an AP would redeem ETF shares from the secondary market and convert these shares to underlying assets in order to obtain profits. If an AP spots a price deviation, it is not obligated to act and perform arbitrage. As there are trading costs and uncertainty involved it is up to the AP to decide whether arbitrage would be profitable enough. For this reason, small pricing inefficiencies might occur. (Petäjistö, 2017; Blackrock, 2017)

2.3 Efficient Market Hypothesis

Efficient Market Hypothesis (EMH) has been a widely studied subject ever since it was first presented by Fama (1970). It presents general theoretical framework for the pricing of publicly traded assets as it states that the asset price should reflect all available information, meaning that relevant information should be incorporated to the assets price. Considering this, investors should be able to trade assets at their fair price which would restrict the possibilities for assets to be traded under or over their fair value, hence investors would not be able to beat the market with obtaining abnormal returns. As ETFs are publicly traded, they should also follow EMH and in case of equity ETFs, so should their underlying assets as they are publicly traded too.

Considering three types of information sets, efficiency can be divided into three different levels: weak form, semi-strong form and strong form efficiency. Malkiel (1989,128) and Bodie et al. (2005, 370-373) explains the levels as follows:

The weak form efficiency: If markets are weak form efficient, the prices are said to fully reflect only historical information e.g., prices. Thus, investors are not able to use this kind of information to yield any abnormal returns based on past prices. Weak form efficiency is associated with random walk and therefore prices cannot be predicted. If markets are weak form efficient, technical analysis becomes useless, because historical information would already be impounded in current prices. However, even weak-form efficiency does not suggest that assets are priced at equilibrium all the time. Moreover, it states that strategies using past information are not able to provide constant excess profits.

The semi-strong efficiency: On top of historical information, the semi-strong efficient markets expand its reflection to publicly available information as well. If markets are this form efficient, fundamental company analysis (analysis of balance sheets, income statements etc.) would not yield abnormal returns.

Strong Form efficiency: In addition to historical and publicly available information, this form of efficiency also reflects private information. Thus, it reflects all available information. When strong-form efficiency holds insiders cannot use private information to generate excess returns.

A famous quote by John Maynard Keynes: *"markets can stay irrational longer than you can stay insolvent"* indicates that in the real world, prices do not necessarily act as EMH suggests. Vast amount of empirical research on EMH argues that violations against theory and different forms of it do exist. Examples of research against EMH include exploiting different market anomalies, possibilities to conduct arbitrage and behavioral finance. Each viewpoint represents potential challenges for proponents of EMH. Considering evidence found regarding EMH it is reasonable to conclude that EMH does not hold all the time. (Ross et al. 2013, 452–457).

2.3 Literature review

Compared to common stocks or mutual funds, ETFs have relatively shorter history. Still over the past couple decades, ETFs have grown to be one of the most successful investment vehicles. The exponential growth has been accompanied by numerous new research considering different topics about ETFs. in their study Charupat and Miu (2013) reviewed previous literature about ETFs which covered three different main topics of research examining the different aspects of ETFs. The three main topics are pricing efficiency, tracking ability and performance and lastly effects on underlying securities. Pricing efficiency is the main discussion in this thesis, but because the data used consists of index replicating ETFs, tracking ability will be adopted as well. The main purpose of examining pricing efficiency is to understand how ETFs are priced and whether the creation/redemption process is effective or not. Characteristics of this kind of studies are that they usually involve testing whether there is any kind of deviation to be found between ETF prices and their NAVs, as well as how quickly possible deviation disappears. (Charupat and Miu, 2013) The ETF performance on the other hand is usually evaluated by using tracking errors which represents the differences between the ETF returns and the returns of the respective index the fund replicates.

2.3.1 Pricing efficiency

Ever since the first ETFs were incepted, pricing efficiency has been a widely studied subject with the addition of examining the possibilities of arbitrage. Pricing efficiency has been mainly studied by examining whether ETF prices deviate from NAV on average. Ackert and Tian (2000) examined pricing efficiency with daily data of SPDR, also called as spider, which tracks the performance of S&P 500. They also examined SPDR MidCap, which tracks S&P MidCap 400 and investigated the difference between the two ETFs. In all simplicity they found no economically significant mispricing and these ETFs are relatively efficiently priced in comparison to close-end funds. Elton et al. (2002) took their study a step further and studied also the mispricing persistency of SPDR with autoregression. Their findings were that on average the price deviation was under 1.8 basis points per year and that pricing differences disappear within one trading day, indicating that pricing is efficient and that the process of creation/redemption is effective.

Ever since the SPDR was found to be an efficient investment vehicle compared to mutual funds, many different ETFs have emerged. One category of new ETFs was so-called country ETFs or International ETFs which for example traded in Europe, but the respective index to track is in the US consisting of companies domiciled in the US. In their study, Engle and Sarkar (2006) examined the magnitude and persistence of premiums with both domestic and international ETFs. As they used intra-day data, they were able to find that pricing persistence of domestic ETFs disappeared in minutes. The same for international ETFs would take few hours. Regarding international ETFs, the findings of this study were similar to the fundings of Elton et al (2002). Domestic ETFs were found to be priced very accurately, close to their NAVs. International ETFs were found to be less actively traded which reflected negatively to the pricing accuracy. Delcoure and Zhong (2007) studied only international ETFs and found similar results as Engle and Sarkar (2006). The results show economically significant premiums, but persistence to be only temporary. They also suggest that economically significant premiums are associated with ineffective arbitrage activities, meaning that creation/redemption process does not work effectively, and price deviation persistence lasts longer.

More recent research has found similar results as previous literature. Petäjistö (2017) studied intraday data of both domestic and international ETFs, that are tracking municipal bonds and junk bonds. He had similar findings as Engle & Sarkar (2006), domestic ETFs are more liquid and hence relatively efficiently priced whereas international ETFs were more illiquid and presented insignificant premiums. Petäjistö (2017) also introduced a new method to examine mispricing which addressed the problem of stale NAVs, a problem which occurred for international ETFs. With this correction, the international ETFs were found to have economically significant premiums.

A good majority of studies have gained large popularity on ETFs that are domiciled in the world's largest market, the US. As the industry progresses and markets develop, smaller ETF markets outside the US have caught the eyes of researchers. For example, Kayali and Ozkan (2012) investigated mispricing of domestic ETFs in Turkey. They found statistically significant mispricing but came to a conclusion, that as larger the price deviations between ETF price and NAV are rare, therefore the more limited the arbitrage opportunities are. Charteris (2013) on the other hand studied mispricing of both domestic ETFs in South Africa and international ETFs issued in Europe. The results suggest that the domestic ETFs were found efficiently priced and in most cases the deviations disappeared within one trading day. The international ETFs persisted for one trading day which indicates that these were also relatively efficiently priced. The possible cause for efficiency as further studied and the reason was found to be efficient arbitrage process. Kumar (2018) reached similar results in Indian market as ETFs are found to be priced efficient and did not provide arbitrage opportunities for market participants.

The first domestic ETFs were listed in Deutsche Börse in Germany in 2000. Still there are not too many studies that investigate pricing efficiency of domestic ETFs in Europe in terms of price deviations between ETF price and NAV and the persistency of possible deviation. In their study, Kreis and Licht (2017) investigated European indexes tracking ETFs listed in German stock exchange XETRA. Their results show that ETF prices regularly deviate from NAV and thus show significant excess returns. Deviations can be explained by the ETF pricing process, meaning that creation/redemption process drives the deviation. Further they added a simple trading strategy in order to exploit the inefficiencies and found that the data shows significant gross returns through the whole time period, but net returns turn negative after trading costs are included.

2.3.2 ETF performance

Tracking error calculations are not a new phenomenon as tracking errors of ETFs can be calculated the same way as it has been calculated for mutual funds. Roll (1992) and Pope and Yadav (1994) have introduced four of the most commonly adopted methods to measure tracking error. The first method is calculated as the average absolute return differences between and ETF and the respective index. The second method is a root-mean-square deviation and the third is the standard deviation. The fourth method is calculated as a regression between ETF returns and index returns where tracking error is estimated as residual standard errors. The first three methods do not tell whether the fund under- or overperforms the index as values used are absolute values whereas the fourth method includes the intercept value in which a negative value indicates underperformance and positive value indicates overperformance.

Given the mentioned methods for calculating tracking error, a vast majority of studies focus on ETFs that tracks both domestic and international indexes. Shin and Soydemir (2010) conducted a study on 26 ETFs, of which 20 were Morgan Stanley Capital International (MSCI) Country funds and six were U.S. equity market funds. Their findings indicate statistically significant tracking errors and that tracking errors are significantly persistent. Adjusted returns are generally lower than benchmark returns, meaning that with a passive investment strategy, ETFs do not outperform their respective benchmark. Blitz and Huij (2012) examined passive ETFs have passive exposure to global emerging markets. Their result indicate that emerging markets show higher levels tracking error compared to ETFs that track developed market indexes. This is seen to be cause by the fact that cross-sectional dispersion in stock returns is larger in emerging markets. Emerging market ETFs are found to exhibit larger tracking errors compared to developed markets but on average, ETFs from both markets display similar long run underperformance.

Elia (2011) investigated the tracking ability of both equity ETFs and ETFs that are pooled of derivatives and swaps, also called as synthetic ETFs. The results show that both ETF types in Europe show significant tracking error, but synthetic ETFs perform better compared to traditional Equity ETFs in terms of tracking error as well as tax efficiency. Elia (2011) also included emerging markets into his study and found that tax efficiency is more efficient in emerging markets. Traditional equity ETFs are seen to still perform better returns than synthetic ETFs as synthetic ETFs underperform compared both the respective benchmark and equity ETFs. A more recent study from Almelu and Goya (2022) examined 27 ETFs domiciled in India. Most of the 27 ETFs outperformed their respective benchmark but also shows a significant average tracking error.

3 Data

The time period of this study is four years from February 2015 to February 2019. First objective of data gathering was to gather a minimum of five European indexes and then search ETFs that replicates the chosen respective indexes. The data sample used in this study consists of equity ETFs which are passively managed and domiciled in Europe. Passive management means that the ETF is not trying to over-perform the index it replicates. Data is collected as price adjusted, meaning that dividends are considered in the closing prices, because some ETFs distribute their dividends, and some invest them back to the fund. As the sample of data used includes also ETFs from outside the Euro currency, this data was collected in euro. Due to the limitations of available NAV data for the time period used in this thesis, a sample of nine ETFs were selected which makes 9396 data points for the whole data sample and 1044 data points per ETF. The data was collected from Thomson Reuters DataStream, EIKON and ETF distributor's websites. General information of selected ETFs is presented in Table 1.

Table 1. General information of the selected ETFs. Market denotes country of stock market as three letter abbreviation, FRA = France, LUX = Luxembourg, IRE = Ireland, SPA = Spain, NOR = Norway. Dist. means distributing and Acc. means accumulating. TER is abbreviation of Total Expense Ratio.

