TECHNISCHE UNIVERSITÄT MÜNCHEN

LEHRSTUHL FÜR BETRIEBSWIRTSCHAFTSLEHRE -
FINANZMANAGEMENT UND KAPITALMÄRKTE

THREE ESSAYS ON EQUITY EXCHANGE-TRADED FUNDS

Friedrich Karl Yorck Henri Osterhoff

Vollständiger Abdruck der von der Fakultät für Wirtschaftswissenschaften der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Wirtschaftswissenschaften
(Dr. rer. pol.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Sebastian Schwenen

Prüfer der Dissertation:
1. Univ.-Prof. Dr. Christoph Kaserer
2. Univ.-Prof. Dr. Reiner Braun

Die Dissertation wurde am 30.03.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Wirtschaftswissenschaften am 15.06.2016 angenommen.

To my brother Georg

"So who still believes markets don't work?

Apparently it is only the North Koreans, the Cubans and the active managers."

Rex Sinquefield

SUMMARY

Using data on both equity ETFs listed in Germany and their underlying stocks, this thesis provides empirical evidence on key ETF-market interdependencies. First, it finds the tracking ability of an ETF to be significantly affected by the liquidity of its underlying stocks. Contrary to the notion that ETFs are shielded from creation- and redemption-related transaction costs, the thesis also shows that creation/redemption activity does affect tracking ability. Both effects might be attributable to imperfect replication of index weights. Second, the dissertation provides evidence that the hiring of an additional market maker significantly improves ETF liquidity, most probably because increased competition between existing market makers causes liquidity costs to decrease. Moreover, it finds market makers who rely predominantly on algorithmic or high-frequency trading to be systematically better in providing liquidity for ETFs. Third, the dissertation shows that creations/redemptions have a highly significant and economically viable effect on abnormal returns of the underlying stocks in the closing auction, an effect that is particularly pronounced in small stocks and on bullish trading days. Furthermore, the results suggest that, given the size of the potential additional income from exploiting this inefficiency, authorised participants not only have the opportunity but also the motivation to do so by actively manipulating prices during the closing auction.

TABLE OF CONTENTS – SUMMARY

TABLE OF CONTENTS

NOMENCLATURE

AP	authorised participant
ATT	average treatment effect on the treated
AuM	assets under management
CAC 40	Cotation Assistée en Continu 40
CAGR	compound annual growth rate
CBOE	Chicago Board Options Exchange
CCP	central counterparty
CDAX	Composite DAX
cf.	confer
cit.	cited
CRT	cost of round-trip
DAX	Deutscher Aktienindex
DMM	designated market maker
e.g.	exempli gratia
EMH	Efficient Market Hypothesis
EMU	Economic and Monetary Union of the European Union
et al.	et alii
ETF	exchange-traded fund
€	euro
HAC	heteroscedasticity and autocorrelation consistent
HFT	high-frequency trading
iNAV	indicative net asset value
ISE	Italian Stock Exchange
ISIN	International Securities Identification Number
KKMDB	Karlsruher Kapitalmarktdatenbank
LHS	left-hand side
MDAX	Mid-Cap-DAX
MPT	Modern Portfolio Theory

NASDAQ	National Association of Securities Dealers Automated Quotations
NAV	net asset value
NYSE	New York Stock Exchange
OTC	over the counter
p.a.	per annum
PR	price return
PSM	propensity score matching
RHS	right-hand side
SA	Société Anonyme
S&P	Standard & Poor's
SDAX	Small-Cap-DAX
SE	Societas Europaea
SPDR	Standard & Poor's Depository Receipts
SPY	SPDR S&P 500 ETF Trust
TecDAX	Technology DAX
TER	total expense ratio
TIPs	Toronto 35 Index Participation Units
TR	total return
TTT	Taiwan Top 50 Tracker
US	United States
US$	United States Dollar
VIF	variance inflation factor
vs.	versus
XLM	XETRA Liquidity Measure

LIST OF FIGURES

List of Tables

1. INTRODUCTION

Since their first appearance some 25 years ago, exchange-traded funds (ETFs) have undergone a truly remarkable transition from a mere niche product to one of the most successful innovations in asset management – an innovation whose profound and indeed disruptive impact on the asset management industry can only be compared to the advent of index funds, hedge funds or the mutual fund itself (McKinsey & Company, 2011). The wide range of ETF products available today provides access to a myriad of asset classes, geographical markets, and strategies that were once out of reach for retail or small institutional investors.

ETFs are predominantly passively managed, listed open-end funds that try to replicate the returns of an underlying benchmark portfolio, typically a particular equity or fixed-income index. Their overall structure allows investors to participate in the performance of the respective underlying basket with a level of transparency and at a cost usually not equalled by any equivalent mutual index fund. In addition to portfolio transparency and comparatively smaller management fees, ETFs also have the important advantage over mutual funds that their shares trade continuously throughout the day, at prices determined by supply and demand rather than at the net asset value (NAV) based on closing prices (Engle & Sarkar, 2006). The high fungibility of ETFs at market-determined prices also allows for quick and normally cost-effective entry and exit from an asset class, strategy or market (BaFin, 2012). Drawn by these prospects[1] of market access, transparency and liquidity at a relatively lower cost, both institutional and retail investors have ploughed billions of dollars into ETFs in recent years: As of December 2015, total global assets under management (AuM) in ETFs amount to US$ 2.81 trillion, up from a mere US$ 74 billion in 2000 (BaFin, 2012; Deutsche Bank Research, 2016).

1 See Hill, Nadig and Hougan (2015) for a detailed discussion of the benefits of ETFs compared to mutual funds.

Yet, while ETF growth remains a key driver of the asset management industry, the market shows clear signs of an incipient maturity. With most of the market opportunities and easy gains already exploited, it will become more difficult for ETF providers to attract enough capital to sustain the explosive growth of previous years. As competition intensifies and the market consolidates, the prospect looms that issuers will have to offer innovative products, such as active and Smart Beta ETFs, or retool existing products at the expense of their direct competitors. As a result, ETF providers are faced with the strategic challenge of differentiating their ETFs from rival products through price and quality. For seemingly homogenous products such as two ETFs mimicking the same reference index, the ability to track their benchmark and their market liquidity are central aspects of market quality and, aside from cost, are the key differentiators on which most institutional investors base their ETF selection decisions (Greenwich Associates, 2014). In light of this investor orientation, a better understanding of what drives ETF liquidity and tracking ability is essential for issuers in finding a competitive edge and securing their market positions. Two of the three essays presented in this thesis will therefore elaborate separately on these two critical components of market quality to determine in what ways they are affected by external market forces.

Although the US$ 2.81 trillion held in ETFs represent a minor fraction of total global AuM, they are invested in an extremely actively traded financial product, which for some represents the very epitome of the era of high-frequency trading (HFT) and its excesses.[2] On US stock exchanges, for instance, ETFs already account for more than

[2] Two exemplary events, which not only highlighted ETFs' crucial new role for financial markets but also raised concerns over whether the impact of ETFs on capital markets as a whole has yet been fully grasped, are the 2010 Flash Crash and the aftermath of China's Black Monday in 2015. During the 2010 Flash Crash, US markets experienced a very steep drop of almost 1,000 points in less than half an hour. Investigation revealed that exchange-traded products accounted for approximately two-thirds of the 21,000 cancelled trades during this period (SEC, 2010). In the aftermath, ETFs' potential role in transmitting price dislocations in distressed markets was widely discussed (OFR, 2013). On 24 August 2015, the Dow Jones Industrial Average suffered a 1,000-point plunge within the first five minutes of trading amid fears about China's economic slowdown. During that short period, many ETFs suffered even greater losses, trading at steep discounts to the actual NAV of their underlying portfolios. Furthermore. ETF trading was halted more than 1,000 times during the day, due to limit-up/limit-down rules imposed after the 2010 Flash Crash, accounting for approximately 85% of all trade halts on 24 August (SEC, 2015).

26% of the total daily equity trading volume, a proportion that can become substantially larger during periods of high volatility (Flood, 2015; Deutsche Bank Research, 2016). The question of whether ETFs themselves have gained enough momentum to have an impact on their underlying markets has been raised previously, but it remains unclear whether the structural design of ETFs causes fund-induced trading to have any effects on the underlying markets that are not otherwise observable in mutual funds. The third essay in this dissertation contributes to the understanding of these possible effects by extending the knowledge on how creation- and redemption-related transactions affect underlying stocks and their returns.

ETFs are truly complex constructs that are shaped by a myriad of dynamic interrelations with related markets and market participants; given that research on some of these dependencies has just commenced, it would be highly presumptuous and ultimately futile to attempt to provide an all-encompassing overview of all these interdependencies in the course of only three essays. However, it is the intention of this dissertation to shed light on some of the more important among the countless relationships within the ETF ecosystem and thus to contribute to a better understanding of ETFs in general.

1.1 BACKGROUND, LANDSCAPE AND FUNCTIONING OF ETFS

1.1.1 THE RISE OF PASSIVE INVESTING

The success of ETFs and similar passively managed products would not have been possible without the incorporation of indexing into institutional investment some 40 years ago (Hill, Nadig, & Hougan, 2015). The theoretical groundwork for index investing had been laid even earlier with concepts like the Modern Portfolio Theory (MPT) introduced by Markowitz (1952) and later augmented by Sharpe (1964) and Lintner (1965) among others and the Efficient Market Hypothesis (EMH) formulated by Fama (1970). Markowitz (1952) postulates that in a world where asset returns follow a distribution described by their first two moments of mean and variance, investors should also optimise their portfolio weights based only on two parameters, namely risk and expected return. Expanding this conjecture on portfolio diversification, Sharpe's (1964) and Lintner's (1965) theories on capital asset pricing ultimately imply that in equilibrium and according to their risk preferences all investors will hold a combination of a risk-free asset and a market portfolio, which represents the mean-variance-efficient portfolio on the efficient frontier that is tangential to the capital allocation line. Fama (1970), on the other hand, provides the theoretical backing for the key assumption underlying MPT that assets are correctly priced by the market at all times.

One profound implication of these works is that in the long run and on a risk-adjusted basis, an active investment cannot systematically outperform a passive investment in an optimally diversified market portfolio. Over the years, this theoretical notion has been substantiated by a multitude of empirical studies (e.g. Malkiel, 1995; Gruber, 1996; French, 2008; Fama & French, 2010; Del Guercio & Reuter, 2014). One of the key reasons for this finding is that the cost of active investing is substantially higher than that of passive investing. In his seminal essay "The Arithmetic of Active Management", Sharpe (1991) makes the case that:

> "If 'active' and 'passive' management styles are defined in sensible ways, it must be the case that (1) before costs, the return on the average actively managed dollar will equal the return on the average passively managed dollar and (2) after costs, the return on the average actively managed dollar will be less than the return on the average passively managed dollar. These assertions will hold for any time period. […] To repeat: Properly measured, the average actively managed dollar must underperform the average passively managed dollar, net of costs. Empirical analyses that appear to refute this principle are guilty of improper measurement." (pp. 7-8)

In light of these insights, the first passively managed investment vehicles emerged that did not focus on generating an outperformance over a pre-defined benchmark but rather on mimicking the benchmark-return as closely and cost-efficiently as possible.