ETF	Listing	Inception	Market	Use of Profits	Index	TER
Lyxor CAC 40	Euronext	13.12.2000	FRA	Dist.	CAC 40	0.25 %
Amundi CAC 40	Euronext	28.2.2003	FRA	Dist.	CAC 40	0.25 %
BNP Easy CAC 40	Euronext	7.3.2005	FRA	Dist.	CAC 40	0.25 %
Lyxor Euro Stoxx 50	Euronext	19.2.2001	FRA	Dist.	Euro Stoxx 50	0.20 %
UBS Euro Stoxx 50	SIX Swiss	29.10.2001	LUX	Dist.	Euro Stoxx 50	0.15 %
iShares FTSE 100	London	27.4.2000	IRE	Dist.	FTSE 100	0.07 %
BBVA Ibex 35	Madrid	7.10.2006	SPA	Dist.	IBEX	0.39 %
XACT OBX	Oslo	7.4.2005	NOR	Acc.	OBX	0.30 %
DNB OBX	Oslo	1.3.2005	NOR	Acc.	OBX	0.31 %

Table 2 presents the descriptive statistics for individual ETF prices and NAVs. Descriptive statistics consist of mean value, standard deviation, minimum and maximum value, skewness, and kurtosis. Skewness illustrates whether the data sample is symmetrically distributed, and kurtosis illustrates a level of peakness, meaning that if most of the data points are in the center the data is said to be peaked. A desired value for skewness is between -1 and +1, otherwise the data is substantially skewed. Same goes for kurtosis, values under -1 indicates that the data is too flat and over 1 indicates that the data is too peaked. (Hair et al. 2017, p. 61)

In terms of daily prices and daily NAVs, all ETFs exhibit a kurtosis over -2 and below 2. This is acceptable to prove univariate normal distribution. Negative kurtosis values indicate that some of the data is platykurtic and positive kurtosis indicate that some of the data is leptokurtic. In univariate level, some of the data is seen as too peaked, but when we look at the kurtosis of the whole data sample, it is well higher than -1 and lower than +1. Same goes for skewness, overall ETF prices and NAVs reaches skewness over -1 and under +1 which indicates of normally distributed data.

Table 2. Descriptive statistics for individual ETFs in terms of price and NAV.

ETF Price						
ETF	**Mean**	**Std.**	**Min.**	**Max.**	**Skew.**	**Kurt.**
Lyxor CAC 40	48,87	3,84	38,57	56,21	-0,30	-1,04
Amundi CAC 40	67,15	7,00	50,83	79,53	-0,07	-1,26
BNP Easy CAC 40	8,16	0,65	6,40	9,33	-0,26	-1,07
Lyxor Euro Stoxx 50	32,73	2,47	26,46	38,22	-0,14	-1,02
UBS Euro Stoxx 50	33,27	2,50	26,72	38,26	-0,28	-1,08
iShares FTSE 100	8,34	0,56	6,99	9,97	0,82	0,47
BBVA Ibex 35	9,82	0,89	7,79	11,93	0,63	0,14
XATC OBX	6,70	0,98	4,66	8,97	0,52	0,04
DNB OBX	9,82	0,89	4,53	8,88	0,22	-0,81
All	24,63	21,12	4,53	79,53	0,90	-0,42
NAV						
Lyxor CAC 40	48,81	3,83	38,89	56,20	-0,30	-1,03
Amundi CAC 40	67,05	6,95	50,71	79,45	-0,07	-1,25
BNP Easy CAC 40	8,16	0,65	6,39	9,32	-0,26	-1,07
Lyxor Euro Stoxx 50	32,72	2,47	26,45	38,22	-0,14	-1,02
UBS Euro Stoxx 50	33,26	2,49	26,56	38,20	-0,28	-1,09

iShares FTSE 100	8.31	0,56	6,97	9,96	0,84	0,53
BBVA Ibex 35	9,82	0,89	7,76	11,88	0,22	0,53
XACT OBX	6,70	0,99	4,56	8,94	0,16	-0,92
DNB OBX	6,65	0,98	4,53	8,87	0,16	-0,91
All	24,60	21,08	4,53	79,45	0,89	-0,43

4 Methodology

This chapter introduces methods used in this thesis for empirical analysis and the reasoning behind selection of these methods implemented. Firstly, the daily returns of ETFs and their underlying benchmark indices will be calculated as they are later used for tracking error calculations. Returns based on market prices of ETFs reflect the actual (realized) returns to the ETF investors and incorporate even the effect of demand and supply conditions in the ETF market and the resulting pricing inefficiency (deviation between price and NAV of ETFs) if any. After calculating the returns, the price deviation between ETFs and NAVs are calculated which is also called premium/discount, depending on whether the value is positive or negative. Return difference will also being calculated which is a deviation between daily ETF returns and daily index returns which the ETF replicates. These calculations are the base for possible detection of pricing inefficiency. After the premium / discount calculations, the obtained values are evaluated with paired samples t-test in order to find whether the ETFs obtain any statistically significant mispricing. The purpose of section 4.4 is then to find whether the possible mispricing is persistent by using an autoregressive model. After the mispricing calculations the relationship between the ETF prices and their respective NAV is evaluated with OLS regression by testing whether they are in unity. Lastly the firstly calculated ETF returns, and Index returns are used to calculate tracking error.

When it comes to ETF mispricing and pricing efficiency, the literature review suggests previously applied econometric methods to detect this phenomenon. In this study we adopt similar approach as other studies as mispricing is rarely tested focusing on just one method but multiple methods. With multiple econometric methods a comprehensive analysis can be constructed and draw a conclusion. Table 2 summarizes the research questions and the methods used to find an answer to the formed research questions with their related literature.

Table 3. Methodology used to find answers to research questions.

Model	Research Question
Paired sample t-test	Q1: Is there evidence of significant mispricing between ETF prices and NAVs?
AR with 1 to 5 lags	Q2: Is there evidence of persistent mispricing?
OLS Regression	Q3: Is there evidence of strong relationship between the ETF price and its NAV?
Tracking Error	Q4: Is there evidence of tracking error between ETF returns and index returns?

4.1 Returns

One part of this study is to find how well the selected ETFs replicate their respective index. In order to find answers to this question, daily returns are calculated. The daily ETF returns, and index returns will be calculated for the whole time period from February 2015 to February 2015. The daily returns are calculated with the following equations:

$$R_{ETF} = \frac{ETF_t - ETF_{t-1}}{ETF_{t-1}} * 100 \qquad (1)$$

$$R_{INDEX} = \frac{INDEX_t - INDEX_{t-1}}{INDEX_{t-1}} * 100 \qquad (2)$$

In equation (1) $\mathrm{R_{ETF}}$ represents the daily return for an ETF and $\mathrm{ETF_t}$ is the closing price of an ETF at time t. In equation (2) $\mathrm{R_{INDEX}}$ denotes the daily return of an index and $\mathrm{INDEX_t}$ represents the index price at time t. For both equations the subscript t-1 represents the price of the previous day.

4.2 Premiums and Discounts

When something trades in premium, it is said to be trade at a higher value than what it is compared to. If ETF is trading in discount, it is just vice versa, it trades in lower value than what it is compared to. In this thesis it is a key variable since it is used as the core when defining whether an ETF is mispriced or not. Here premium/discount is calculated with equation 3 and 4.

$$PD_{EUR} = ETF_t - NAV_t \tag{3}$$

$$PD_{REL} = \frac{ETF_t - NAV_t}{NAV_t} * 100 \tag{4}$$

Premium/discount calculated in equation (3) is simply a deviation between ETF trading price and NAV of the given day and equation (4) is the relative deviation. $\mathrm{PD_{EUR}}$ denotes premium/discount in euro, PD_{REL} denotes the relative premium/discount, $\mathrm{ETF_t}$ is the closing price of ETF at time t and $\mathrm{NAV_t}$ is the net asset value of the equities at time t.

In this thesis the relationship between ETF and the benchmark index it replicates, is studied as well. In terms of premiums and discounts, the relationship between ETF returns and benchmark index returns will be described simply as return deviation and is calculated with the following simple equation:

$$RD_t = R_{ETF,t} - R_{index,t} \tag{5}$$

Where $\mathrm{RD_t}$ denotes return deviation at time *t*, $\mathrm{R_{ETF,t}}$ daily ETF return at time *t* and $\mathrm{R_{index,t}}$ the daily index returns at time t.

4.3 Paired Samples t-test

Discovering mispriced ETFs relative to fund's NAV can be done on the basis of premium/discount from equation (3). As average premium/discount can be calculated for each individual ETF a t-test can be deployed to determine whether premium/discount differs significantly from zero. T-test is any statistical hypothesis test in which the test statistic follows a student's t-distribution under a null hypothesis. Student's t-test is one of the most widely used statistical tests to compare two groups. It essentially compares the mean of a sample to a population mean, using either a known population standard deviation or sample standard deviation. It tells how significant a difference between group means is, and if those differences could have happened by chance (Stone, 2010).

Since the data samples used in this thesis are linked together, a paired samples t-test (also known as dependent t-test) is the most suitable t-test to be used. A comparison is made between two data sample means. The following equation is used as suggested by Martin & Bridgmon (2012):

$$t = \overline{D} / \sqrt{\Sigma D^2 - ((\Sigma D)^2 / n) / (n(n-1))} \tag{6}$$

where, t is t-value, $\overline{\mathrm{D}}$ represents the average difference between the two samples used, ΣD represents total sum of differences between the two samples, and n is number of observations. The calculation will be made between the daily ETF prices and their respective NAV, and later in section five between the ETF return and index return.

The null hypothesis of the t-test is that there is no significant difference between the means of two dependent groups. In other words, the mean of paired differences equals zero in the population. The alternative hypothesis is that the mean of paired differences does not equal zero in the population, meaning that there is significant difference. This conclusion can be made by comparing the obtained t-value with equation (6) to the critical t-value or the p-value to the chosen confidence level. If the obtained t-value exceeds the critical t-value, the null hypothesis is rejected. Similarly, if the p-value is lower than the chosen confidence, the null hypothesis can be rejected. (Tähtinen et al. 2020) The hypotheses can be tested as one-tailed or two-tailed. One-tailed hypothesis testing basically means that emphasis of significance level (alpha) is put on one side of normal distribution, hence the region of rejection is either left or right. It is utilized when the direction of sample difference matters. Two-tailed test of null hypothesis puts emphasis of significance level to both sides of normal distribution, meaning that the direction of the two means does not matter. (Ringwalt et al. 2011) Hence, two-tailed test is utilized in this thesis.