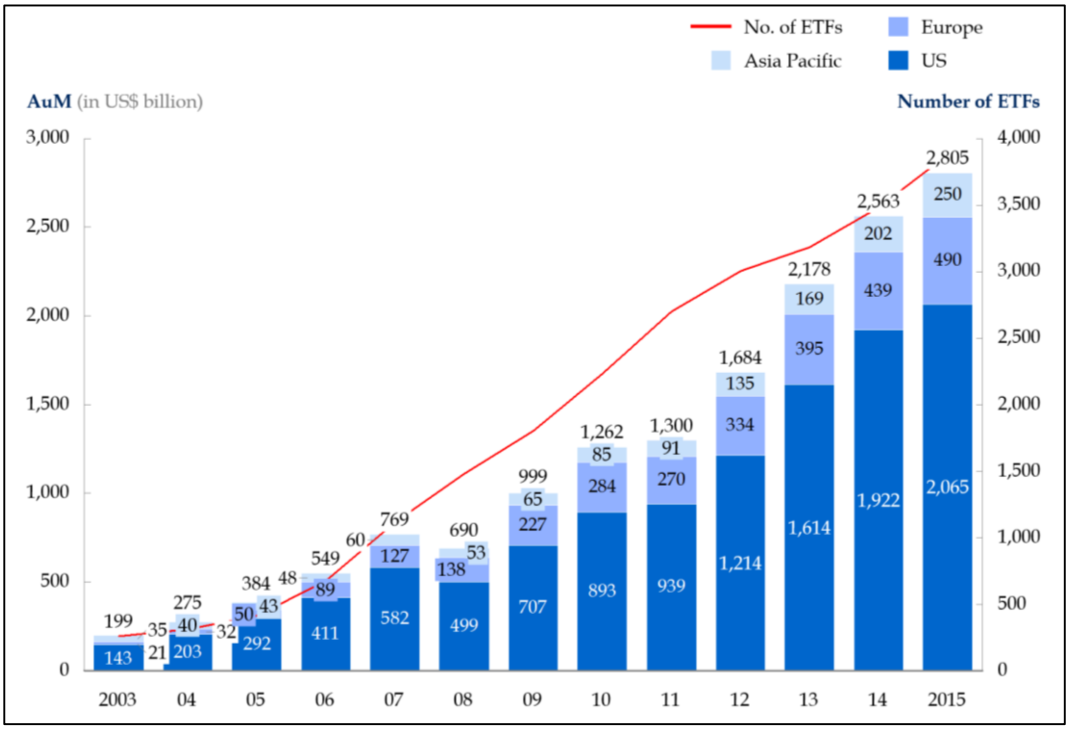
After the launch of the first index mutual fund by John Bogle's Vanguard Group in 1975, passively managed funds quickly gained momentum; by 2014, they accounted for over 20% of total funds AuM in the US and 14% globally (BCG, 2015; ICI, 2015). The growing popularity of index-based investment not only resulted in an increasing share of passively managed vehicles in total AuM, but also had other far-reaching effects on the asset management industry, as it forced asset managers and financial advisers around the globe to improve their precision and value proposition (Hill, Nadig, & Hougan, 2015).

1.1.2 The ETF Landscape – Beginning of The Second Act

Although the 1993 launch of the SPDR S&P 500 Index ETF (SPY) is often considered the ETF industry's date of birth, the very first exchange-traded, index-linked fund had actually been introduced three years earlier, when the Toronto 35 Index Participation Units (TIPs) were listed on the Toronto Stock Exchange in 1990.[3] Thousands of new ETFs followed in the years to come. In 2003, ten years after SPY's launch, 100 ETFs were listed; today there are more than 3,800 products covering all sorts of asset classes, strategies and markets. While the first ETFs largely replicated local equity indices, providers later expanded their product offerings to cover different regions, sectors and investment styles. In 2002, the first fixed-income ETFs were introduced, a product line that now offers access to a wide range of bond indices for different ratings, durations, issuer types, currencies and regions (Daley, Dorencz, & Bargerstock, 2010). Money market instruments, currencies, commodities and exotic products like hedge fund strategies and ETFs on leveraged or inverse indices, followed shortly thereafter. More recently, strategies such as Smart Beta and actively managed ETFs have drawn increased attention from investors.

[3] Whereas TIPs met the fate of many other funds in being merged and consolidated, SPY today remains the largest and oldest ETF in the world and is widely recognised to be the role model for ETF design.

Figure 1.1: Global ETF regional asset growth, 2003–2015

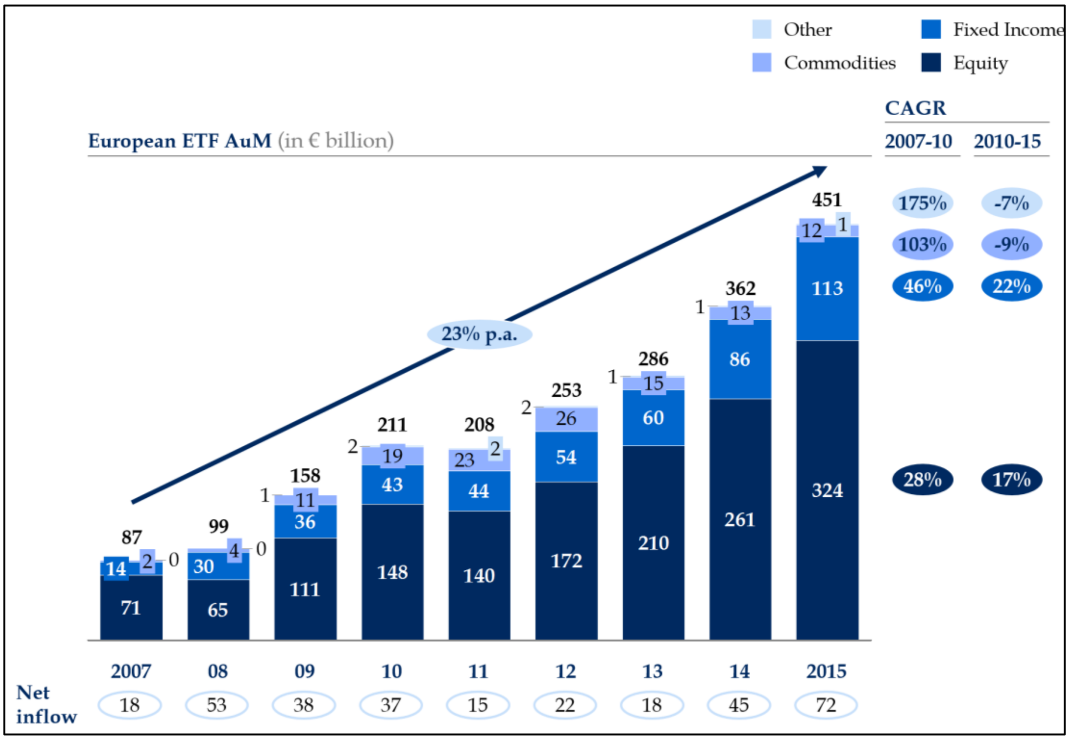


This figure reports global ETF regional asset growth in US$ billion for the period 2003–2015 (LHS). For reasons of visual simplicity, Latin America and Middle East & Africa are omitted in the figure; as of 2014, these two regions had combined AuM of only approximately US$ 11 billion in ETFs. The red line illustrates the number of listed ETF products over time (RHS). (Source: Deutsche Bank Research, 2016)

Figure 1.1 illustrates the global growth, divided by region, of ETFs and their total AuM over the last 13 years. Between 2003 and 2015, ETF assets in the US grew on average by 27% per year, compared to 32% and 19% average annual growth in Europe and Asia Pacific respectively. However, growth has slumped considerably over the last five years, especially in Europe, where it fell to approximately 17%, lagging behind developments in the US (24%) and Asia Pacific (nearly 30%). Despite its global expansion over the years, the market for ETFs today is still clearly dominated by the US, which accounts for almost three quarters of all AuM – US$ 2.07 trillion compared to US$ 490 billion in Europe and US$ 250 billion in Asia Pacific. To put these figures into the broader fund industry context, ETFs still represent only a small fraction of the total fund market; in the US, ETFs comprise some 16% of the total mutual fund industry, while in Europe it is merely just over 3% (Deutsche Bank Research, 2016).

Nevertheless, due to their high liquidity throughout the day, ETFs already account for over 26% and 9% in equity trading on US and European stock exchanges respectively.

Figure 1.2: European ETF asset growth by asset class, 2007–2015



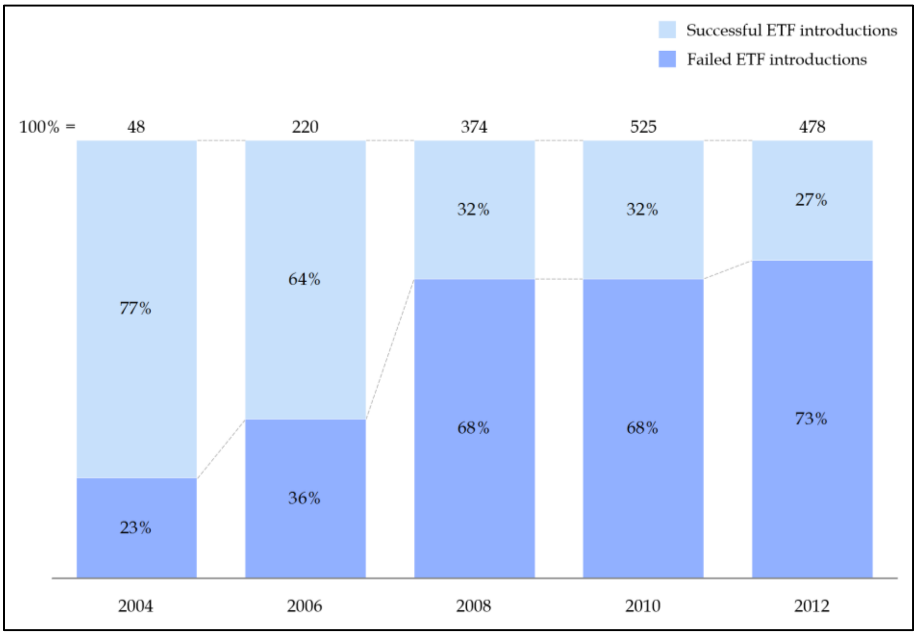
This figure reports European ETF asset growth by asset class in € billion for the period 2007–2015. The compound annual growth rate (CAGR) for each asset class for the periods 2007–2010 and 2010–2015 and annual net cash inflow in € billion are also reported. (Source: Deutsche Bank Research, 2016)

Turning to Europe, the focus region of this dissertation, the absolute growth of assets during the 2007–2015 period was driven largely by equity ETFs, with fixed-income products gaining considerable momentum over the last couple of years amid a global investment trend towards this asset class, especially corporate and high-yield bonds (see Figure 1.2). Notwithstanding the recent relative growth in fixed income products, equity ETFs still make up more than 72% of total AuM and account for approximately two thirds of net cash flows into ETFs from 2013 through 2015. Comparing the inflow of € 72 billion in 2015 to the overall absolute asset increase of € 89 billion, it is clear that the AuM growth is not being driven by bullish market conditions alone but also by fresh money. The rebound in net inflows after 2013 reflects in part the fact that in the aftermath of the global financial crisis, European banks initially focused on deposit account sales in order to fulfil regulatory mandates to improve their respective balance sheets and only gradually returned to market fund products (BCG, 2015). The current levels of net asset inflow can be expected to continue for the near future, given that, according to a recent École des Hautes Études Commerciales du Nord survey among European ETF investors, some 60% of respondents stated their intention to increase ETF investment in the coming years; in two thirds of these cases, the main motive was to substitute actively managed funds (EDHEC, 2015). Moreover, the authors of the report suggest that ETFs are not only benefiting from a general trend towards index investing, but also that a growing share of investors appears to favour ETFs over other indexing instruments such as futures or total return swaps and increasingly roll over their exposure to ETFs that offer equivalent exposure.

While ETFs continue to be a key driver of growth in the investment management business globally, the recent slowdown in asset growth may be heralding a second act, with a more competitive landscape and product proliferation among the strategic challenges facing the entire industry (McKinsey & Company, 2011). Although the number of competitors in the ETF industry has been relatively stable, there are several large banks and asset managers that have yet to decide whether to enter the market by starting their own ETF product lines. Since most of the existing players remain small, substantial industry consolidation remains a real possibility. The rising number of ETF products, coupled with the growing share of unsuccessful product launches, furthermore implies a gradual market saturation, especially with easy gains having already been realised. Figure 1.3 illustrates how difficult it has become for ETF providers globally to attract sufficient assets for new products; whereas more than 77% of all ETFs in 2004 succeeded in collecting US$ 100 million or more in the first 24 months after their launch, only 27% managed to do so in 2012. This global trend also holds true for the German market, where the share of successful ETFs dropped from 89% to 24% between 2004 and 2012 (see Appendix A). Despite growing difficulties in obtaining sufficient funding for their new products and the increasing number of fund shutdowns or mergers, several ETF providers still appear ready to attempt to differentiate themselves from their competitors through broader and more innovative product offerings for a wider range of uses, across different horizons and in varying market conditions (Hill, Nadig, & Hougan, 2015).

In summary, ETFs have seen a remarkable growth in all regions of the world since their arrival in the 1990s. With regard to share of assets, they not only gained on existing passive products such as index mutual funds but also – and in Europe even predominantly so – on actively managed instruments. However, after years of unprecedented development, the first signs of declining growth and increasing market saturation make an intensifying competitive environment and market consolidation more likely than ever before.

Figure 1.3: Success of new ETF introductions by year (Global)



This table reports the success of global new ETF introductions by year. Successful ETF introductions are defined as launches that secure US$ 100 million AuM at any point in their first two years. The total number of product launches per year is also reported at the top of each bar. (Source: Morningstar)

1.1.3 HOW ETFS WORK – THE KEY FUNDAMENTALS

At first glance, the basic design of an ETF appears to be similar to that of an index mutual fund – both are registered as open-end funds and have the same objective. However, they exhibit crucial structural differences, such as how they replicate their respective benchmarks, how their shares are issued and redeemed and how their shares are traded. This section offers a brief overview of how ETFs work and of the fundamental components that make ETFs unique.

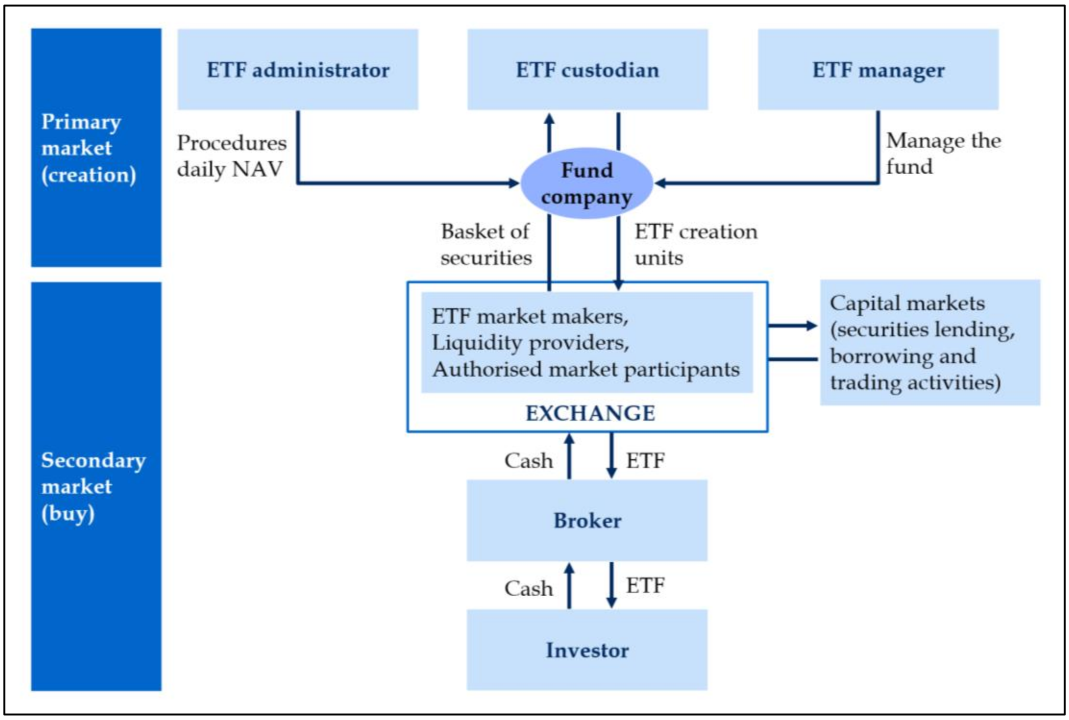
Regardless of the benchmark to be mimicked, the management of an ETF first has to decide whether to replicate the respective risk-return profile physically or synthetically. *Physical replication* entails that the fund invest directly in the securities in the underlying reference basket to replicate its returns. If the underlying securities are liquid enough, the fund manager can buy all index constituents according to their respective weights, which is known as *full replication*. However, if the constituent securities in the basket are too illiquid, if their overall number is too high or if the index is frequently adjusted, full replication might be too costly or otherwise impractical. In this case, the fund manager might decide to invest only in a selection of constituent securities, typically those with the highest correlation to the overall index, using techniques like sampling or optimisation. For *synthetic replication*, the fund enters a swap deal with a counterparty, often the ETF provider's parent bank, to mimic the benchmark's performance. In the *unfunded* swap model, the ETF invests the cash from investors in a basket of substitute securities that do not necessarily have to be part of the reference index and then delivers the return of these securities to the swap counterparty in exchange for the reference index's performance (Johnson, Bioy, & Rose, 2012). Under the structure of a *funded* swap model, the ETF does not hold a substitute basket but instead delivers the cash to a single swap counterparty, which in turn commits itself to deliver the index performance and to post collateral with a third-party custodian. While synthetic replication bears some counterparty-risk,[4] it allows

---

[4] For an examination of the operational frameworks of exchange-traded funds and potential channels through which risks from synthetic ETFs to financial stability can materialise, see Ramaswamy (2011).

the mimicking of even highly illiquid or complex indices.[5] After a very successful start in Europe, synthetic ETFs' overall share in net inflows and in new product launches has been in constant decline for several years. Although they still account for approximately one half of all listed ETFs in Europe, they make up only 26% of all ETF AuM.

Figure 1.4: Overview of the primary and secondary markets for ETFs



The chart lays out the process of creating ETF shares in the primary market and buying it in the secondary market, indicating participants involved in this transaction flow. The redemption process as well as the sell transaction are not pictured. Redemption is the reverse of creation: Here, the market maker swaps ETF creation units with the ETF custodian for the underlying basket of securities. In a sale of ETF shares on the secondary market, ETFs are exchanged for cash. (Source: EDHEC, 2015)

ETFs can be traded on primary and secondary markets (see Figure 1.4). On the secondary market, investors can trade ETF shares just like listed stocks through an exchange or over the counter (OTC). As these transactions have no direct effect on the ETF's portfolio composition or its number of shares outstanding, they generally go

---

5 See Section 2.2.2 for a discussion of research on the relative tracking ability of synthetic ETFs compared to physically replicating ETFs.

unnoticed by the fund management. However, transactions on the primary market always entail the creation of new shares or the redemption of existing shares and hence do have a direct impact on the ETF and its underlying portfolio.

The mechanism of creating and redeeming shares is a central, if not the most essential, pillar of ETF functioning (Abner, 2010). For reasons of practicality, ETF shares can only be created or redeemed in large, pre-specified blocks of shares, so-called creation units, which range in size from one thousand to several hundreds of thousands of shares. The process is executed by authorised participants (APs), usually trading desks at investment banks, brokers or independent trading houses, who are mandated by the ETF provider to manage the creation/redemption process of one specific or of several ETFs. Balancing supply and demand for the ETF on the secondary market by creating or redeeming shares on the primary market, APs act at the nexus between these two spheres.

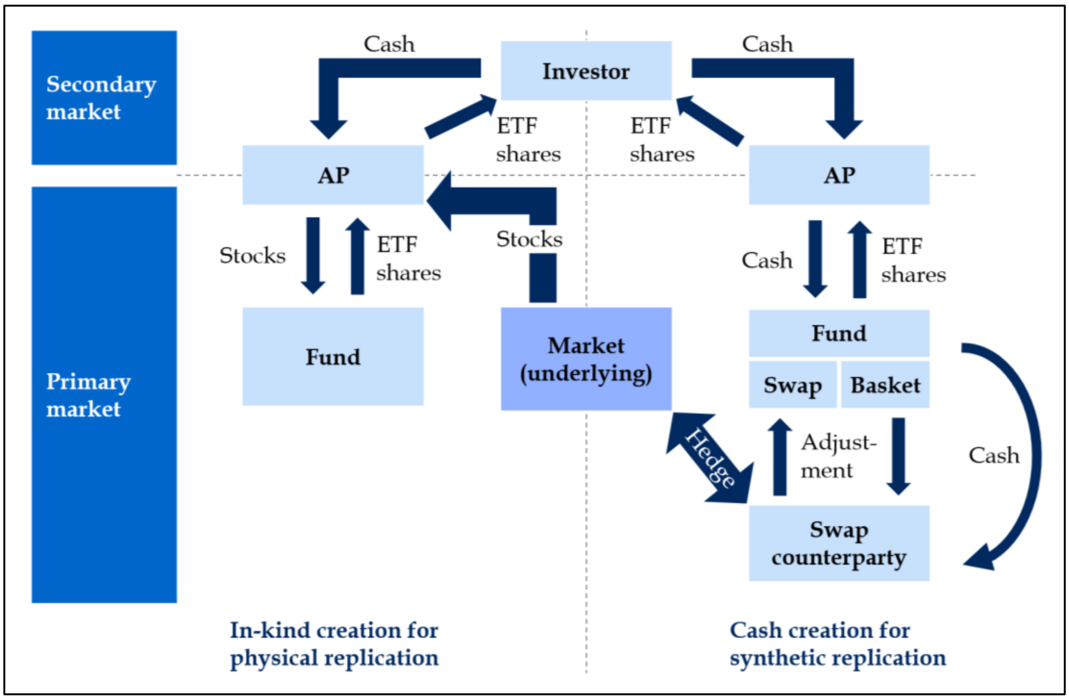
Share creation[6] can be either *cash* or *in-kind*, meaning that in return for ETF shares, the AP has to deliver either cash or a pre-determined selection of securities (known as the creation basket) to the ETF's management (see Figure 1.5). A cash creation is similar to the mechanism used by mutual funds in that it leaves the task of investing the cash into the underlying securities to the fund. APs deliver cash to the ETF according to the creation units they want to create. The fund then either invests the money directly in the underlying reference basket or, in the case of synthetically replicating ETFs, in a basket of collaterals (directly in the case of an unfunded swap or via the single swap-counterparty in the case of a funded swap), subsequently adjusting the swap. While cash creation is the method of choice for synthetic ETFs, it is less common for physically replicating ETFs; their shares are predominantly created through the in-kind mechanism. The in-kind creation process starts each business day with the ETF issuer publishing a file detailing the exact composition of the creation basket that the fund expects in return for one creation unit of ETF shares. Depending on the number

---

6 For reasons of simplicity only the creation process is described in the following sections. The redemption of ETF shares is the reverse of the share creation process.

of units that an AP wants to create, the AP first buys the securities specified in the creation basket file on the secondary market and then delivers them to the ETF. In return, the AP receives the corresponding number of creation units, which are then either kept on the books or sold on the secondary market. The actual process of exchanging the creation basket against shares takes place at the end of the day. Based on the AP's overall aggregated net exposure from market making activities for the ETF throughout the trading day, one net creation or one net redemption and the corresponding delivery of underlying securities are executed.