4.4 Mispricing persistency

Charteris (2013, 20) argues that the existence of premium / discount does not immediately indicate that the ETF is mispriced or that the pricing is inefficient, only unless mispricing is persistent. Petäjistö (2017) argues that mispricing significance could be caused by the volatility of premium/discount, therefore measuring the persistency of possible mispricing has an important role in this thesis. As authorized participants have the ability to practice creation / redemption of ETF shares in order to keep the ETF price in line with the NAV, in a sense mispricing persistency measures the efficiency of this process. Mispricing creates opportunities for arbitrageurs to practice arbitrage and capitalize from the creation / redemption process not working accordingly. In this thesis mispricing persistency is being modeled with autoregressive (AR) model. The AR-model can be used to estimate whether past premium discount/premium could be used to forecast today's premium/discount or if the past premium/discount has explanatory power of today's premium/discount. Following equation (7) of order n describes the AR model, denoted as AR(n), used:

$$PD_{REL,t} = \alpha + \beta_1 PD_{REL,t-1} + \beta_2 PD_{REL,t-2} + \cdots + \beta_n PD_{REL,t-n} + \varepsilon_t \tag{7}$$

In equation (7) $\mathrm{PD}_{\mathrm{REL,t}}$ denotes relative ETF premium / discount at time t. The variables on the right-side of the equal sign with beta coefficients represent the lagged values of premium/discount, except α is the intercept and ε_t represents the error term which essentially means white noise or randomness. In this model the variable $\mathrm{PD}_{\mathrm{REL,t}}$ is a dependent variable, meaning that its dependent on the right-hand side beta coefficients, which are linked to the past values the variable took in previous days (Brooks, 2008, p.215). Because of the connection between beta coefficients and lagged values, the interest of the AR(n) model lies in the beta coefficients. Premium/discount persistency depends on whether beta coefficients are significant or not. If the first beta value is statistically insignificant, it indicates that premium / discount disappears within one trading day. If the first beta is statistically significant, this means that the premium / discount persists at least for one day. Continuing this logic, if the second beta is significant but the third beta is not, premium/discount persists at least for two days. The null hypothesis of mispricing is not persistent will be rejected if premium /

discount persists at least one day, meaning that at least the first beta is significant, because investors would be able to capitalize on this inefficiency. In terms of modelling, the reasoning of Charteris (2013) is followed and modelling should start by using just one lag. This continues with additional lags until the last beta coefficient is not significant. Also, if any beta coefficient turns insignificant after adding an additional lagged value, the modelling is stopped. The maximum number of lags used in this thesis is five as it simulates one trading week. Each ETF will be modelled individually.

Before modelling with autoregression can begin, one must test the data for stationarity, because stationarity is assumed by AR-model. This can be done with Augmented Dickey-Fuller test (ADF-test), which is a commonly used unit root test. Essentially ADF-test is testing if time series data has a stochastic trend. The following equation suggested by Cheung et al. (1995) is implemented:

$$\Delta x_t = \mu + \gamma t + \alpha x_{t-1} + \sum_{j=1}^{k-1} \beta_j \, \Delta x_{t-j} + \mu_t \tag{8}$$

In equation (8) Δ denotes the difference operator for variable x which is the time series, μ is the intercept constant, γ is the slope coefficient on time trend t, Δx_{t-j} represents the lagged values of the dependent variable x, α is a coefficient of lagged x_{t-1}, and μ_t is white noise. (Cheung et al., 1995) ADF-test does not straight forward test stationarity, it rather tests whether the process root coefficient equals zero which means that the time series has a unit root. If a time series has a unit root, it is said to follow a stochastic trend. This is the null hypothesis. The alternative hypothesis is that the time series does not have a unit root and therefore the data is stationary. The data used in this equation is a time series of the price deviation of ETFs and NAVs for each individual ETF. (Cheung et al., 1995) As price data is not always stationary, therefore in order to reach stationarity a method called data differencing has to be implemented. Differencing essentially means that price difference of ETF and NAV at time t is subtracted from price difference of $t-1$.

4.5 Ordinary Least Squares Regression

Ordinary Least Squares (OLS) linear regression is one of the most popular linear regressions used in regression analysis. OLS estimates the relationship between a dependent variable and one or multiple explanatory variables. With the use of linear regression, it is desired to find and understand the relationship between the dependent and independent variables. Kumar (2018) and Kayali (2007) used similar approach in their studies, hence their suggested formula is used in this study as presented in following equation:

$$y = \alpha + \beta_1 x + \varepsilon \quad (9)$$

In equation (9) y represents the dependent variable which in this case is the ETF price, α the intercept value, β_1 is the slope of the regression line, x is the independent variable which in this case is the NAV, and ε is the random error. The purpose of the OLS regression is to find the best alpha and beta estimates so that the squared distance from predicted response and actual response to reaches minimum from all possible regression coefficient choices. To simply put, the OLS method tries to find parameter estimates with choosing a regression line that is the closest to all data points by minimizing the sum of squared residuals. (Yan & Su. 2009, 9-11)

4.6 Tracking Ability and Performance

Equity ETFs are essentially created to either passively or actively replicate their respective benchmark index. In theory, replication is possible but there are many factors' that indices do not take into consideration, such as portfolio creation and management as well as other market factors that have influence on ETF prices. These factors make it hard for ETFs to reach similar returns as their benchmark and can cause tracking inefficiencies. To measure the magnitude of possible tracking inefficiency, different calculations of tracking error can be deployed. Before making a decision to invest into an ETF, tracking error calculation is a vital part of that decision. (Johnson et al. 2013)

When going through related literature, one can notice that there is not a generally accepted method for calculating tracking errors, different industry participants define tracking error in different ways. Generally tracking error can be defined as Harper et al. (2006) defines it; as a deviation between the ETFs performance in terms of returns and respective their benchmark indices, which essentially gives information whether ETF has over performed its benchmark in terms of returns. This has been defined as return deviation (RD) in section 4.2 as equation (5) in this thesis. In order to reach a comprehensive outlook of ETFs relationship to its benchmark in terms of tracking error, this thesis follows similar approach as Alamelu and Goyal (2022) and Elia (2011) used in their papers.

To measure the magnitude of possible tracking error, three different methods are chosen. First tracking error calculation is presented as equation (10) which is a measure of mean absolute return differences between daily ETF returns and benchmark returns, simply titled as TEABS. The absolute return difference tells how far away from each other the two calculated points are, therefore, it provides information about the real magnitude of the difference in terms of distance. In equation (11) tracking error is being calculated as standard deviation of the average absolute difference. The third method is calculated as a regression as in equation (12), where the tracking error is the residual standard error of the linear regression of ETF returns and index returns.

$$TEABS_{i,t} = \frac{\sum_{t=1}^{n} |R_{i,t} - R_{b,t}|}{n} \tag{10}$$

$$TESD_{i,t} = \sqrt{\frac{1}{n-1} \sum_{t=1}^{n} (TEABS_{i,t} - \overline{TEABS}_i)^2} \tag{11}$$

$$R_{i,t} = \alpha + \beta_1 R_{b,t} + \varepsilon \tag{12}$$

In equation (10) $R_{i,t}$ represents the daily ETF return at time t, $R_{b,t}$ is the daily benchmark index at time t and n is the number of observations. Simply put, $|R_{i,t} - R_{b,t}|$ represents average absolute difference between ETF return and index return. In equation (11) $TEABS_{i,t}$ represents the absolute difference in terms of returns at time t and $\overline{\mathrm{TEABS}_i}$ is the average absolute difference of returns.

In equation (12) $\mathrm{R}_{i,t}$ and $\mathrm{R}_{b,t}$ have the same meaning as in equation (10), but adding to this, α represents the intercept value, β_1 is the beta coefficient and ε is the error term. The alpha coefficient indicates excess return that the ETF has obtained relative to the respective index, although the coefficient values are likely to reach low insignificant values due to the passive replication nature of ETFs. Beta coefficient is an estimate to which extent the ETF is able to replicate the respective index, meaning that beta value of 1 represents full replication, beta coefficient value over one is an indication of outperformance and beta coefficient under one means underperformance which is also an indication that ETF is composed differently. (Milonas and Rompotis, 2012) The analysis of these regression values is considered to be the performance factor as these values give quite much information about how the ETF performs compared to its respective index.

5 Empirical Results

5.1 Premium and Discount

In this thesis the ETF mispricing is being studied on a daily level with a time span of four years, from February of 2015 to February of 2019. Figure 3 illustrates the trend of a daily comparison of ETF closing prices and NAV for every individual ETF. When the daily ETF prices and NAV are compared, it can be noted that ETF prices follow closely to their respective NAV during the whole four-year time period as the blue line dominates the in every figure. However, some ETFs show discrepancies as can be seen from table 4 and table 5 later on. The relative price deviation can be as low as 0.009 % on one day and as high as 9.96 % on other. Similarly, the price deviation in euro currency can be as high as 2.60 euro and as low as 0.01 euros in terms of distance from zero as a value of zero means that ETF price and NAV are equal. The price deviations are illustrated as a frequency distribution table 1, where the frequencies are presented for relative deviation and in euro deviation. The table gives insight about the magnitude of price deviations of the whole data set as maximum number of observations is 9396 data points

As can be seen from the frequency distribution table 4 below, over 8500 data points have a daily mispricing between the interval of [-0.5%;0.5%] and this makes it roughly 90 % of the whole data sample. The further the distance goes from zero, negative or positive, the lower the frequency. Hence, the interest lies on the so-called outlier values. One observation obtains a value equal to or under -5 %, meaning that the one observation was trading at discount of 5 % or over. Similarly, three observations reached a value equal to or greater than 5 %, meaning that an ETF or ETFs traded at premium of 5 % or over during the time period. In Appendix X the frequency distribution is illustrated for every ETF individually.

In terms of euro deviation, over 90 % of the data points obtained a daily mispricing between -0.2 euro and 0.2 euro. As for the outlier values 17 observations has a value equal or under -0.8 euro, meaning that the ETF or the ETFs traded at discount of 0.8 euro over. Similarly, 34 observations obtain a value equal to or over 0.8 euro, meaning that the ETF or ETFs traded at premium of 0.8 euro or over.

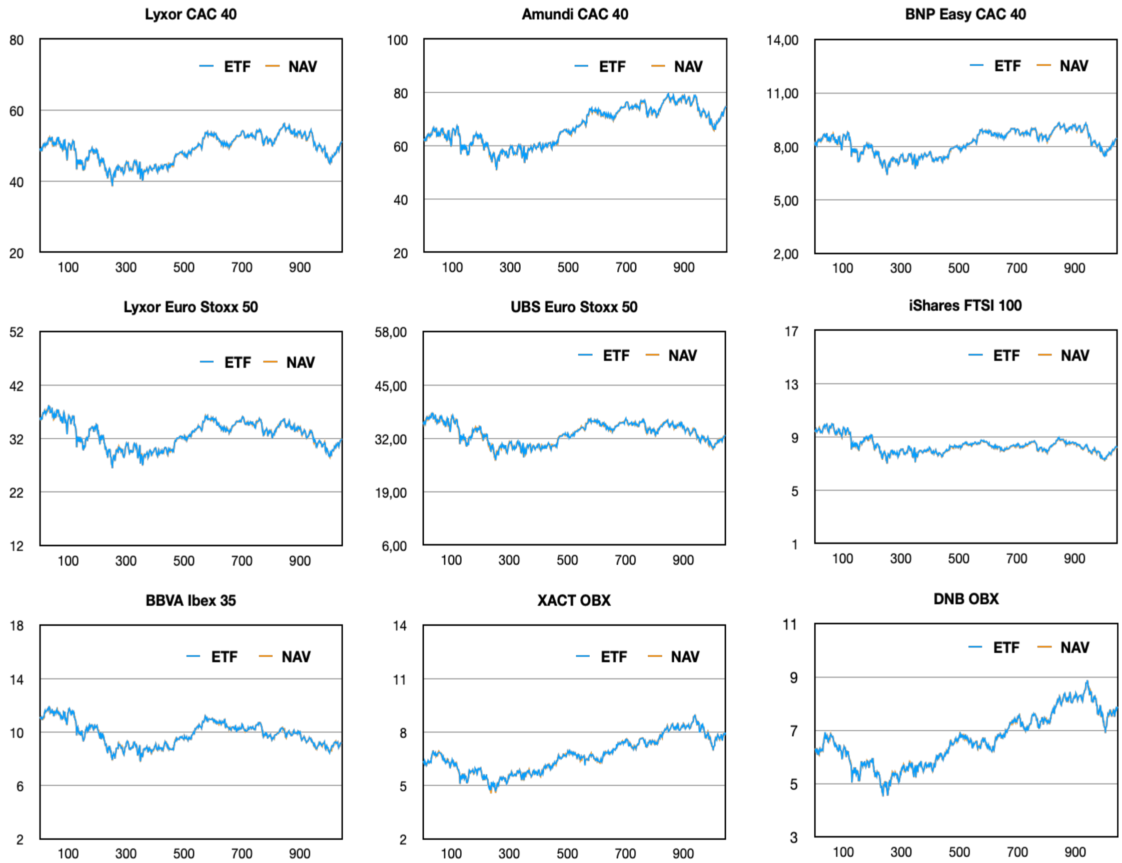

Figure 3. Average closing values of ETFs and NAVs.