Figure 1.5: The cash and in-kind creation process for ETFs



(Source: Kim, 2014)

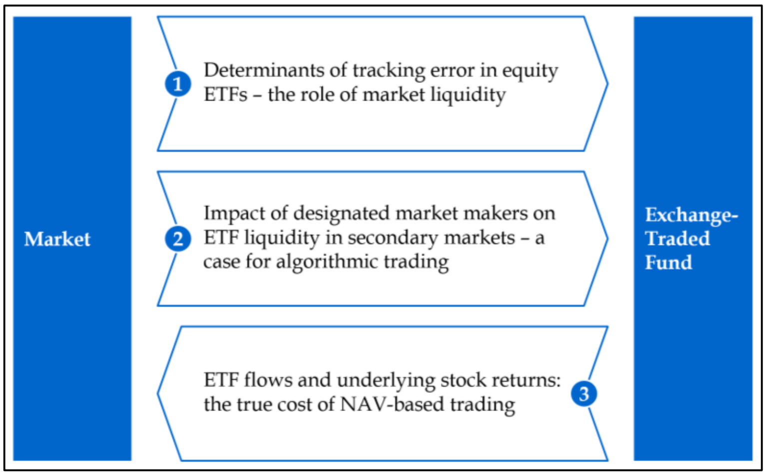
While the net creation (or redemption) ultimately takes place once a day, market makers can still quote bid-ask spreads for the ETF throughout the trading day. Since the details of the creation basket required for delivery or redemption at the end of the day are publicly known, an indicative net asset value (iNAV) of the ETF can be determined at any point during the day (Hill, Nadig, & Hougan, 2015). This circumstance allows the APs and other investors to exploit arbitrage opportunities by simultaneously trading ETF shares and their underlying creation baskets. If an ETF trades at a premium to its NAV, for instance, an AP could sell the ETF short in the secondary market and simultaneously go long in the underlying securities of the creation basket. To cover the short position, the AP could then instantly create new ETF shares and use the long position in the underlying securities for the creation basket to be delivered. Over time, the arbitrage trading will bring prices back to equilibrium. The process of creation/redemption and the AP's role as a link between primary and secondary markets are essential in ensuring that market-determined ETF prices are always close to the value of the underlying portfolio; indeed, they are much closer than intraday prices of exchange-traded mutual funds (Engle & Sarkar, 2006).

1.2 RESEARCH QUESTIONS AND CONTRIBUTION

As ETFs have entered the centre stage of asset management, they have drawn considerable attention not only from investment professionals but also from academic scholars. Charupat and Miu (2013) offer a comprehensive overview of the current state of research on ETFs, identifying three main strands in the literature: (i) investigations of gaps between an ETF's market price and its NAV and of how well the fund's creation/redemption process works in arbitraging away these differences; (ii) analyses of ETF performance, often by conducting comparative analyses of the tracking ability of a set of ETFs and competing products; (iii) examinations of the effects of ETF trading on related securities such as underlying stocks or derivatives on the reference index.

This thesis contributes to the last two strands of literature, in which previous attempts have left research on some key areas still in the incipient stage. Using extensive data on equity ETFs listed in Germany and their underlying stocks, each of the three essays presented in this thesis separately provides first-hand empirical evidence on specific key ETF-market interdependencies. Two essays elaborate on how external market factors affect certain aspects of ETF quality, thus contributing to research strand (ii) and the literature focussing on how successful ETFs are in managing external influences to provide their advertised value proposition. The first of the two essays sheds new light on the relationship between stock market liquidity and ETF tracking ability (Chapter 2), while the second one analyses the impact of hiring additional market makers on ETF liquidity and determines whether market makers that rely predominantly on algorithmic or high-frequency trading have a structural advantage in providing liquidity (Chapter 3). While these two essays study external market forces' impact on ETF characteristics, the third essay takes a somewhat reverse inside-out perspective by determining how creation- or redemption-induced equity trading affects underlying stock returns (see Figure 1.6). In doing so, it contributes to research strand (iii) and the literature on the potential effects of ETF trading on related securities.

Figure 1.6: Overview of ETF-market interactions analysed in the thesis



This chart lays out an overview of the ETF-market interactions analysed in the three essays presented in this thesis and the main direction of the observed effects.

In addition to the collective theme of ETF-market interrelations, all three papers share the fact that their datasets arise solely from equity ETFs listed in Germany. Despite the recent growth of other sectors, equity ETFs remain by far the largest and most important product line for the industry, accounting for more than three quarters of overall ETF AuM in Europe. Although some studies suggest considerable variation in ETF characteristics across different regions (e.g. Shin & Soydemir, 2010; Svetina, 2010; Wong & Shum, 2010), academic research on developed markets outside the US is scarce at best. Despite the importance of Germany's ETF market – with a market share of 25% in 2014, second only to the London Stock Exchange in Europe and the fourth-largest ETF marketplace in the world (Deutsche Bank Research, 2014) – it has been grossly understudied thus far, and further research on its microstructure is sure to be worthwhile. Furthermore, a clear focus on one underlying asset class and one regional market ensures that neither structural differences between underlying baskets nor regulatory differences between exchanges, such as those affecting settlement cycles or designated market maker (DMM) duties, bias the results presented in the essays or the conclusions drawn from them.

1.2.1 Determinants of Tracking Error in Equity ETFs – The Role of Market Liquidity

A fundamental quality for any ETF is its ability to track the performance of its benchmark index as closely as possible. Up to now, studies have mostly clustered around three factors that significantly drive the tracking error in ETFs, namely the total expense ratio (e.g. Elton et al., 2002; Agapova, 2011; Blitz et al., 2012), changes in index composition and the index replication strategy (e.g. Gastineau, 2002; Frino et al., 2004; Aber et al., 2009) and dividend payments (e.g. Elton et al., 2002; Frino et al., 2004; Blitz & Huij, 2012).

The first essay identifies stock market liquidity as an additional factor that affects the tracking error of an equity ETF. The impact of stock market liquidity on an ETF's tracking ability has largely been ignored in the literature, while the few studies that do acknowledge its potential effect rely predominantly on ETF bid-ask spreads for capturing market liquidity (e.g. Milonas & Rompotis, 2006; Delcoure & Zhong, 2007; Shin & Soydemir, 2010). However, using this proxy makes it impossible to grasp fully all dimensions of liquidity, especially market depth, or to comprehend the liquidity effect of individual stocks on an ETF's tracking ability (Krogmann, 2011; Hendershott & Riordan, 2013). In contrast to these studies, Deutsche Börse's volume-weighted spread XETRA Liquidity Measure (XLM) is used throughout this essay. XLM measures the order-size-dependent liquidity costs of a round-trip transaction for individual stocks, taking the entire depth of the limit order book into account (cf. Stange & Kaserer, 2011; Rösch & Kaserer, 2013). Using this measure for each stock in the underlying portfolios of all observed ETFs should allow for a more detailed view of the liquidity costs of individual stocks in an ETF's underlying basket and their effects on tracking error.

The essay provides empirical evidence that the tracking ability of a physically replicating equity ETF is significantly affected by the liquidity of individual stocks in its underlying portfolio, both directly and in interaction with portfolio adjustments. Even after separately controlling for cash holdings, portfolio adjustments and creation/redemption, market liquidity of stocks still has a strongly significant and independent effect on ETF tracking error. In these cases, the observed independent liquidity effect appears to represent the liquidity cost borne by the ETF for its attempts to optimise the weights of its underlying portfolio.

The analysis also shows that creation/redemption activity affects tracking ability, contrary to Gallagher and Segara's (2006) and Gastineau's (2004) notion that ETFs are immune from creation/redemption-related transaction cost. Two explanations are proposed for this effect. First, imperfect replication of index weights in the ETF portfolio during a creation or redemption might result in minor deviations from the index composition and hence in tracking error with regard to index performance. Second, inadequate accounting treatment of the daily attribution of income and fees to or from the fund's NAV, such as income from securities lending, could have a significant effect on tracking error. Together with the findings on portfolio adjustments, the results for the effect of creation and redemption on tracking error furthermore imply that any changes in the composition of the underlying basket affect the ETF's tracking ability, regardless of whether they are caused by a creation or redemption or other portfolio adjustments.

Moreover, the results show that aside from the liquidity cost of stocks, the creation and redemption of shares and portfolio adjustments, management fees, dividend yield, cash distributions to ETF investors and cash holdings also have a significant and sometimes economically substantial effect on ETF tracking error.

1.2.2 IMPACT OF DESIGNATED MARKET MAKERS ON ETF LIQUIDITY

As with any other security, liquidity is a central aspect of market quality for ETFs, and an increasing number of ETF exchanges rely on DMMs to ensure liquidity beyond endogenous levels (cf. Anand et al., 2009). Over the past couple of years, a general trend has emerged among ETF issuers to outsource market making activities completely or partly to external providers.

Chapter 3 deals with this trend and contributes to the understanding of the relationship between external liquidity provision and the liquidity of equity ETFs. The essay first analyses whether hiring an additional DMM has a measurable effect on the liquidity cost of ETFs, before determining whether external DMMs that predominantly rely on algorithmic or high-frequency trading are systematically better at providing liquidity than non-HFT types of market makers.

The notion that DMMs do improve liquidity has been corroborated by various studies (e.g. Nimalendran & Petrella, 2003; Venkataraman & Waisburd, 2007; Anand et al., 2009; Menkveld & Wang, 2013). The same is true for the perceived effect of algorithmic and high-frequency trading on market liquidity, with several studies showing a significant positive correlation between the advent of HFT and liquidity (e.g. Hendershott et al., 2011; Riordan & Storkenmaier, 2012; Hasbrouck & Saar, 2013; Hendershott & Riordan, 2013). However, research on these two phenomena with a focus on ETFs is scarce at best.

The essay presented in Chapter 3 is among the first studies with a clear focus on ETFs to elaborate on the effects of additional DMMs and their operational design on market liquidity. Compared to most other academic works that examine the impact of market making on liquidity at the level of individual stocks, the essay's research design has the advantage of a reduced risk of endogeneity; since the impact is observed at the level of aggregated portfolios, potential uncontrolled liquidity effects of individual stocks should be mitigated. As noted above, most existing studies on DMMs' effect on liquidity commonly rely on ETF bid-ask spreads to approximate market liquidity. In doing so, they fail to address market depth as a decisive factor in liquidity. Therefore, Deutsche Börse's XLM data is again applied to capture all relevant aspects of market liquidity.

The results corroborate the view that contracting an additional market maker immediately and substantially reduces liquidity costs, regardless of whether the DMM applies HFT techniques. Hence, it appears that, contrary to the prevailing view of underlying basket liquidity being the key if not the sole driver of ETF liquidity (e.g. Kittsley & Edrosolan, 2008; Agrrawal & Clark, 2009; Roncalli & Zheng, 2014), there are other factors that determine an ETF's liquidity level. The fact that one additional market maker always has a significant effect on an ETF's liquidity cost structure suggests that the DMMs already operating prior to the new hire do not reduce liquidity cost to the lowest possible levels. One reason postulated in the essay is that by adding another market maker to the pool, the ETF provider increases competition between individual DMMs, which leads to decreased spreads and increased liquidity. The results presented in the essay do not support Anand et al.'s (2009) notion of "liquidity [begetting] liquidity" (p. 1447) through increased AuM and resulting growth in trading volume. Therefore, it appears that the liquidity cost reduction is indeed generated primarily by the previously postulated inter-DMM competition effect. The finding of unchanged AuM one year into treatment also suggests that investors do not immediately honour this higher market quality with larger investments.

According to the models tested in Chapter 3, the estimated annual liquidity cost reductions per ETF from adding one non-HFT or HFT market maker are € 70,000 and € 238,000 respectively. The difference between non-HFT and HFT DMMs, both in terms of liquidity cost in XLM basis points and in terms of absolute cash figures, is highly significant and economically substantial and serves as evidence that HFT market makers are systematically better at providing liquidity. Yet, given that the outperformance only becomes significant two or three months after hiring the additional DMM, it appears that the HFT market maker needs some time to exploit its full potential.

1.2.3 Effect of ETF Flow on Underlying Stock Returns

The effect of fund flows on stock prices has been confirmed by several studies (e.g. Edelen & Warner, 2001; Yu, 2005; Coval & Stafford, 2007; Jotikasthira et al., 2012). However, despite the structural differences between ETFs and mutual funds, especially the in-kind creation/redemption process for ETF shares, the relationship between ETF flows and their underlying stocks has received far less attention, with the exceptions of studies by Kalaycıoğlu (2004) and Staer (2014).

By shedding new light on the relationship between ETF flow-related trading volume and stock returns, the third essay extends the research on ETFs' impacts on related securities. Previous studies usually describe the effect of ETFs on their respective underlying stocks at an aggregated index or market level. However, more granularity is necessary in order to identify the different levels of influence that ETFs have, depending on the size and liquidity of a stock. This essay contributes to the existing research gap by using intraday tick data of the entire stock universe of the DAX index family, which allows for an examination of ETF flow effects on returns at the most granular level possible, individual stocks.