Table 4. Frequency distribution of daily relative premium/discount. On the left-hand side is frequency distribution for relative premium/discount and on the right-hand side is the premium/discount in euro currency. Dist = percentage distribution of the whole data set.

% Deviation	Frequency	% Dist.	€ Deviation	Frequency	% Dist.
≤ -5 %	1	0.01 %	≤ - 0.8	17	0.18 %
> - 5 % and ≤ - 2.5 %	21	0.22 %	> -0.8 and ≤ -0.6	27	0.29 %
> - 2.5 % and ≤ - 1 %	145	1.54 %	> -0.6 and ≤ -0.4	60	0.64 %
> - 1 % and ≤ - 0.5 %	170	1.81 %	> -0.4 and ≤ -0.2	149	1.59 %
> - 0.5 % and ≤ 0 %	3216	34.23 %	> -0.2 and ≤ 0	3300	35.12 %
> 0 % and ≤ 0.5 %	5339	56.82 %	> 0 and ≤ 0.2	5376	57.22 %
> 0.5 % and ≤ 1 %	297	3.16 %	> 0.2 and ≤ 0.4	324	3.45 %
> 1 % and ≤ 2.5 %	167	1.78 %	> 0.4 and ≤ 0.6	72	0.77 %
> 2.5 % and ≤ 5 %	37	0.39 %	> 0.6 and ≤ 0.8	37	0.39 %
≥ 5 %	3	0.03 %	≥ 0.8	34	0.36 %

5.2 Paired sample t-test

The EMH States that share prices reflect all available information and that stocks trade at their fair market value. Hence, in theory, ETFs should be efficiently priced, and one should not be able to detect any mispricing. Also, creation/redemption process should keep ETF trading price close to its NAV, which indicates that mispricing should not occur. Table 5 provides detailed analysis of ETF premiums and discounts in euro compared to their NAVs. The results are presented individually for each ETF. Statistical significance is tested for average premium/discount with paired sample t-test. The null hypothesis is that the mean of paired differences equals zero, which indicates that ETFs would be efficiently priced if the null hypothesis is accepted. Significance is being evaluated with two-tailed p-value.

Price deviations of the data sample variate between -1.44 euro and 2.61 euro. UBS Euro Stoxx 50 ETF obtained both the lowest minimum value and the highest maximum value of the whole dataset. This means that on one day the ETF traded at discount of 1.44 euro and at premium of 2.61 euro on another day during the four-year time period. It is also the most volatile ETF of the whole data set, obtaining a standard deviation of 0.3816. Comparing the obtained values individually, it can be noted that there are some substantial differences in premiums/discounts. For example, compare to Lyxor Euro Stoxx 50 which replicates the same index as UBS Euro Stoxx 50, the maximum day premium of UBS Euro Stoxx 50 is seven times higher than the respective value obtained by Lyxor Euro Stoxx 50. Similarly, it is a hundred times higher than the respective value obtained by DNB OBX, which also obtained the lowest maximum value of 0.02 euro and the highest minimum value of -0.03 euro were obtained by DNB OBX ETF. It is also the only ETF, that on daily average trades at discount obtaining an average of -0.0014 euro. Considering the total average, the data sample obtains an average maximum value of 0.5804 euro and average minimum value of -0.3382. The data sample obtains a total average daily premium of 0.24 euro compared to NAV, indicating that the ETFs of used in this study are trading at daily premium on average during the four-year time period.

Table 5. Premium/discount statistics with t-test for individual ETFs. Values are in euros. * Means statistical significance at 95% confidence level. Min = minimum value, Max = maximum value, std = standard deviation of daily difference of ETF price and NAV, BA = daily average bid-ask spread is a proxy for trading costs, and obs = number of observations.

ETF	Min.	Max.	Mean	Std.	t-value	p-value	BA	obs.
Lyxor CAC 40	-0.2515	0.3788	0.0535	0.0586	**29.4532***	< 0.001	0.0335	1044
Amundi CAC 40	-0.2237	0.8777	0.0995	0.1016	**31.6356***	< 0.001	0.0977	1044
BNP Easy CAC 40	-0.1070	0.1601	0.0041	0.0113	**11.5926***	< 0.001	0.0217	1044
Lyxor Euro Stoxx 50	-0.3536	0.3724	0.0088	0.0318	**8.9942***	< 0.001	0.0375	1044
UBS Euro Stoxx 50	-1.4410	2.6090	0.0191	0.3816	1.6178	0.1060	0.0497	1044
iShares FTSE 100	-0.0361	0.0842	0.0264	0.0147	**58.0625***	< 0.001	0.0026	1044
BBVA Ibex 35	-0.2106	0.2515	0.0013	0.0228	1.8704	0.0617	0.3235	1044
XACT OBX	-0.3900	0.4700	0.0023	0.0330	**2.2875***	0.0224	0.0772	1044
DNB OBX	-0.0300	0.0200	-0.0014	0.0085	**-5.2073***	< 0.001	0.0140	1044
Average	-0.3382	0.5804	0.0237	0.0738	**3.9680***	< 0.001	0.0731	1044

When interpreting the results of the paired t-test, the p-value is being compared to the chosen 95% confidence level (alpha 0.05). P-values under 5% are considered to be statistically significant and thus the null hypothesis that mean of paired differences equals zero can be rejected. Based on this, it can be seen from table 5 that seven out of nine ETFs have p-values under 5%, which means that the mispricing of these ETFs is found to be statistically significant, hence we fail to accept the null hypothesis with the 95% confidence level. Only two ETFs out of eleven have p-values over 5%, meaning that these ETFs can accept the null hypothesis and therefore the average difference between the ETF price and its respective NAV for these ETFs is not significant. Rather interesting finding is that the ETFs that obtained statistically insignificant values are UBS Euro Stoxx 50 and BBVA Ibex 35, ETFs of which first obtained the highest standard deviation of all the ETFs and the latter on the other hand obtained the lowest average premium. Another interesting finding is that DNB OBX is the only ETF that on average seems to trade at discount during the four-year time period. When comparing the mean relative return differences of BBVA Ibex 35 and DNB OBX, it is interesting to see that the distance difference to zero is only 0,0001 and only DNB OBX is statistically significant. As t-test relies on normal distribution, non-normally distributed data can get unreliable results. This explains the result of UBS Euro Stoxx 50 as it is not normally distributed compared to rest of the data. The insignificance of BBVA Ibex 35 can be explained by the fact that the average difference between daily price and NAV is almost zero.

The daily relative deviations of ETFs compared to the respective NAV are presented in table 6. The daily relative deviations variate between -5.9816 % and 9.9816 %. What is different to table 3 is that XACT OBX obtained both the lowest minimum value and largest maximum value of the relative difference of the whole data set, indicating that it traded at discount of 5.9816% on one day and at premium of 9.9816 % on some other day relative to the NAV. Although the variation XACT OBX obtained a standard deviation of 0.0057 which is above average. UBS Euro Stoxx 50 obtained the second lowest minimum of -4.5342 % and second largest maximum value of 9.2318 % with the highest standard deviation of 0.0119 making this ETF the most volatile in terms of premium/discount3. If the minimum and maximum values of these two ETFs are compared to the average minimum and maximum value of the whole data sample, the differences are large as the average minimum value is -1.8103 % and the average maximum value is 3.1625 %. Even though XACT OBX obtained the lowest minimum value the highest maximum value of the whole data set, when the daily average relative differences are compared together, the daily return differences are constant for this ETF. This can be determined from the daily average relative difference of 0.0456 % which is below average and from the low average standard deviation of 0.0057, even though it is above the data sample average. The daily average relative difference between ETFs and their respective NAVs of the whole data set is 0.0830 and the average standard deviation is 0.0031.

Table 6. Premium/discount statistics with t-test for individual ETFs. Values are in percentages. * Means statistical significance at 95 % confidence level. Min = minimum value, Max = maximum value, std = standard deviation of daily difference of ETF price and NAV, BA = average daily bid-ask spread is a proxy for trading costs, and obs = number of observations.

ETF	Min.	Max.	Mean	Std.	t-value	p-value	BA	obs.
Lyxor CAC 40	-0.4485	0.7337	0.1084	0.0012	**29.6800***	< 0.001	0.0671	1044
Amundi CAC 40	-0.3138	1.3272	0.1422	0.0015	**31.5144***	< 0.001	0.1396	1044
BNP Easy CAC 40	-1.4085	1.7459	0.0485	0.0014	**1.5960***	< 0.001	0.2512	1044
Lyxor Euro Stoxx 50	-1.0912	1.1379	0.0270	0.0010	**8.8322***	< 0.001	0.1116	1044
UBS Euro Stoxx 50	-4.5342	9.2318	0.0632	0.0119	1.7195	0.0858	0.1504	1044
iShares FTSE 100	-0.4548	1.0326	0.3191	0.0018	**58.5321***	< 0.001	0.0314	1044
BBVA Ibex 35	-2.0799	2.9198	0.0142	0.0024	1.8880	0.0593	3.2815	1044
XACT OBX	-5.9816	9.9576	0.0456	0.0057	**2.5858***	0.0098	1.1530	1044
DNB OBX	-0.4246	0.3759	-0.0215	0.0013	**-5.2465***	< 0.001	0.2143	1044
Average	-1.8103	3.1625	0.0830	0.0031	**17.1285***	< 0.001	0.6001	1044

In terms of statistical significance, the null hypothesis of two sample means of paired differences equals zero can be rejected if p-value falls below 5 %. As in table 3, for seven out of nine ETFs the relative difference is found statistically significant, meaning that the daily average relative difference of these ETFs compared to their respective NAVs are significantly different from zero. The same ETFs as in table 5, UBS euro Stoxx 50 and BBVA Ibex 35 are found to be statistically insignificant, meaning that we fail to reject the null hypothesis. These insignificances can be explained with the same explanations as above. The data of UBS Euro Stoxx 50 ETF is non-normally distributed compared to the rest of the data. Even though the relative values of XACT OBX seem to be large, the insignificance can still be explained by the fact that the difference of the ETF price and NAV is close to zero. The large relative values indicates that the euro difference is relatively large compared to the NAV.