The key variable in this approach is the abnormal stock return in the closing auction in the German stock market. By using this controlled environment for the study, it is less likely that general market movements or breaking news drive abnormal returns in the observed sample. Furthermore, the abnormal return is calculated against the MSCI EMU ex Germany index, which encompasses the equity market of the entire European Monetary Union with the exception of Germany. This immunises the market return and the applied market model from the very effect of flow-related transactions in the sample that the essay is examining. What is more, by concentrating on the last minutes of trading, the study contributes to research on the price effects of trading around the close. Contrary to most papers that rely on prosecuted cases of manipulation to determine the effect on stock prices (e.g. Comerton-Forde & Putniņš, 2011), the model applied in this essay takes all forms of last-minute trading into account, thereby allowing for a broader and less biased view on the effects of trading around the close.

The results reveal that ETF flow-related stock transactions significantly affect stock prices and show that creations/redemptions of ETFs that replicate indices in the DAX index family have a highly significant and economically viable effect on abnormal returns of underlying stocks in the closing auction. The effect is particularly pronounced in small stocks and on trading days that are generally bullish. One explanation for the persistence of this effect is that APs might be able to exploit such pricing inefficiencies; given the substantial additional earnings, active price manipulation might be an attractive option. Therefore, the key conclusion is that dealing ETF shares on the primary market, for example through NAV-based orders, might entail hidden costs that have not been recognised until now in the literature and perhaps not even by investment professionals.

1.3 STRUCTURE

The remainder of this dissertation is structured as follows. Chapter 2 is dedicated to the first essay on ETF performance and the role of market liquidity in determining ETF tracking ability. Chapter 3 presents the findings on the role of HFT and non-HFT market makers in liquidity provision for ETFs in secondary markets. Chapter 4 focuses on the impact of ETF-induced market transactions on underlying stock returns. Chapter 5 offers some concluding remarks and an outline of potential avenues for future research.

2. Determinants of Tracking Error in German ETFs – The Role of Market Liquidity

Abstract

In this paper, we attempt to identify the determining factors of daily tracking error in physically replicating ETFs in the German DAX index universe, with a special focus on the liquidity of individual underlying stocks. It has been argued that market liquidity should not affect the tracking error of an ETF because creation or redemption of ETF shares is usually performed in-kind through authorised participants. We find that the daily tracking error significantly depends on the liquidity of underlying stocks. Moreover, we show that the liquidity effect cannot be explained by the creation/redemption of ETF shares alone. We argue that this effect might be attributable to imperfect replication of index weights: Either the in-kind basket delivered in the course of creation/redemption does not perfectly match the benchmark index weights, or the internal rebalancing of weights results in liquidity cost. We also show that portfolio adjustments, cash holdings, dividend yield and cash distributions from an ETF to its investors affect the tracking error.

***Keywords*:** ETF, Exchange-Traded Fund, Tracking Error, Liquidity Cost, Market Liquidity, XETRA Liquidity Measure, Creation and Redemption

***JEL Classification*:** G12, G14, G15, G23

***Authors*:** Christoph Kaserer, Friedrich Osterhoff

***First Author*:** Friedrich Osterhoff

***Current Status*:** Accepted; forthcoming in *Managerial Finance*

2.1 INTRODUCTION

Over the past few years, exchange-traded funds (ETFs) have experienced a remarkable development from being a mere niche product to becoming "one of the most successful innovations in the history of investment" (Charupat & Miu, 2013, p. 427). As a result, they have drawn considerable attention from both researchers and investors. For any ETF trying to replicate the performance and risk of an underlying benchmark a decisive if not defining quality is its ability to track its corresponding benchmark as closely as possible. Although a growing body of literature has confirmed the significant impact of an ever-increasing number of factors on ETF tracking ability, research on some potential key determinants still appears to be in its inception, especially for ETFs in developed markets outside the US, such as that of Germany.

Our intention in this paper is to shed new light on the relationship between stock market liquidity and the performance of the ETF industry. More specifically, we will investigate the extent to which stock market liquidity affects ETFs' tracking error. It should be noted that so far, this relationship has been more or less ignored in the literature. In fact, pertinent studies identify three factors which significantly drive an ETF's tracking error, as follows: (i) the total expense ratio (e.g. Elton et al., 2002; Agapova, 2011; Blitz et al., 2012), (ii) changes in index composition and the index replication strategy (e.g. Gastineau, 2002; Frino et al., 2004; Aber et al., 2009)), and (iii) dividend payments (e.g. Elton et al., 2002; Frino et al., 2004; Blitz & Huij, 2012).

The impact of stock market liquidity has mostly been ignored based on the presumption that the creation/redemption of ETF shares is usually performed in-kind through authorised participants (APs), and that this should shelter the ETF from any market frictions like transaction costs. A very few studies acknowledge that market liquidity might nevertheless be an issue. However, these papers predominantly rely on ETF bid-ask spreads to capture market liquidity (e.g. Milonas & Rompotis, 2006; Delcoure & Zhong, 2007; Shin & Soydemir, 2010); this proxy is somewhat flawed given that it measures liquidity merely at the aggregate fund level and that it does not take market depth into account (cf. Krogmann, 2011; Hendershott & Riordan, 2013).

In this study, we aim to contribute to a better understanding of the impact of market liquidity on the performance of the ETF industry, or more specifically, on the tracking error of ETFs. To the best of our knowledge, this is the first paper that uses a specific liquidity measure for each single stock underlying an ETF. This measure is Deutsche Börse's volume-weighted spread XETRA Liquidity Measure (XLM). It measures the order-size-dependent liquidity costs of a round-trip for individual stocks, taking the entire depth of the limit order book into account (cf. Stange & Kaserer, 2011; Rösch & Kaserer, 2013). Applying XLM should allow for a more elaborate view of the liquidity costs of individual stocks in the underlying portfolio of an ETF and its effects on tracking error.

Our findings extend the literature by corroborating the view that the liquidity of individual stocks in the underlying portfolio of an ETF has a considerable impact on its tracking error. This finding emerges even though the ETFs under investigation in this paper predominantly use in-kind redemption or creation of shares through APs. In fact, even after controlling for creation/redemption, the liquidity effect remains basically unchanged. Therefore, the relationship between market liquidity and ETF tracking ability seems to be rather intriguing. We suggest some explanations in this paper, although we are not able to isolate any specific channel due to data limitations. Moreover, we are also able to show that besides the liquidity cost of stocks and the process of creation and redemption of ETF shares, portfolio adjustments, management fees, dividend yield, cash distributions to ETF-investors and cash holdings also have a significant and sometimes substantial effect on an ETF's tracking ability. Finally, by using an orthogonalisation technique, we show that the effects of total expense ratio, basket liquidity and distributions on tracking error are highly non-linear.

The remainder of this paper is structured as follows: In Section 2.2, we give a brief overview of the relevant literature with a focus on the current state of research on potential determinants of tracking error. Section 2.3 comprises the empirical part of the study. Here, we describe the data and methodology (2.3.1) and subsequently present (2.3.2) and critically discuss (2.3.3) our findings. Finally, Section 2.4 provides concluding remarks, especially with regard to potential future fields of research.

2.2 LITERATURE REVIEW

Tracking error can be broadly defined as the deviation of an ETF's price or net asset value (NAV) return from its corresponding benchmark index return. Price deviations from the ETF NAV in the form of premiums or discounts are quite common, yet given the arbitrage opportunities of daily creation and redemption of ETF shares, they can be expected to remain within rather tight bounds. In contrast, tracking error – that is, NAV return deviating from the return of the corresponding underlying index – can accumulate over time and hence significantly affect the long-term performance of ETFs (Charupat & Miu, 2013).

2.2.1 TRACKING ABILITY OF ETFS IN DIFFERENT MARKETS

Due to its maturity in terms of assets under management (AuM), overall trading volume and product range, the US ETF-market has undoubtedly drawn the most attention from researchers over the last few years. This is also true in terms of empirical work on tracking ability; indeed, various studies confirm the existence of tracking error in US-traded ETFs. Elton et al. (2002), for instance, investigate the SPY-ETF – the very first ETF in the US – for the period between the fund's inception in 1993 and 1998 and find evidence for significant tracking error with the corresponding S&P 500 index, with an average of 28 basis points per annum. Blume and Edelen's (2003) study on S&P 500-tracking index mutual funds finds tracking errors to amount only to a few basis points per year, while Elton et al.'s (2002) study highlights potential differences between conventional mutual funds and their respective ETF counterparts. Aber et al. (2009) analyse four iShares ETFs that track broad US equity indices and compare them with their corresponding conventional index mutual fund counterparts in terms of price volatility and tracking ability. They show that both fund types have approximately the same degree of co-movement with their respective benchmarks in their sample but that they differ slightly in their tracking ability. On average, conventional index funds beat their corresponding ETF counterparts in terms of tracking error. This result is inconsistent with Agapova's (2011) findings for US ETFs and index mutual funds; she too observes some tracking error in the ETFs for the period from 2000 to 2004, yet, at least on a gross-of-fees basis, they generally track their underlying benchmarks more closely than their conventional fund counterparts. On a net-of-fees basis, both assets turn out to have zero tracking error (Agapova, 2011).

As to European-traded ETFs, the existence of tracking error has also been confirmed by various studies. Milonas and Rompotis (2006) study the Swiss ETF market and find evidence for substantial tracking error with an approximate average of 1.02% per year, attributing this considerable magnitude to the lack of full-replication ETFs in Switzerland. Blitz et al. (2012) analyse the relative performance of 40 passively managed ETFs and mutual index funds listed in Europe and covering all major global stock markets. They find index funds and ETFs to underperform their respective benchmarks by 50 to 150 basis points per annum. Elia (2012) provides further evidence for tracking error in European-traded funds with his investigation into the index tracking ability of 48 European ETFs, covering 20 different benchmark indices for the period between September 2007 and August 2011. Moreover, he shows that the mere occurrence of tracking error is independent of the replication method used by a given fund. This insight is supported by the findings of Meinhardt et al. (2012), who analyse the daily returns of 326 synthetic and 95 fully replicating euro-denominated ETFs listed on the Frankfurt Stock Exchange for the period from January 2010 to August 2011; they identify tracking errors in all types of ETF, regardless of the method of replication chosen. However, Elia's (2012) and Meinhardt et al.'s (2012) findings do not necessarily challenge Milonas and Rompotis' (2006) suggestion that the magnitude of tracking error in Swiss ETFs can still be attributed to the method of replication.

There are a few other empirical studies on the German ETF market besides Meinhardt et al. (2012). Rompotis (2012), for example, investigates the performance and trading characteristics of 43 ETFs traded on XETRA during the period from 2003 to 2005 and finds significant evidence that average tracking error amounts to 0.54% per year, indicating an insufficient replication of benchmark portfolios in German-listed ETFs. Kundisch and Klein (2009) observe the daily returns and tracking ability of several DAX certificates and one DAX ETF for the period from 2001 to 2006, finding the ETF exhibits tracking error. However, it vanishes almost completely after 2004, a development that the authors attribute to changed trading times at XETRA from 2003 onwards. Finally, Fischer et al. (2013) show that both ETFs and index mutual funds that try to replicate the DAX-index do not perfectly match the return profile of their benchmark, with index mutual funds exhibiting higher tracking error on average.

Over the past years, there has also been a growing literature on ETFs' tracking ability in markets beyond Europe and the US. Gallagher and Segara (2006) study the performance and trading characteristics of ETFs in Australia and find evidence for tracking error, but it is relatively small and low in frequency. Observing ETFs listed at the Hong Kong stock exchange, Chu (2011) obtains results that suggest that tracking errors in Hong Kong are comparatively higher than those documented in the US and Australia. These results are reconfirmed in a subsequent, extended study (Chu, 2013). Another analysis for emerging markets is conducted by Lin and Chou (2006), who investigate daily data from June 2003 to March 2005 for Taiwan's first ETF, the Taiwan Top 50 Tracker (TTT), and confirm the existence of tracking error.

Given all this evidence, tracking error is a prevalent if not universal feature in ETFs, regardless of the marketplace. However, having presented ample evidence for ETFs' being unable to track their respective benchmarks perfectly, one must strive to determine the factors that drive tracking errors.

2.2.2 DETERMINANTS OF TRACKING ERROR IN ETFS

The literature describes a wide array of factors that have a measurable effect on the tracking ability of ETFs. Due to the number of factors, the following section aims to provide some structure by clustering the relevant identified determinants along broader categories. However, given the interconnectivity between some of these parameters, it is sometimes difficult to draw definitive lines.

One widely recognised factor affecting tracking error is management fees: All other things being equal, the higher the expense ratio of a fund, the more an ETF can be expected to underperform its underlying index and hence, the larger the tracking error should be (Charupat & Miu, 2013). This view is supported by the vast majority of research: Elton et al. (2002), Lin and Chou (2006), Rompotis (2006; 2011), Agapova (2011), Elia (2012), Blitz et al. (2012) and Meinhardt et al. (2012), among others, show that an ETF's expense ratio is key to explaining its tracking error. In contrast to this majority, Rompotis (2012) cannot verify the relationship between expense ratio and tracking error to be statistically significant for his sample of German ETFs. Moreover, while Chu (2011) finds the magnitude of tracking error for ETFs listed on the Hong Kong Stock Exchange to be positively related to the expense ratio of a fund (and negatively related to fund size), he observes a negative relationship between expense ratio and tracking error in a later study of 21 ETFs traded in Hong Kong between 2009 and 2011 (Chu, 2013). He explains this rather unintuitive outcome by noting that his analysis does not differentiate between fully replicating ETFs and synthetic ETFs, and hence it does not account for the potential impact of replication strategy on tracking ability via expenses and transaction cost. Although this might be the reason for variation in absolute magnitude in tracking error, it does not sufficiently explain the negativity of the relation.

There is also evidence for a regional impact on tracking error, most probably due to variation of economic development and global financial integration across markets. In their study of 26 ETFs of major US, European and Asian equity indices for the period from 2004 to 2007, Shin and Soydemir (2010) come to the conclusion that US ETFs exhibit the lowest levels of tracking error. This is also consistent with the findings of Meinhardt et al. (2012) and Svetina (2010), who corroborate relatively lower tracking error magnitude for the US. Wong and Shum (2010), on the other hand, find for all investigated markets that tracking errors tend on average to be always positive, with the highest means and standard deviations found in the United States.

Another factor that appears to have a measurable effect on tracking error is cross-country trading, in which ETFs track non-domestic indices. Johnson (2009), for example, investigates the determinants of tracking errors between foreign ETFs trading on a US exchange and their respective foreign country indices, concluding that variables such as foreign index positive returns relative to the US-index and simultaneous trading between foreign and US markets were significant explanatory variables in the correlation coefficients between ETFs and their underlying home indices. Svetina (2010) also analyses US markets and finds tracking errors to be more pronounced in ETFs that invest in multinational indices compared to those investing only domestically. Analysing a sample of international iShares, Rompotis (2011) finds evidence that ETFs which invest in international capital markets have higher tracking errors than domestically invested ETFs. He argues that "[this] trend might be due to higher expenses charged by international ETFs, the time difference between the trading hours of USA and local markets out of the USA, and the greater risk to which international ETFs are usually exposed" (Rompotis, 2011, p. 34). Blitz and Huij (2012) examine the monthly performance of a sample of European- and US-listed ETFs that provide exposure to broad emerging markets equity indices from inception to December 2010, finding that ETF tracking errors are substantially higher than previously reported levels for developed markets. The ones with statistical replication techniques are especially prone to high tracking errors, especially during periods of high cross-sectional dispersion in stock returns. Shin and Soydemir's (2010) findings for MSCI country ETFs suggest that exchange rates are the only statistically significant determinant of tracking error. Volatility of daily market price is significant only for one type of tracking error definition and after controlling for country level effects it is found to be no longer statistically significant in any case (Shin & Soydemir, 2010).

Another factor that has a measurable impact on tracking error is the index replication strategy chosen. For Australian index mutual funds, Frino and Gallagher (2002) show that besides fund cash flows, market volatility and transaction costs, index replication strategy also significantly affects tracking error. This is further supported by Frino et al.'s (2004) findings, which also suggest that tracking errors in fully replicating and large funds are smaller. The implication that synthetic replication is less effective in tracking a benchmark is striking, since one of the key arguments used by ETF providers in favour of synthetic replication is its supposed effectiveness in tracking the underlying index. However, several studies confirm Frino et al.'s (2004) findings for ETFs. Rompotis (2012), for example, shows that the relatively pronounced tracking error found in his sample of German ETFs is due partly to the fact that these funds do not adopt full replication techniques. Milonas and Rompotis (2006) come to similar conclusions regarding the Swiss ETF-market. Chu's two studies (2011; 2013), suggest that synthetic ETFs have higher tracking errors on average than physically replicating ETFs. He argues that ETF managers may face difficulties in finding derivatives that are exact matches of the securities included in their respective benchmark indices, which would lead to tracking errors (Chu, 2013). Meinhardt et al. (2012) examine euro-denominated ETFs, finding that fixed-income, full replication ETFs have smaller tracking errors than any equity ETF, regardless of replication method. They provide further evidence that, contrary to conventional wisdom, synthetic equity ETFs do not have smaller tracking errors than their full replication counterparts. However, in the case of fixed-income products, synthetic ETFs have a better ability to track their respective benchmarks. Overall, they suggest that tracking error is generally determined by risk, volume, total expense ratio, and – depending on the definition of tracking error – spread and dividends. However, there are other studies with outcomes that favour synthetic replication; for his sample, Elia (2012) shows that synthetically replicating ETFs tend to exhibit a lower tracking error and higher tax efficiency, especially when the ETF is tracking an emerging market index. Hence, it might be the case that synthetic replication is only a favourable choice in cases where benchmark indices exhibit high market frictions, such as those due to stock illiquidity or regulatory issues.

Frino et al. (2004) use monthly data for the years 1994 to 1999 and show that tracking error in index mutual funds for the S&P 500 index is significantly related to index revisions, share issuances, spin-offs, share repurchases, index replication strategy and fund size. Gastineau (2002) finds for equity index funds tracking the Russell 2000 and S&P 500 indices that changes in index composition (and to a lesser extent corporate actions) have a significant effect on tracking error due to the transaction cost involved in the necessary rebalancing of the underlying portfolio. He also argues that better timing of changes in portfolio composition because of index adjustments can lead to improved returns for the investor and even outperformance of the benchmark. Yet, he acknowledges that fund managers might be constrained in their ability to deviate from precise index replication. The conjecture that tracking error magnitude is affected by inflexible replication strategies due to fund-managers' reluctance to alter their portfolio composition before the official date of index adjustment, for example, has also been posited by Blume and Edelen (2003) for index mutual funds and by Gastineau (2004) and Aber et al. (2009) for ETFs in a study of four iShares ETFs tracking broad US equity indices. In terms of share issuance in ETFs, Gallagher and Segara (2006), claim that at least in the case of creation/redemption in-kind – that is, delivery of the underlying basket in exchange for ETF shares – ETFs do not have to bear any liquidity cost and hence should not be affected in terms of tracking ability. Gastineau (2004) further points to the fact that ETF providers tend to charge a fee to the AP for the creation or redemption of shares, which should cover any other transaction cost.

Frino and Gallagher (2001) identify dividend payments as another factor with a significant impact on tracking error in passive funds. This is also verified by Frino et al. (2004) for their sample of index funds tracking the S&P 500. For the US-traded SPY-ETF tracking the S&P 500 index, Elton et al. (2002) show that the main cause of tracking error besides total expense ratio is forfeited return due to delayed reinvestment of cash dividends. Chu (2013) also finds that dividend yield has a positive impact on tracking error and claims that delays in receiving dividends and costs incurred in re-investment erode ETFs' ability to replicate index performance. For their sample of European ETFs, Blitz et al. (2012) find the explanatory power of dividend withholding taxes for fund underperformance with respect to its benchmark to be at least on par with fund expenses.[7] Applying these findings, Blitz and Huij (2012) show that emerging market equity ETFs' expected returns are equal to their respective gross benchmark index returns minus expense ratio and dividend taxes. Lin and Chou (2006) identify three factors that determine tracking error in Taiwan's first ETF, as follows: (i) cash dividends, whose impact becomes particularly obvious during peak dividend pay-out-season; (ii) management expenses – indeed, these represent the main factor causing the gap between two different tracking error series; and (iii) stock replacement operations due to, for example, index adjustments in the underlying benchmark.

With regard to the effect of market liquidity on tracking error, previous research commonly focusses on widely accepted proxies such as trading volume and bid-ask spread.[8] Kundisch and Klein (2009) show that the trading volume of the analysed ETF is negatively correlated with its tracking error; this means that on average, tracking error tends to decrease with increasing trading-volume. In contrast, Chu (2013) identifies trading volume, dividend yield and market risk to be positively related to tracking error magnitude for his sample of Hong-Kong-traded ETFs. The outcome that trading volume positively affects tracking error may be unintuitive. Yet Rompotis (2006) also presents results for his international sample of iShares, which suggest a significant positive relationship between trading volume and tracking error, although it is very small in absolute terms. Closely linked with ETF-trading-volume is its bid-ask spread. The findings of several studies suggest a positive effect of spreads on

---

7 As a consequence, Blitz et al. (2012) argue that with tracking error being inevitable, active fund managers should refrain from using a "paper" index as benchmark; rather, they should use a corresponding passive fund, which already incorporates these frictions, such as an ETF or mutual index fund. This view is supported by Kostovetsky (2003), who argues that tracking error itself is difficult to model, since there is no true benchmark for comparison. In his view, any performance comparison with paper indices is fallacious because it assumes efficient paper transactions at all times.

8 See Stoll (2003) for a detailed discussion of the bid-ask spread as indicator of the cost of trading and the illiquidity of a market.

tracking error. For instance, Milonas and Rompotis (2006), Delcoure and Zhong (2007) and to a lesser extent, Shin and Soydemir (2010) all verify that a fund's tracking error is positively affected by the bid–ask spread. Rompotis (2012) and Meinhardt et al. (2012) come to similar conclusions for the German ETF market. While Kostovetsky (2003) sees market liquidity in terms of bid-ask spread as one of the main determinants of tracking error in common index mutual funds, he rejects liquidity cost as a source of tracking error in ETFs based on the assumption that in-kind creation/redemption of ETF shares through authorised participants should shield the fund from any cost.

2.2.3 Current State of Research and Potential Research Gaps

Although several studies have suggested that there is indeed variation in ETFs' tracking ability across different markets, research on developed markets outside the US is still rather scarce. Given the importance of Germany's ETF market – with a market share of 25% in 2014, second only to the London Stock Exchange in Europe and the fourth-largest ETF marketplace in the world (Deutsche Bank Research, 2014) – further research on its microstructure could indeed prove valuable.

While an ever-growing body of literature confirms a significant impact of market liquidity on funds' tracking ability, the effect is usually addressed with proxies such as bid-ask spread or trading volume and normally on the aggregate ETF level.[9] By these means, however, it is impossible to either fully grasp all dimensions of liquidity, encompassing market breadth, market depth, immediacy of execution and market resiliency (Krogmann, 2011), or comprehend the liquidity effect of individual underlying stocks in the ETF portfolio.

Our research intends to contribute to existing literature in two ways: First, we extend the empirical evidence on market liquidity as a determinant of ETFs' tracking ability by capturing the liquidity impact of underlying stock in our analysis. We do so by

[9] See, for example, Delcoure and Zhong (2007), among others.

using Deutsche Börse’s unique volume-weighted XLM. Here, price impact information is used as a measure of the cost of immediate demand for liquidity by investors placing an order (Krogmann, 2011). XLM measures the order-size-dependent liquidity costs of a round-trip, whilst taking the entire depth of the limit order book into account, and condenses all daily market impact information for each individual stock into a single figure. Second, we examine the impact of creation and redemption of shares on an ETF’s tracking ability. Gallagher and Segara (2006) and Gastineau (2004) both present arguments against this mechanism being a potential source of daily tracking error. However, to our knowledge, their claim has not yet been empirically tested.