It can be concluded, that during the four-year time period there are days when the ETFs are significantly mispriced compared to their respective NAV, and that the mean of paired sample differences are statistically significant. This indicates that there are days when the authorized participants could profit through creation / redemption process as mispricing exposes the ETFs for arbitrage possibilities. Charters (2013) found similar, statistically significant, average relative deviations in her study as the magnitude of the relative price deviation are similar to the findings of this study. One must keep in mind that there are also trading costs involved in arbitrage and the role of bid-ask spread is to play as a proxy for these costs. The wider the bid-ask spread the more expensive the exploitation of mispricing becomes. In this aspect there are some individual ETFs that obtain an average mispricing higher than the bid-ask spread which would indicate that the mispricing of these ETFs is economically significant. These ETFs are Lyxor CAC 40, Amundi CAC 40 and iShares FTSE 100. When comparing the mean mispricing of the whole data sample of the four-year time period, the bid-ask spread exceeds the average daily mispricing which is an indication that the average mispricing of the whole data sample is not economically significant, although statistically significant.

5.3 Mispricing persistency

ETFs can exhibit high premium/discount, and this could be due to high volatility, but might not last long for investors to capitalize on. Hence modeling mispricing persistency plays a key role in this thesis. Mispricing persistency was modelled with AR-model for each individual ETF. Since AR-models assumes for time stationarity, ADF-test was implemented to test for stationarity and the results can be seen in appendix 3. The null hypothesis is that the time series has a unit root, meaning that the data is non-stationary. As can be seen, the p-value for every individual ETF is under 0.01 which means that they fall below the 95% significance level, meaning that the null hypothesis of the unit root test can be rejected, and the data sample is stationary.

Based on the results found by use of ADF-test, the AR-model can be utilized. The procedure of AR modelling followed such logic where a future value was regressed to its previous value until the next lag obtained an insignificant beta coefficient. The modelling continued up to the fifth lag, which was used as the maximum of lags as it replicates one trading week. If any added lag made the previous beta coefficient insignificant, the modelling stopped. The results of AR model can be seen in table 7. for each ETF individually. The table consists of the intercept value and beta coefficients from one to five representing the lags. In the table, p-value is marked in parenthesis and the coefficient value is above.

Following the alpha coefficient interpretation of Charteris (2013), a significant alpha coefficient value alpha can be interpreted as significant average premium/discount. The results show that eight out of nine ETFs trades on average premium in comparison to their respective NAV of which five are statistically significant. Only one ETF, DNB OBX obtain negative alpha coefficient value, meaning that it trades at discount on average, compared to its respective NAV. These values are similar to average price deviations presented in table 5. and table 6, only difference being that XACT OBX did not obtain statistically significant average price deviation in autoregression.

Table 7. The results of autoregressive model with 1 to 5 lags for each ETF. * Means that the value is significant with 95% confidence level. Blank cell after a significant coefficient value indicates that the added lag made previous beta insignificant, and the modeling is stopped.

ETF	α	β1	β2	β3	β4	β5
Lyxor CAC 40	**0.0052*** **(0.0005)**	**0.4589*** **(0.0000)**	**0.1807*** **(0.0000)**	**0.1154*** **(0.0008)**	**0.0897*** **(0.0038)**	
Amundi CAC 40	**0.0126*** **(0.0000)**	**0.5941*** **(0.0000)**	**0.1428*** **(0.0000)**	0.0403 (0.1945)		
BNP Easy CAC 40	**0.0018*** **(0.0000)**	**0.5739*** **(0.0000)**	-0.0171 (0.5833)			
Lyxor Euro Stoxx 50	**0.0173*** **(0.0000)**	**0.6220*** **(0.0000)**	**0.1642*** **(0.0000)**	0.0403 (0.1945)		
UBS Euro Stoxx 50	0.0006 (0.9668)	**0.0735*** **(0.0181)**	0.0038 (0.9032)			
iShares FTSE 100	**0.0036*** **(0.0000)**	**0.5735*** **(0.0000)**	**0.2066*** **(0.0000)**	**0.0860*** **(0.0055)**		
BBVA Ibex 35	0.0010 (0.1567)	**0.2711*** **(0.0000)**	0.0335 (0.2806)			
XACT OBX	0.0019 (0.0584)	**0.1858*** **(0.0000)**	-0.0009 (0.9776)			
DNB OBX	**-0.0012*** **(0.0000)**	**0.0815*** **(0.0089)**	0.0206 (0.5076)			

In this autoregression analysis the beta coefficient represents the persistency of premium/discount. Based on the overall results, it can be seen that every ETF in the data sample have at least one significant beta, which indicates that mispricing persists at least for one trading day. In four out of nine ETFs the second lagged value has explanatory power, which counts almost half of the sample. ETF pricing is considered to be inefficient if premium/discount persistency endures over one day, but Charteris (2013) argues that two-day persistence exposes ETFs for only moderate arbitrage opportunities, and it is not a major sign of inefficiency. Two ETFs have more than two significant lagged beta coefficients. For iShares FTSE 100 the third lagged value has explanatory power, meaning that the premium/discount is persistent at least for three days. Similarly, for Lyxor CAC 40 the fourth lagged value have explanatory power over the ETF price, which indicates that the premium/discount is persistent at least for four days. Continuing the logic of the argument made by Charteris (2013), longer premium/discount persistence exposes iShares FTSE 100 and Lyxor CAC 40 ETFs to higher risk of arbitrage possibilities through creation/redemption process. Hence, they are considered to be inefficiently priced. It should be noted that high persistence does not necessarily mean high mispricing when average premium/discount is low. If an ETF exhibits low average premium/discount and low volatility of price deviations it may not be attractive for authorized participants.

5.4 OLS regression test for relationship

We have now computed and analyzed whether there is significant mispricing to be detected and the persistence of possible mispricing. With the OLS linear regression, the relationship between the ETF price and NAV can be analyzed. The results of the OLS regression are illustrated in table 6. Statistical significance of a coefficient is determined with 95% confidence level, meaning that if p-value exceeds 5% this value is insignificant.

Table 8**.** OLS regression results. * Means statistical significance with 95% confidence level and the t-values are for testing H0: beta = 1

ETF	α	t-value	β	t-value	R^2
Lyxor CAC 40	**-0.0752***	-3.286	**1.0026***	-6.14	1.000
Amundi CAC 40	**-0.3584***	-14.5236	**1.0072***	-18.91	1.000
BNP Easy CAC 40	**0.0153***	-3.5192	**1.0024***	-5.02	1.000
Lyxor Euro Stoxx 50	0.0048	0.3644	1.0001	-0.73	1.000
UBS Euro Stoxx 50	**0.3195***	2.0253	**0.9910***	-1.91	0.997
iShares FTSE 100	**0.0250***	3.6893	1.0002	-1.06	0.999
BBVA Ibex 35	0.0079	1.0144	0.9993	0.03	0.999
XATC OBX	**0.0297***	4.2616	**0.9959***	2.88	0.999
DNB OBX	-0.0033	-1.827	1.0003	-1.36	1.000

Since the relationship between the ETF price and NAV is being studied, significant betas are expected. We are analyzing if beta coefficient obtains a value of one which is an indication of unity. The null hypothesis is that beta equals one. A close unity and an insignificant alpha coefficient indicate a close relationship between the ETF price and NAV. The results presented in table 8 suggests that only two out of nine ETFs of the data sample obtains a beta coefficient value below one. These ETFs are UBS Euro Stoxx 50 and XATC OBX. This means that for example, if NAV of UBS Euro Stoxx 50 moves by one unit UBS Euro Stoxx 50 ETF price would move to the same direction but by 0.9910 units. XACT OBX price would move 0.9959 and so on. Of these three mentioned ETFs, only UBS Euro Stoxx 50 obtains a statistically significant beta value, meaning that it is statistically different from the value of one. The same logic goes the other way as well. When NAV moves by one unit, Lyxor CAC 40 moves to the same direction but by 1.0026 units. Six out of nine ETFs obtain a beta value above, and three of these ETFs, Lyxor CAC 40, Amundi CAC 40 and BNP Easy CAC 40, are found to be statistically significant from one at the 95 % significance level. Therefore, the beta values of these three mentioned ETFs are seen to be statistically different from one. None of the ETFs in the data sample are in unity as none of the ETFs obtains a value of on. Overall, four out of nine ETFs are seen to be insignificantly different from one, meaning that these ETFs are close to unity. The relationships between ETF prices and NAVs are illustrated in figure 4, which is a scatter plot with red regression line. It can be noted that the ETF prices and NAVs are not very scattered, and they are quite linear compared to the red regression line. UBS Euro Stoxx 50 is an exception as the daily ETF prices and NAVs are more scattered compared to others, but still obtain a linear form.

As only one explanatory variable was used in this regression the R-squared (R^2) represents the correlation between ETF price and its NAV. For every ETF, the R-squared reaches a value of 1.000 or very close to it. It also evaluates the scatter of the data points around regression line. As UBS Euro Stoxx 50 obtained an R-squared value of 0.997, it can be seen from figure 4 that the data points are more widely scattered than for the other ETFs as they obtained values between 0.999 and 1.000. The higher the R-squared value is the smaller are the differences between the observed data and the fitted values. Overall, considering the significant betas that are close to unity and R-squared values of 1.000 or close to it, these indicate a strong linear relationship between daily NAV and daily ETF closing prices.

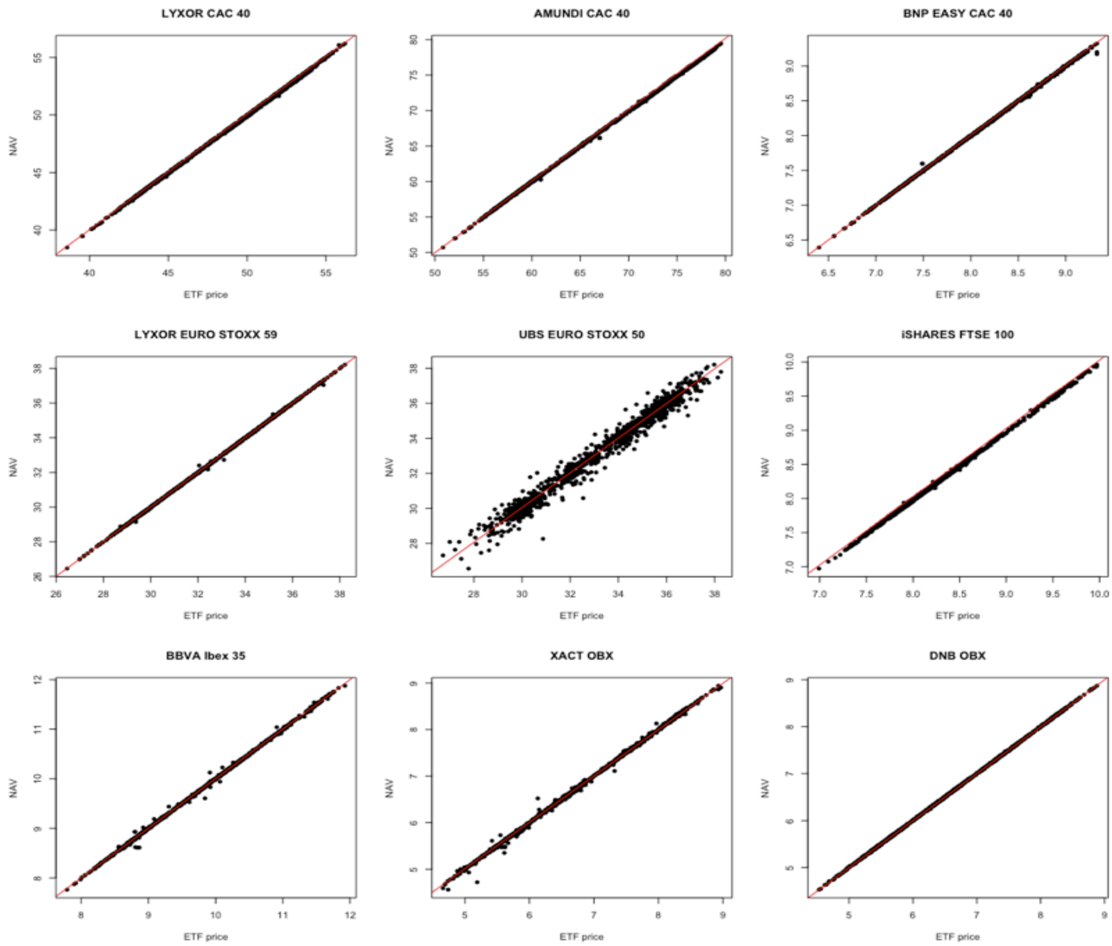

Figure 4. OLS regression results where red line represents the regression line.