2.3 EMPIRICAL PART

The aim of this section is to describe our research design and subsequently report and critically discuss the results. Our study focusses on XETRA-traded funds that track equity indices in the DAX index universe, comprising Germany’s large-cap index DAX, mid-cap index MDAX, small-cap index SDAX and technology index TecDAX, as well as related sub-indices and strategy indices based on the constituents of one of the other named indices.

The overall design of our study is geared towards answering: (i) whether the liquidity cost of individual underlying securities has an impact on an exchange-traded fund’s tracking ability and (ii) whether there are additional significant effects from cash holdings, accrued dividends, cash distribution to ETF investors, the process of daily share creation and redemption or portfolio adjustments on ETF tracking error – either independently or in interaction with liquidity cost.

2.3.1 DATA AND METHODOLOGY

DATA

Except for annual expense ratios, all data are collected on a daily basis for the time period of 1 July 2003 to 31 October 2013. Since the XLM is only calculated for stocks within the universe of the DAX index family, our sample is constrained to ETFs replicating one of these indices. In the given period, a total of 22 XETRA-listed ETFs tracked relevant indices. In order to calculate a daily weighted market liquidity proxy for each ETF portfolio, we have to match daily fund holdings with the respective daily XLM data on round-trip liquidity cost for individual securities. For this reason, we are constrained to physically replicating funds in our sample. Overall, ten ETFs on XETRA fulfil this requirement, two of which have to be taken out of the sample due to insufficient data availability. Thus, the final sample consists of eight ETFs that track a total of seven different indices within the DAX universe. As of April 2014, the eight funds in the sample covered roughly 23% of total ETF trading-volume (approximately 40%, including OTC trading) on Deutsche Börse's XETRA trading platform, with the latter covering 70% of trading-volume in DAX index universe ETFs.[10]

In terms of use of income, ETFs can either be reinvesting or distributing. That is, fund management has to decide upon inception whether to accumulate or distribute the cash dividends that the fund receives from its equity investments. Accordingly, its tracking ability is measured against a performance or price-index respectively. An overview of the eight funds, their International Securities Identification Number (ISIN), their use of income and their respective benchmark is given in Table 2.1. Here, accumulating performance or total return indices are marked with TR, and distributing price indices are marked with PR.

[10] Data provided by Deutsche Börse. OTC trading volume based on all settled transactions conducted via Clearstream OTC Cascade Functionality.

Table 2.1: Overview of ETFs in observed sample

Fund name	ISIN	Benchmark	Use of income
Deka DAX® (ausschüttend) UCITS ETF	DE000ETFL060	DAX® (PR)	Distributing
Deka DAX® ex Financials 30 UCITS ETF	DE000ETFL433	DAX® ex Financials 30 (PR)	Distributing
Deka DAX® UCITS ETF	DE000ETFL011	DAX® (TR)	Reinvesting
Deka DAXplus® Maximum Dividend UCITS ETF	DE000ETFL235	DAXplus® Maximum Dividend (PR)	Distributing
iShares DAX® UCITS ETF (DE)	DE0005933931	DAX® (TR)	Reinvesting
iShares DivDAX® UCITS ETF (DE)	DE0002635273	DivDAX® (PR)	Distributing
iShares MDAX® UCITS ETF (DE)	DE0005933923	MDAX® (TR)	Reinvesting
iShares TecDAX® UCITS ETF (DE)	DE0005933972	TecDAX® (TR)	Reinvesting

This table reports the full name (as given in the official prospectus), ISIN, respective benchmark index (TR for total return and PR for price return index) and use of income for each of the eight physically replicating ETFs in the DAX index universe which represent the individual panels in regressions (2.4) to (2.6).

Daily data on prices, returns and dividends for all relevant benchmark indices and their respective underlying stocks have been compiled from Thomson Reuters' Datastream. Daily ETF fund data, particularly portfolio holdings (including cash and derivatives), NAV, assets under management, total shares outstanding, shares created or redeemed, cash distributions to investors and total expense ratio were obtained manually from the websites of the respective ETF providers. Where missing, information on historical expense ratios is supplemented by data from the Morningstar Fund Database.

Deutsche Börse has kindly provided daily XLM data for all securities in the DAX index universe. The XLM is a volume-weighted spread which is automatically calculated by the XETRA trading system for each individual security from the visible and invisible parts of the limit order book, including so-called "iceberg orders". Daily values of the XLM for each stock are calculated by XETRA as the equal-weighted average of all available minute-by-minute volume-weighted spread data points for several hypothetical standardised trading volumes (e.g. € 10,000, € 25,000, € 50,000 and

€ 100,000), thereby providing the relative liquidity cost of a round-trip for the respective trading volumes.[11]

Daily observations are only taken into account if all necessary information is available. Trading days with ETFs holding any assets other than equity or cash – derivatives such as certificates or options, for instance – are also taken out of the sample.

METHODOLOGY

In a first step, we calculate daily returns for all eight ETFs and their respective benchmark indices. Analogous to the majority of literature, we use the NAV return for the examination of an ETF's tracking error to its benchmark. One reason is that given a high-frequency trading environment and differences in exchange closing times for ETF and index trading, it is almost impossible to perfectly match daily ETF closing prices with the corresponding index prices. Another more important reason for using NAV returns is that contrary to quoted price returns, they are not biased by premiums or discounts that have not been arbitraged away. Using price instead of NAV returns would bear the risk of wrongfully attributing differences between the ETF return and benchmark return to tracking ability that are actually caused by non-arbitraged NAV-price deviations. The daily NAV return of ETF i and the return of its corresponding benchmark-index are expressed by formulae (2.1) and (2.2) respectively:

$$NR_{i,t} = \frac{NAV_{i,t} - NAV_{i,t-1}}{NAV_{i,t-1}} \tag{2.1}$$

$$IR_{i,t} = \frac{Index_{i,t} - Index_{i,t-1}}{Index_{i,t-1}} \tag{2.2}$$

[11] Further theoretical background on XETRA Liquidity Measure is provided by Hachmeister (2007), Stange and Kaserer (2011), and Rösch and Kaserer (2013).

There are actually numerous ways to define and calculate tracking error. Most studies refer to the methods brought forward by Roll (1992) and Pope and Yadav (1994). The latter authors further argue that high frequency data bear the risk of overestimating tracking error. This is why most studies base their analyses on weekly or monthly data. However, Meinhardt et al. (2012) show in their analysis of the German ETF market that the risk of overestimation of tracking error is just as high in lower-frequency data. We intend to shed light on the short-term effects of market liquidity on tracking error, and hence, as in Meinhardt et al. (2012) and Qadan and Yagil's (2012) work, our research design is based on an estimator that reflects daily tracking ability. Due to their overall structure, including the ability to create and redeem shares throughout the trading-day, ETFs are increasingly used for short-term investments and hedging strategies. In light of their potential use as short-term investment vehicles, we deem it relevant to scrutinise and better understand the effect of inter-day changes in basket liquidity on daily tracking ability. With this focus, we have to deviate from previous studies' research design, where tracking error is mostly calculated as the standard deviation of differences between benchmark and NAV over a period of time. For the calculation of daily standard deviations, intraday matching of the benchmark index and NAV returns would be required. Yet, given that the intraday NAVs that are being reported are merely indicative figures (iNAV), this procedure would be unreliable. We therefore resort to an alternative method used and corroborated by Milonas and Rompotis (2006), Shin and Soydemir (2010) and Chu (2011; 2013), among others, where we calculate daily tracking error for ETF *i* at day *t* as the absolute daily deviation of *i*'s NAV-return from its corresponding benchmark-index' return:

$$TE_{i,t} = \left| NR_{i,t} - IR_{i,t} \right| \qquad (2.3)$$

In this section, we perform a panel regression, with fund fixed effects to control for ETF-inherent characteristics. The model with the proposed factors to explain tracking error *TE* can be expressed as follows:

$$TE_{i,t} = (\alpha + v_i) + \beta_1 \cdot TER_{i,t} + \beta_2 \cdot XLM_{i,t} + \beta_3 \cdot RelCash_{i,t} + \beta_4 \cdot DivYield_{i,t} + \beta_5 \cdot DISTR_{i,t} + \beta_6 \cdot RelNetCR_{i,t} + \beta_7 \cdot PortAdj_{i,t} + \varepsilon_{it} \quad (2.4)$$

where v_i is the fixed effect of individual ETF i and *TER* is the annual total expense ratio that is charged by the issuer of exchange-traded fund i on day t. As corroborated by most of the literature, this is one of the key determinants of fund-tracking ability, and in line with previous studies, we expect *TER* to be positively related to tracking error.

XLM is a weighted average of the daily XETRA Liquidity Measure figures of all stocks held in ETF *i's* portfolio at day t (according to the stocks' respective weights in the index). Hence, it serves as a proxy for the average liquidity of ETF *i's* underlying securities. Illiquid securities are expected to have higher round-trip cost (expressed as per cent of trading volume), which should consequently drive up the overall portfolio cost. This, in turn, should have a positive impact on an ETF's daily tracking error.

A fund's daily cash holdings relative to its total assets under management are controlled for by *RelCash*. This factor should also be positively related to tracking error, for a fund's cash holdings cannot track the underlying benchmark and should hence contribute to deviations from benchmark return. Thus, tracking error should be higher in funds that have relatively higher cash holdings. The reasons for holding cash at all can be manifold: For example, the fund can face difficulties obtaining/selling the underlying securities in the aftermath of an index revision. Another reason can be that cash inflows from dividends cannot be paid out until a certain date. Dividends on securities held in the portfolio, calculated as yield to current NAV, are also separately accounted for in the model by *DivYield*. Again, the relation between this factor and tracking error should be positive, since dividends represent forfeited portfolio returns unless they are immediately reinvested. While *DivYield* accounts for cash inflows from underlying assets to the fund, we control for distribution of cash from the ETF to its shareholders by *DISTR*, which is calculated as the sum distributed to each ETF share on day t relative to the NAV of ETF i at t. By transferring cash to investors, a distributing ETF should be able to come closer to its price-index benchmark. Hence, we expect the tracking error to be negatively affected by cash distribution.

One of the most important mechanisms of an ETF is the opportunity to create and redeem shares throughout the trading day. With *RelNetCR*, a fund's daily net creation/redemption of shares is expressed as absolute relative change in current assets under management. It is calculated as the net change in shares outstanding on day t, multiplied with the closing NAV (i.e. the net transaction volume caused by the creations/redemptions on the trading day) and then divided by current assets under management. Although the creation/redemption of shares is usually performed in-kind through authorised participants, hence shielding the ETF and its assets from the largest part of transaction cost, we still expect *RelNetCR* to have a measurable positive effect on tracking error, due to imperfect index replication in the in-kind-basket delivered by or to the AP and possible delays in the settlement of the in-kind transaction. We also address the impact of market transactions beyond the dimension of creation and redemption in our model by controlling for changes in the composition of the constituents in the underlying basket through the dummy variable *PortAdj*. Given that this kind of adjustment results in transaction costs, we expect tracking error to be higher on these days.