5.5 ETF Tracking Ability & Performance

In this section the daily and monthly ETF returns are observed in comparison to daily index returns it replicates in order to find how well the ETF performs compared to the respective index. This is being referred as tracking efficiency. As one form of measuring tracking efficiency is a basic comparison of ETF returns and index returns, figure 5 illustrates a comparison of daily ETF returns and daily index returns. It can be noted that how different the return deviations are. Four of the nine ETFs follow a narrower path, and they seem less volatile, but they also experience more aggressive spikes

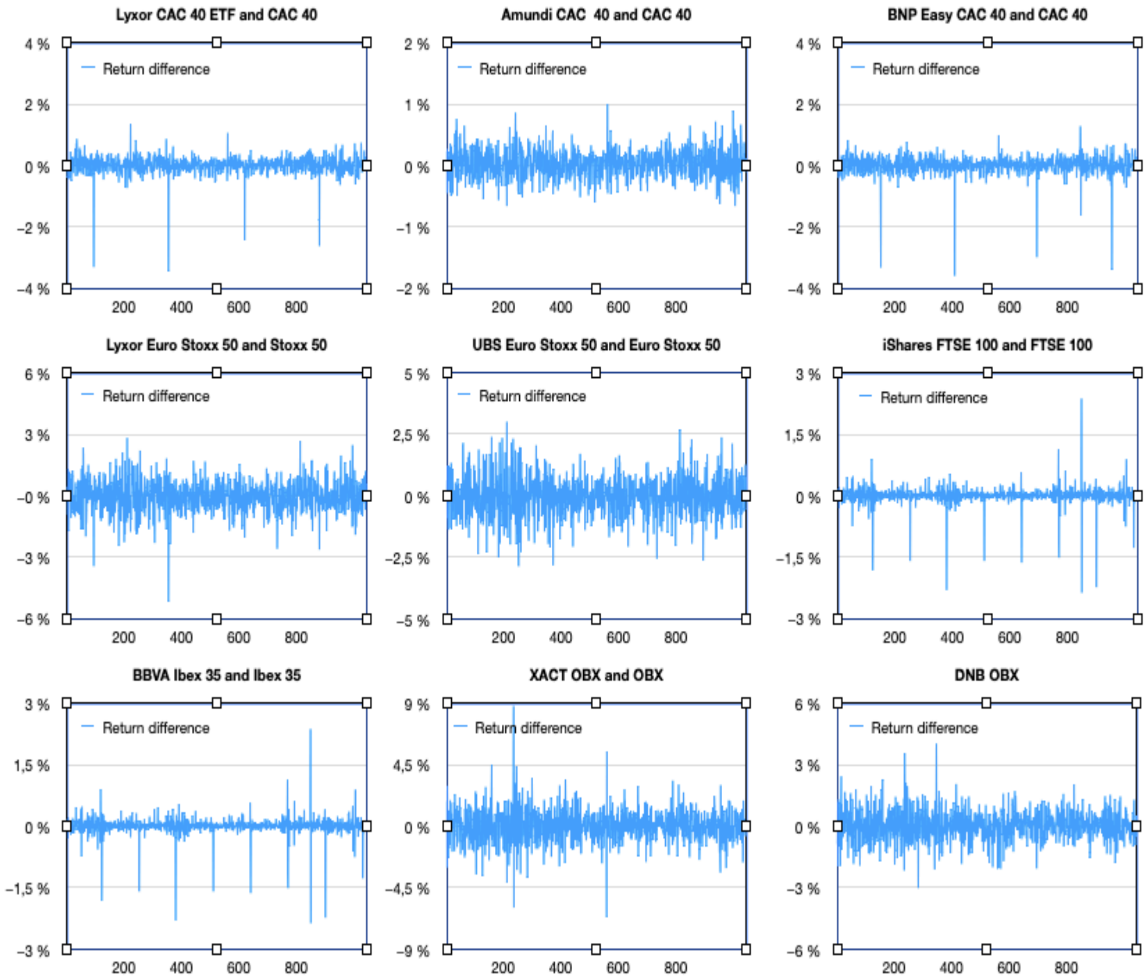

Figure 5. Daily return differences between the ETF and the respective index.

To put this figure in numbers, first a paired sample t-test is implemented in order to investigate whether sample means of ETF returns and index returns are significantly different from zero. The Hypothesis is that the mean of paired differences equals zero, which would indicate that ETFs would be track their respective indexes efficiently if the null hypothesis is accepted. Significance is being evaluated with two-tailed p-value. Statistical significance is measured with 95 % confidence level. Table 9 provides a detailed analysis of return deviations between ETFs and their respective indexes.

Table 9. Daily return difference statistics between ETF returns and Index returns. *Means statistical significance with 95% confidence level and the null hypothesis is rejected. Values are in percentages.

ETF	Min.	Max.	Mean	Std.	t-value	p-value	obs.
All	-7.3394	8.7571	0.0040	0.0074	0.6363	0.5247	9387
Lyxor CAC 40	-3.4357	1.3396	0.0127	0.0029	1.4110	0.1585	1043
Amundi CAC 40	-0.6580	0.9852	0.0259	0.0022	3.7439	**0.0002***	1043
BNP Easy CAC 40	-3.5763	1.2664	0.0124	0.0031	1.2740	0.2030	1043
Lyxor Euro Stoxx 50	-5.1638	2.7955	-0.0289	0.0076	-0.1100	0.9124	1043
UBS Euro Stoxx 50	-2.8575	2.9883	-0.0259	0.0080	-0.1649	0.8690	1043
iShares FTSE 100	-3.6106	2.4626	-0.0188	0.0060	0.1062	0.9155	1043
BBVA Ibex 35	-2.3614	2.3577	0.0000	0.0024	0.4959	0.6201	1043
XACT OBX	-6.6438	8.7571	0.0289	0.0119	-0.2215	0.8247	1043
DNB OBX	-7.3394	6.1966	0.0254	0.0090	-0.2454	0.8062	1043

Considering the overall data sample, daily return differences are between an interval of -7.34 % and 8.76 %. A negative value means that ETF gained lower return on a certain day compared to the index, therefore positive value indicates ETF over performed the index on a certain day. An interesting point is that both these minimum and maximum values belong to ETFs that replicate the same index, XATC OBX and DNB OBX. In the eyes of an investor, for a daily minimum and maximum values these differences might seem as significant but as one can see from the table that individual average return differences are between -0.029 % & and 0.029 % during the whole time period. Average the daily ETF returns are for the whole data sample are close to index returns at the difference is only 0.0040 %. When comparing other ETFs that replicate the same index the results indicate that these ETFs have similar return differences on average as both, Lyxor Euro Stoxx 50 and UBS Euro Stoxx 50, gained a negative average return differences with only 0.003 % from each other as well as similar standard deviations. Also, CAC 40 replicating ETFs gained similar return differences, except for Amundi CAC 40, which have lower average return difference and lower standard deviation. Amundi CAC 40 is also the only ETF that have a statistically significant average return difference as the p-value fall well below 5%. Hence the paired-sample t-test indicates that the average return difference between Amundi CAC 40 ETF and CAC 40 index are different from zero and the null hypothesis can be rejected. This significance could be explained with the fact of low standard deviation and high average return difference compared to its respective index.

5.5.1 Tracking Error

Previous literature suggests that higher frequency, such as daily data, can cause overestimation of tracking error. This is, because daily data could cause negative serial correlation between ETFs and their respective indexes (Elia, 2011). This can be noted when comparing R-squared values of table 9 to appendix 8. The monthly returns obtain significantly better adequacy in terms of R-squared values. For this reason, monthly data is used in all the implemented tracking error methods. Table 10 illustrates the results of three tracking error calculations that were implemented for every ETF individually. The first column represents the tracking error as absolute difference between ETF returns and index returns on monthly average, and the second column represents tracking error as standard deviation of the absolute difference of returns. The third column presents tracking error as residual standard error (TESE) of a regression where ETF return was the dependent variable and index price the explanatory variable.

When we look at TEABS values we can see that the tracking error as absolute difference in terms of monthly average returns vary between 0.3353 % and 0.7363 % between the ETFs, Amundi CAC 40 obtaining the lowest absolute difference in returns and UBS Euro Stoxx 50 the highest absolute difference compared to their respective index. This means that the distance monthly average returns of Amundi CAC 40 is 0.3353% from the returns of CAC 40 index. Similarly, the distance for UBS Euro Stoxx 50 and Euro Stoxx 50 index is 0.7364 %. The absolute difference gives a good indication of the magnitude of the return difference between ETFs and their respective indexes. When considering the volatility of the monthly absolute difference of returns, TESD values are relevantly low as the standard deviations variate between 0.0038 % and 0.0095 % for all the ETFs, thus all ETFs obtaining similar standard deviation values. This would indicate that the distribution of the monthly absolute return differences of the sample ETFs does not variate much. In addition, values of TESE are nearly identical to those of TESD. As suggested by Frino and Gallagher (2002) TESD value should provide similar values as the standard residual error TESE.

Table 10. Tracking error results. TEABS is tracking error in form of absolute difference, TESD is tracking error standard deviation and TESE is tracking error in form of residual standard error from $R_{i,t} = \alpha + \beta_1 R_{b,t} + \varepsilon$

ETF	TEABS	TESD	TESE
Lyxor CAC 40	0.5014 %	0.0083 %	0.0094
Amundi CAC 40	0.3353 %	0.0043 %	0.0048
BNP Easy CAC 40	0.6083 %	0.0095 %	0.0114
Lyxor Euro Stoxx 50	0.5051 %	0.0080 %	0.0092
UBS Euro Stoxx 50	0.7364 %	0.0079 %	0.0106
iShares FTSE 100	0.5543 %	0.0038 %	0.0068
BBVA Ibex 35	0.6289 %	0.0058 %	0.0083
XACT OBX	0.6338 %	0.0081 %	0.0101
BNB OBX	0.5040 %	0.0075 %	0.0088
Min.	**0.3353 %**	**0.0038 %**	**0.0048**
Max.	**0.7364 %**	**0.0095 %**	**0.0114**

5.5.2 ETF Performance

Table 11 illustrates the results of the regression where daily ETF return is the depend variable and daily index return is the explanatory variable. As Alamelu and Goyal (2022) suggests alpha coefficient indicates relative excess returns gained by the ETF in comparison to its respective index and beta coefficient is an estimate of whether ETF returns are in unity with returns of the respective index. As the beta coefficient represents the slope, it is the estimate for systematic risk. Therefore, it implies the sensitivity of an ETF to the movement of the respective index. It is also the indicator of replication strategy. A beta coefficient of one would indicate a full replication and that they are in unity. A significant beta different from unity indicates that the ETF does not practice full replication strategy.

As we delve deeper into the obtained regression values, one can notice that most of the alpha coefficients do not exhibit statistically significant values with the 95% confidence level as expected. Amundi CAC 40 is the only ETF that presents a significant alpha coefficient, meaning that it obtains relative excess returns compared to the respective CAC 40 index. A negative alpha would indicate an ETF underperformance compared to the respective index. When we look at beta coefficient values, it can be noted that three out of nine ETFs, BNP Easy CAC 40, iShares FTSE 100 and BBVA Ibex 35, obtain a value of higher than one, meaning that these ETFs are above unity. Therefore, when their respective index moves one unit, the ETF moves the amount of the beta value to the same direction, BBVA Ibex 35 would move 1.0534 units and so on. As the null hypothesis is that beta equals one, it can be seen from table 11 that only one ETF with beta above one is statistically significant at the 95 % significance level. This ETF is BBVA Ibex 35, meaning that the beta value is significantly different from value of one. Six out of nine ETFs obtained beta coefficient under unity, meaning that these ETFs move to the same direction as the index moves by one unit, but to a lesser extent. Three out of these six ETFs that obtain beta values below one, are statistically significant, meaning that they are statistically different from value of one at the 95 % confidence level. Overall, four out of nine ETFs obtain a statistically significant beta coefficient. Statistically significant beta value is an indication that the ETF does not fully replicate the respective index. An interesting notion is that how differently ETFs can move even if they track the same respective index, BNP Easy CAC 40 ETF being the example as it obtains beta value over one as Lyxor and Amundi ETFs does not.

Table 11. ETF performance regression results. * Means statistical significance with 95% confidence level. t-values are for H0: beta = 1.

ETF	α	t-value	β	t-value	R^2
Lyxor CAC 40	0.0000	0.0334	**0.9350***	-1.8357	0.9406
Amundi CAC 40	**0.0026***	3.6118	0.9908	-0.5093	0.9855
BNP Easy CAC 40	-0.0004	-0.2475	1.0217	0.5061	0.9284
Lyxor Euro Stoxx 50	-0.0003	-0.1834	**0.9392***	-1.8529	0.9490
UBS Euro Stoxx 50	0.0010	0.6133	**0.9497***	-1.3304	0.9348
iShares FTSE 100	0.0000	0.0129	1.0009	0.0332	0.9686
BBVA Ibex 35	0.0005	0.3992	**1.0534***	1.9226	0.9703
XACT OBX	0.0023	1.5092	0.9838	-0.4974	0.9540
DNB OBX	0.0025	1.9022	0.9934	-0.2312	0.9654

Overall, it can be understood that daily returns deviate between –7.34 % and 8.76 %, but on average the return deviation is rather low 0.0040 % during the whole time period. This means that considering the whole data sample, the ETFs obtain higher daily returns on average compared to the indexes. Individually, three out of nine ETFs obtained lower average daily returns compared to their respective index as the average price deviation is negative over the four-year time span. Paired-sample t-test found the return deviations to be statistically insignificant, except for Amundi CAC 40. The significance of Amundi CAC 40 could be explained by high average return deviation and low standard deviation compared to its respective index. To conclude the tracking error calculations, it is obvious that funds does not perfectly replicate their respective index as the absolute return differences vary between 0.3353 % and 0.7364 % and the beta coefficients of the regression are not in unity, but very close.

In order to understand the magnitude of the results of TEABS, a comparison to previous research is made. Frino & Ghallager (2000, 2001) documented tracking error between 0.039 % and 0.110 % in terms of average monthly absolute return difference for the US ETFs and for Australian ETFs the values vary between 0.074 % and 0.224 %. Shin and Soydemir (2010) found variation for European ETFs between 0.020 % and 0.084 %. Ackert et al. (2017) found that average monthly absolute return difference in New Zealand varies between 1.20 % and 1.13 %. Tripahti and Sethi (2021) find tracking error for Indian ETFs to variate between 0.22 % 2.03 %. It can be noted that some markets have a large variation in terms of TEABS. Ackert et al. (2017) concluded, that monthly tracking errors in terms of absolute return difference are quite significant and that ETFs and that the ETFs used in their study does not perfectly replicate their respective index. Tripathi and Sethi (2021) came to a similar conclusion. Comparing the findings of this thesis to previous research, it can be concluded that the ETFs used in this thesis do not perfectly replicate their respective indexes and that the tracking errors in terms of absolute return differences are significant.

Summary & Conclusion

The objective of this thesis was to discover whether the ETFS are efficiently priced compared to their underlying assets and how well does the ETFs track their respective indexes. The pricing efficiency was measured with euro difference and relative difference between the closing price of an ETF and its respective NAV. The data consisted of nine European ETFs with daily data for a time period of four years from February 2015 to February 2019. This thesis has given theoretical insight of market efficiency and about Exchange Traded Funds in general. The null hypothesis of this thesis was that the ETFs are efficiently priced and that there is no significant tracking error between the ETFs and their respective indexes. To find answer to the null hypothesis, four research questions was formed.

The first research question handled the phenomenon of mispricing in terms of the euro difference between the ETF price and the respective NAV as well as in terms of relative mispricing of the ETF compared to the respective NAV. The mispricing of seven out of nine ETFs were found to be statistically significant as the mean difference between the price of these seven ETFs were found to be significantly different from zero compared to the NAV. Time to time ETFs provided some arbitrage opportunities for APs to capitalize on. Still most ETFs were not found to be economically significant as the daily average mispricing was compared to the bid-ask spread which was used as a proxy for trading costs.

Finding evidence for mispricing continued to examine whether there is evidence of persistent mispricing which was tested with AR-models. The results of autoregression suggests that for most of the ETFs the mispricing was persistent for at least two days. Charteris (2013) argues that two-day persistency is not an indication of severe pricing inefficiency, but it exposes an ETF for arbitrage opportunities that investors could exploit with active trading strategies. These findings lead to a conclusion that mispricing of ETFs is persistent to some degree.

The third research question considered the relationship between the ETF price and the NAV. This was evaluated with OLS regression. Four out of nine ETFs obtained statistically significant beta coefficients, which indicates that these are statistically different from one. The important thing about the beta coefficient was that it also represented whether the ETF price and the NAV were in unity. All the beta estimates obtained a value close to one, some below and some a little bit above. This indicated that there is a close relationship between the ETF price and the NAV. Also, R-squared values were very close to one which backs up the finding of close unity and indicates a strong linear relationship and that ETF price and the respective NAV are strongly correlated.

As equity ETFs are replicating an index, the last part of this study focused on the tracking ability and performance of the ETFs used in this study. Firstly, the daily return differences were evaluated with t-test and as expected most of the average return differences were found to be statistically insignificant, meaning that the mean return difference was not significantly different from zero. As suggested by previous literature of Elia (2011), three tracking error measures were implemented for monthly return differences. The values of the three tracking error methods used were compared to the findings of related research and the conclusion was that the ETFs do not perfectly replicate their respective index and the presented tracking errors are significant.

Overall, it can be concluded that the ETFs used in this study provides arbitrage opportunities from time to time as there are days when ETFs trade at substantially high premium/discount. This is backed by the findings of AR-model as the mispricing is persistent for at least two-days for most of the ETFs. Although the mispricing was found to be statistically significant, the magnitude of the mispricing on daily average is low for most of the ETFs and falls below the trading costs making the mispricing economically insignificant. The European ETFs used in this study are found to be efficiently priced, but as the tracking error calculations revealed, the ETFs does not replicate their respective index perfectly and perform significant tracking errors.

The ETF industry is growing rapidly, and new ETFs are flowing to the market constantly. Crypto ETFs being one of the newest members of the ETF family. This creates researchers plenty of new opportunities to conduct new studies. As daily price deviation was found among ETFs in this study, it would be interesting to dive deeper and study what determinants were the cause for the price to deviate. Also, forming a trading strategy would give better insight whether arbitrage would have been profitable for investors and in what magnitude.

Arguably a trading strategy would be better to construct with high-frequency data as the price of an ETF fluctuates throughout the day as any instrument in the stock market. As for new research it would be interesting to study different types of ETFs domiciled in Europe, as many existing studies considering European ETFs have been conducted by using international ETFs. It is known that the price of international ETFs tend deviate more than local ETFs. Therefore, using ETFs from the same market or markets that have time-zones close to each other would be more comparable in order to evaluate the state of the market in terms of mispricing.

REFERENCES

Ackert, L. F. & Tian, Y. S. (2000) Arbitrage and Valuation in the Market for Standard and Poor's Depositary Receipts. *Financial management*. 29 (3), 71–87.

Alamelu, L. L. & Goyal, N. N. (2022) Investment Performance and Tracking Efficiency of Indian Equity Exchange Traded Funds. *Asia-Pacific financial markets*. 1–24.

Antoniewicz, R. & Heinrichs, J. (2014) "Understanding Exchange-Traded Funds: How ETFs Work." ICI Research Perspective 20, no. 5 (September) [online]. Available at: www.ici.org/pdf/per20-05.pdf. [Accessed 22.11.2022]

Ben-David, I., Franzoni, F., & Moussawi, R. (2012). ETFs, arbitrage, and contagion. University of Geneva.

Ben-David, I., Franzoni, F. & Moussawi, R. (2016). Exchange Traded Funds. Working paper no. 22829. (National Bureau of Economic Research).

Blackrock. (2017). A primer on ETF primary trading and the role of authorized participants. [online]. Available at: https://www.blackrock.com/corporate/enat/literature/whitepaper/viewpoint-etf-primary-trading-role-of-authorized-participantsmarch-2017.pdf. [Accessed 22.11.2022]

Blitz, D. and Huij, J. (2012), "Evaluating the performance of global emerging markets equity exchange-traded funds", Emerging Markets Review, Vol. 13, pp. 149-158.

Brooks, C. (2008) *Introductory econometrics for finance*. 2nd ed. Cambridge: Cambridge University Press.

Bundesbank, D. (2018). The growing importance of exchange traded funds in the financial markets. Monthly Report.

Charupat, N. and Miu, P. (2013). Recent developments in exchange-traded fund literature: Pricing efficiency, tracking ability, and effects on underlying securities, Managerial Finance, Vol. 39 No. 5, pp. 427-443.

Cheung, Y.-W. & Lai, K. S. (1995) Lag Order and Critical Values of the Augmented Dickey-Fuller Test. *Journal of business & economic statistics*. 13 (3), 277–280.

Chu, P. K. K. (2011). Study on the tracking errors and their determinants: evidence from Hong Kong exchange traded funds. Applied Financial Economics, 21(5), 309-315.

Deville, L. (2008). Exchange traded funds: History, trading, and research. In Handbook of financial engineering (pp. 67-98). Springer, Boston, MA.

Delcoure, N. & Zhong, M. (2007) On the premiums of iShares. Journal of empirical finance. 14 (2), 168–195.

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

Elia, M. (2012). Tracking Error of Traditional and Synthetic European Exchange-Traded Funds.

Elton, E. J. et al. (2002) Spiders: Where Are the Bugs? *The Journal of business (Chicago, Ill.)*. 75 (3), 453–472.

Fama, E.F., (1970). Efficient capital markets: a review of theory and empirical work. Journal of Finance, 25(2), pp.383-417.

Foucher, I. and Gray, K., (2014). Exchange-traded funds: Evolution of benefits, vulnerabilities and risks. Bank of Canada Financial System Review, pp.37-46.

Frino, A., & Gallagher, D. R. (2001). Tracking S&P 500 index funds. The Journal of Portfolio Management, 28(1), 44-55.

Frino, A., & Gallagher, D. R. (2002). Is index performance achievable? An analysis of Australian equity index funds. Abacus, 38(2), 200-214.

Gabriela ğiĠan, A., (2015). The efficient market hypothesis: Review of specialized literature and empirical research. Procedia Economics and Finance, 32, pp.442-449.

Gastineau, G.L. (2002), Equity index funds have lost their way, Journal of Portfolio Management, Vol. 28 No. 2, pp. 55-64.

Goel, G. & Ahluwalia, E. (2021) Do pricing efficiencies in Indian equity ETF market impact its performance? *Global finance journal*. Vol 49.

Hair, J. F. go. (2017) *A primer on partial least squares structural equation modeling (PLS-SEM)*. Second edition. Los Angeles: Sage.

Harper, J. T., Madura, J., & Schnusenberg, O. (2006). Performance comparison between exchange-traded funds and closed-end country funds. Journal of International Financial Markets, Institutions and Money, 16(2), 104-122.

Hill, J.M., Nadig, D. & Hougan, M. 2015. A comprehensive guide to exchange-traded funds (ETFs). CFA Institute Research Foundation.

Johnson, B., Bioy, H., Kellett, A. and Davidson, L., 2013. On the right track: Measuring tracking efficiency in ETFs. *The Journal of Index Investing*, *4*(3), pp.35-41.

Kayali, M. (2007). Pricing efficiency of ETFs in Turkey: Early evidence from Dow Jones Istanbul 20. International Research Journal of Finance and Economics. pp. 15-22.

Kayali, M.M. & Ozkan, N. 2012. Does the Market Misprice Real Sector ETFs in Turkey? International Research Journal of Finance and Economics, issue 91, pp. 157- 160.

Kreis, Y. & Licht, J. W. (2018) Trading on ETF mispricings. *Managerial finance*. 44 (3), 357–373.

Kostovetsky, L. (2003). Index Mutual Funds and Exchange-Traded Funds (Digest Summary). Journal of Portfolio Management, 29 (480-92).

Kumar, P. (2018) An Investigation of Pricing Efficiency of CPSE ETF in India. *Journal of commerce and accounting research*. 7 (2), 28–.

Lakshmi, V. (2022) Do Exchange Traded Funds in India Have Tracking and Pricing Efficiency? *Advances in decision sciences*. 26 (2), 1–25.

Malkiel, B.G. and Fama, E.F., (1970). Efficient capital markets: A review of theory and empirical work. The journal of Finance, 25(2), pp.383-417.

Martin, W. E. & Bridgmon, K. D. (2012) Quantitative and statistical research methods: from hypothesis to results. San Francisco, CA: Jossey-Bass.

Milonas, N. T. & Rompotis, G. G. (2013) Does intervalling effect affect ETFs? Managerial finance. 39 (9), 863–882.

Naumenko, K., & Chystiakova, O. (2015). An empirical study on the differences between synthetic and physical ETFs. *International Journal of Economics and Finance*, *7*(3), 24-35.

Petäjistö, A. (2017). Inefficiencies in the Pricing of Exchange-Traded Funds, Financial Analysts Journal, vol. 73, issue 1, pp. 24-54.

PricewaterhouseCooper. (2022). ETFs 2026: The Next Big Leap. [online] https://www.pwc.com/gx/en/financial-services/publications/assets/ETF_2026_PwC.pdf [accessed 23.11.2022]

Pope, P. F. & Yadav, P. K. (1994) Discovering Errors in Tracking Error. *Journal of portfolio management*. 20 (2), 27–32.

Poterba, J. M., & Shoven, J. B. (2002). Exchange traded funds: A new investment option for taxable investors (No. w8781). National Bureau of Economic Research.

Ringwalt, C. et al. (2011) The Use of One- Versus Two-Tailed Tests to Evaluate Prevention Programs. Evaluation & the health professions. 34 (2), 135–150.

Roll, R. (1992) A Mean/Variance Analysis of Tracking Error. *Journal of portfolio management*. 18 (4), 13–22.

Ross, S. A., Westerfield, R W. and Jaffe, J. (2013). Corporate finance. 10th ed. McGrawHill.

Ross, A. (2017). Basic and Advanced Statistical Tests: Writing Results Sections and Creating Tables and Figures.

Shin, S. and Soydemir, G. (2010), “Exchange-traded funds, persistence in tracking errors and information dissemination”, Journal of Multinational Financial Management, Vol. 20, pp. 214-234.

Stone, E. (2010) t Test, Paired Samples. Encyclopedia of Research Design 3 p.1560–1565.

Tähtinen, J. et al. (2020) *Tilastollisen aineiston käsittelyn ja tulkinnan perusteita*. 2. uudistettu painos. Turku: Turun yliopiston kasvatustieteiden laitos.

Vanguard. (2013). Understanding synthetic ETFs. [online]. Available at: https://www.vanguard.ca/documents/understanding-synthetic-etfs-tlrv.pdf [Accessed 22.11.2022]

Yan, X. & Su, X. (2009) *Linear regression analysis theory and computing*. Singapore World Scientific.

APPENDICES

Appendix 1. First half of frequency distribution table of price deviations in euro for individual ETFs

€ Deviation	Lyxor CAC 40	Amudi CAC 40	BNP Easy CAC 40	Lyxor Euro Stoxx 50
≤ - 0,8	0	0	0	0
> -0,8 and ≤ -0,6	0	0	0	0
> -0,6 and ≤ -0,4	0	0	0	0
> -0,4 and ≤ -0,2	1	1	0	1
> -0,2 and ≤ 0	211	168	290	359
> 0 and ≤ 0,2	828	690	754	678
> 0,2 and ≤ 0,4	4	179	0	6
> 0,4 and ≤ 0,6	0	1	0	0
> 0,6 and ≤ 0,8	0	2	0	0
≥ 0,8	0	3	0	0

Appendix 2. Second half of frequency distribution of price deviations in euro for every individual ETFs

€ Deviation	USB Euro Stoxx 50	iShares FTSE 100	BBVA Ibex 35	XATC OBX	DNB OBX
≤ - 0,8	17	0	0	0	0
> -0,8 and ≤ -0,6	27	0	0	0	0
> -0,6 and ≤ -0,4	60	0	0	0	0
> -0,4 and ≤ -0,2	144	0	1	1	0
> -0,2 and ≤ 0	270	34	495	650	823
> 0 and ≤ 0,2	260	1010	545	390	221
> 0,2 and ≤ 0,4	130	0	3	2	0
> 0,4 and ≤ 0,6	70	0	0	1	0
> 0,6 and ≤ 0,8	35	0	0	0	0
≥ 0,8	31	0	0	0	0

.

Appendix 3. First half of frequency distributions of relative price difference for every individual ETF

% Deviation	Lyxor CAC 40	Amudi CAC 40	BNP Easy CAC 40	Lyxor Euro Stoxx 50
≤ -5 %	0	0	0	0
> - 5 % and < - 2,5 %	0	0	0	0
> - 2,5 % and < - 1 %	0	0	3	1
> - 1 % and < - 0,5 %	0	0	0	4
> - 0,5 % and < 0 %	212	169	287	355
> 0 % and < 0,5 %	830	868	750	678
> 0,5 % and < 1 %	2	2	2	5
> 1 % and < 2,5 %	0	5	2	1
> 2,5 % and < 5 %	0	0	0	0
≥ 5 %	0	0	0	0

Appendix 4. Second half of frequency distributions of relative price difference for every individual ETF

% Deviation	USB Euro Stoxx 50	iShares FTSE 100	BBVA Ibex 35	XATC OBX	DNB OBX
≤ -5 %	0	0	0	1	0
> - 5 % and < - 2,5 %	19	0	0	2	0
> - 2,5 % and < - 1 %	125	0	9	7	0
> - 1 % and < - 0,5 %	140	0	5	21	0
> - 0,5 % and < 0 %	234	34	482	620	823
> 0 % and < 0,5 %	227	902	536	327	221
> 0,5 % and < 1 %	126	107	7	46	0
> 1 % and < 2,5 %	141	1	3	14	0
> 2,5 % and < 5 %	30	0	2	5	0
≥ 5 %	2	0	0	1	0

Appendix 5. Result of Augmented Dickey-Fuller test for relative price deviation.

ETF	p-value
Lyxor CAC 40	< 0,01
Amundi CAC 40	< 0,01
BNP Easy CAC 40	< 0,01
Lyxor Euro Stoxx 50	< 0,01
UBS Euro Stoxx 50	< 0,01
iShares FTSE 100	< 0,01
BBVA Ibex 35	< 0,01
XACT OBX	< 0,01
DNB OBX	< 0,01

Appendix 5. Daily ETF performance regression results.

ETF	Variable	Coefficient	SE	t-value	p-value	R2
Lyxor CAC 40	Alpha	0,0001	0,0001	1,4219	0,1554	0,931
	Beta	0,9550	0,0081	118,4960	**0,0000***	
Amundi CAC 40	Alpha	0,0003	0,0001	3,8126	**0,0001***	0,959
	Beta	0,9584	0,0061	156,2716	**0,0000***	
2BNP Easy CAC 40	Alpha	0,0001	0,0001	1,2793	0,2011	0,921
	Beta	0,9582	0,0087	110,1185	**0,0000***	
Lyxor Euro Stoxx 50	Alpha	-0,0002	0,0002	-1,0524	0,2928	0,595
	Beta	0,7859	0,0002	39,0859	**0,0000***	
UBS Euro Stoxx 50	Alpha	-0,0002	0,0002	-0,8628	0,3884	0,564
	Beta	0,7834	0,0214	36,6922	**0,0000***	
iShares FTSE 100	Alpha	-0,0002	0,0002	-0,9630	0,3358	0,728
	Beta	0,8705	0,0165	52,7462	**0,0000***	
BBVA Ibex 35	Alpha	0,0000	0,0001	-0,0424	0,9662	0,960
	Beta	0,9758	0,0062	158,4660	**0,0000***	
XATC OBX	Alpha	0,0003	0,0004	0,8024	0,4225	0,298
	Beta	0,8489	0,0404	21,0060	**0,0000***	
DNB OBX	Alpha	0,0003	0,0003	0,8557	0,3924	0,369
	Beta	0,8722	0,0353	24,6907	**0,0000***	