Daily tracking error is also regressed on each independent variable individually to control for multicollinearity and to better determine explanatory power of individual factors. In model (2.5), we include squared terms of the key independent variables to identify potential non-linearity of relations between tracking error and independent variables. The model can be expressed as follows:

$$
\begin{aligned}
TE_{i,t} = {} & (\alpha + v_i) + \beta_1 \cdot TER_{i,t} + \beta_2 \cdot TER_{i,t}^2 + \beta_3 \cdot XLM_{i,t} + \beta_4 \cdot XLM_{i,t}^2 + \\
& \beta_5 \cdot RelCash_{i,t} + \beta_6 \cdot RelCash_{i,t}^2 + \beta_7 \cdot DivYield_{i,t} + \beta_8 \cdot DivYield_{i,t}^2 + \\
& \beta_9 \cdot DISTR_{i,t} + \beta_{10} \cdot DISTR_{i,t}^2 + \beta_{11} \cdot RelNetCR_{i,t} + \beta_{12} \cdot RelNetCR_{i,t}^2 + \\
& \beta_{13} \cdot PortAdj_{i,t} + \varepsilon_{it}
\end{aligned} \tag{2.5}
$$

As suggested in the previous paragraphs, we also posit that the liquidity costs of underlying securities should have a further, indirect effect on other independent variables. The reason for this is that any market transaction, whether caused by a reinvestment of cash, a creation/redemption of shares or a portfolio adjustment, should bear transaction costs. With increasing liquidity costs of underlying assets, the reinvestment of cash, portfolio adjustment or creation/redemption of shares should become more expensive for the respective ETF. *XLMxRelCash* controls for the relationship between relative cash holdings of a fund and the liquidity cost of its underlying basket. With the interaction term of *RelNetCR* and *XLM (XLMxRelNetCR)*, we can test whether it is really the AP who bears all liquidity cost, and *XLMxPortAdj* measures the interaction between liquidity and market transactions that have been initiated by management, for example due to index adjustments. Model (2.5) can hence be augmented by additional terms to account for the interaction between *XLM* and relative cash holdings, portfolio adjustments and the daily net creation/redemption of shares respectively:

$$\begin{aligned} TE_{i,t} = {} & (\alpha + v_i) + \beta_1 \cdot TER_{i,t} + \beta_2 \cdot TER_{i,t}^2 + \beta_3 \cdot XLM_{i,t} + \beta_4 \cdot XLM_{i,t}^2 + \\ & \beta_5 \cdot RelCash_{i,t} + \beta_6 \cdot RelCash_{i,t}^2 + \beta_7 \cdot DivYield_{i,t} + \beta_8 \cdot DivYield_{i,t}^2 + \\ & \beta_9 \cdot DISTR_{i,t} + \beta_{10} \cdot DISTR_{i,t}^2 + \beta_{11} \cdot RelNetCR_{i,t} + \beta_{12} \cdot RelNetCR_{i,t}^2 + \\ & \beta_{13} \cdot PortAdj_{i,t} + \beta_{14} \cdot (XLM_{i,t} \times RelCash_{i,t}) + \beta_{15} \cdot \left(XLM_{i,t} \times RelNetCR_{i,t}\right) + \\ & \beta_{16} \cdot \left(XLM_{i,t} \times PortAdj_{i,t}\right) + \varepsilon_{it} \end{aligned} \tag{2.6}$$

In order to ensure valid statistical inference of all our models, we apply robust standard errors. Driscoll and Kraay (1998) provide a computational method which generates standard errors which are heteroscedasticity and autocorrelation consistent (HAC), as well as consistent for cross-sectional dependence (as cit. in Hoechle, 2007). Furthermore, given the large number of second-order terms and interaction terms in regressions (2.5) and (2.6), we have to account for potential multicollinearity. We do so by centring all independent variables and by means of polynomial orthogonalisation. The latter approach creates a set of squared variables, from which all effects of their respective lower-order terms are removed.[12] That is, the squared terms only represent the purely non-linear effects of the independent variables.

2.3.2 EMPIRICAL RESULTS

We present our empirical findings in three steps. First, we report the descriptive statistics for the tested variables. Subsequently, we test the previously described variables individually and as part of basic models (2.4) and (2.5). Finally, in a third step, we test the full model (2.6) including interaction terms for the full period.

DESCRIPTIVE STATISTICS

The descriptive statistics for the sample ETFs' daily tracking error over time are reported in Table 2.2. For all years, tracking error exhibits a left-tailed skew and substantial leptocurtosis, implying fatter tails than observed in normal distributions. Our sample confirms the general view that on average, ETFs tend to underperform their paper-based benchmark indices: For 10,483 of the total 14,077 observations, the ETFs could not beat their respective benchmarks, underperforming by a daily average of 0.05%. On 3,594 observation days, approximately a quarter of our sample, the ETFs performed equal to or better than their respective benchmarks with an average daily outperformance of 0.15%.

---

[12] For further theoretical background on orthogonalisation, see Golub and Van Loan (1996).

Table 2.2: Descriptive statistics for daily tracking error per sample year

Year	No. of ETFs	Observations	Mean	Median	Std. dev.	Min.	Max.	Skew.	Kurtosis	Raw mean
2003	3	384	0.6035***	0.4229	0.5708	0.0010	3.2906	1.87	7.58	-0.0054
2004	3	768	0.4344***	0.3113	0.3996	0.0023	2.5390	1.62	6.10	-0.0069
2005	4	964	0.2648***	0.2005	0.2688	0.0001	2.4237	1.68	8.35	-0.0018
2006	4	1,020	0.0101***	0.0014	0.0952	0.0000	2.8268	26.00	756.01	-0.0015
2007	4	1,008	0.0100***	0.0014	0.1003	0.0000	2.8482	23.71	645.14	-0.0017
2008	6	1,350	0.0111***	0.0013	0.1319	0.0000	4.5518	30.77	1,044.41	-0.0012
2009	7	1,705	0.0360***	0.0016	0.3101	0.0000	9.1722	20.85	526.66	-0.0046
2010	7	1,785	0.0214***	0.0014	0.1756	0.0000	5.0715	21.22	534.10	-0.0018
2011	7	1,785	0.0238***	0.0016	0.1821	0.0000	4.8078	18.58	418.71	-0.0013
2012	7	1,757	0.0269***	0.0015	0.2282	0.0000	7.1482	23.05	639.37	-0.0022
2013	8	1,551	0.0261***	0.0015	0.1917	0.0000	5.3592	19.59	477.74	-0.0017
Total	**8**	**14,077**	**0.0771*** **	**0.0019**	**0.2746**	**0.0000**	**9.1722**	**10.08**	**192.76**	**-0.0024**

This table contains descriptive statistics for daily tracking error (in %) over time for the full ETF sample. It reports the mean, median, standard deviation, minimum, maximum, skewness and kurtosis for the daily tracking error per sample year. Furthermore, it reports the number of observations and ETFs in the sample that are active in the respective year. In the last column, the raw mean daily tracking error is reported (Raw mean). Tracking error is calculated according to formula (2.3) as the absolute difference between daily NAV and the index return for each ETF in the sample. Raw tracking error is calculated as the difference between the daily NAV and index return for each ETF in the sample, accounting for negativity and positivity of deviations. Years 2003 and 2013 are not complete. The statistical significance of the results being different from zero is based on a two-tailed test at the *10%, **5% and ***1% confidence levels.

Table 2.3 provides some further descriptive statistics on the key variables for the eight observed ETFs. For the period from July 2003 to October 2013, daily tracking error in the observed sample was roughly 0.08% per day. Relative cash positions (*RelCash*) average approximately 0.39% of total AuM per day, with comparably high standard deviation, suggesting considerable variation in daily cash holdings over time and across funds. The ETF with the highest cash-to-assets ratio in our sample holds an average of 2.06% of its total assets in cash, while the ETF with the lowest ratio exhibits an average cash holding of merely 0.01% relative to AuM. Relative net creation/redemption averages 1.63% of AuM per day for the full sample, meaning that on average, fund holdings fluctuate by that net figure per trading day due to newly created or redeemed fund shares. Similar to daily cash holdings, *RelNetCR* is subject to considerable variation over time and across funds, with a maximum daily average net creation/redemption relative to AuM of 4.66% and a minimum average of merely 0.12% of AuM.

Table 2.3: Descriptive statistics for the tested variables in the full sample

Variable	Observations	Mean	Std. dev.	Min.	Max.
TE	14,077	0.0771	0.2746	0.0000	9.1722
TER	14,082	0.3351	0.1587	0.0500	0.5200
XLM	12,444	0.3015	0.3853	0.0508	4.2814
RelCash	14,082	0.3856	1.1143	- 1.0015	10.0687
DivYield	12,910	0.0134	0.0747	0.0000	1.7388
DISTR	14,082	0.0167	0.3248	0.0000	17.9461
RelNetCR	14,079	1.6325	26.6014	0.0000	1,566.3320

This table reports the number of observations, mean, standard deviation, minimum and maximum observations for daily tracking error (*TE*), total annual expense ratio (*TER*), weighted underlying liquidity cost (*XLM*), cash holdings relative to total AuM (*RelCash*), dividend yield to NAV (*DivYield*), distribution to ETF shareholders relative to NAV (*DISTR*) and net creation/redemption of ETF shares expressed as relative change in current total AuM (*RelNetCR*) (all in %).

Except for the effect of portfolio adjustments, our postulations for the relations between the tested independent variables and ETF tracking error are further supported by the pairwise correlations between the variables over the full period, reported in Table 2.4. Although statistically insignificant in one case, the correlations still suggest a negative effect of *DISTR* and a positive effect of *TER*, *XLM*, *RelCash*, *DivYield* and *RelNetCR* on an ETF's tracking error.

Table 2.4: Matrix of pairwise correlations between tested variables

	TE	TER	XLM	RelCash	DivYield	DISTR	RelNetCR	PortAdj
TE	1.0000							
TER	0.1254***	1.0000						
XLM	0.0056	0.6454***	1.0000					
RelCash	0.0467***	-0.1277***	-0.1037***	1.0000				
DivYield	0.1552***	-0.0368***	-0.0347***	0.1090***	1.0000			
DISTR	-0.3707***	0.0054	0.0076	-0.0136	-0.0231***	1.0000		
RelNetCR	0.0356***	-0.0182**	-0.0183**	0.0416***	0.1989***	0.0013	1.0000	
PortAdj	-0.0146*	0.0216**	0.0286***	-0.0037	-0.0113	0.0045	-0.0014	1.0000

The matrix reports the pairwise correlations for daily tracking error (*TE*), total annual expense ratio (*TER*), weighted portfolio liquidity cost (*XLM*), cash holdings relative to total AuM (*RelCash*), dividend yield to NAV (*DivYield*), distribution to ETF shareholders relative to NAV (*DISTR*), net creation/redemption of ETF shares expressed as relative change in current total AuM (*RelNetCR*) and portfolio adjustments (*PortAdj*). The statistical significance of the results being different from zero is based on a two-tailed test at the *10%, **5% and ***1% confidence levels.

Tracking Error Determinants

The outcomes for the basic panel regression models (2.4) and (2.5) are reported in Table 2.5 and Table 2.6 respectively. Tracking error is first regressed on each of the independent variables separately before testing all factors together.

Overall, the two tables report highly significant coefficients for all tested variables. To be more specific, the outcome suggests that daily tracking error of ETFs in the DAX index universe is significantly affected by the liquidity cost of their respective underlying securities, their relative cash holdings, dividend yield, distribution of cash to investors, daily net creation and redemption of shares, as well as portfolio adjustments. The results for the tested factors clearly verify our posited hypotheses, except for that concerning the effect of portfolio adjustments. Significant results for

XLM, in particular, confirm the postulated positive effect of liquidity cost of underlying securities on ETF tracking error.

The results for the squared variables reported in Table 2.6 further suggest that the relations between tracking error on the one hand and total expense ratio, basket liquidity and distributions on the other are significantly non-linear in nature. Except for the total expense ratio, all significant non-linear relations exhibit a concave shape, implying a constantly declining slope in the magnitude of impact of these factors on ETFs' tracking ability. The respective maximum and minimum points of effect on tracking error for identified non-linear effects are reported in Appendix B.

In regression (2.6), we control for potential interactions of liquidity cost with relative cash holdings, the daily creation/redemption process of ETF shares and portfolio adjustments (see Table 2.7). With regard to the key independent variables in the full sample, we obtain results that are similar to the ones observed in Table 2.5 and Table 2.6. That is, all key variables have a statistically significant impact on tracking error, with *TER*, *XLM* and *DISTR* exhibiting non-linear relationships. Similar to the findings in the basic model (2.5), *DISTR* has the greatest single effect with a beta of approximately -0.12, followed by *TER*, *DivYield*, the *XLM* effect of underlying liquidity and *RelCash*.[1] Relative net creation/redemption of shares also has a significant, albeit rather small effect on *TE*. From the insignificance of *XLMxRelNetCR*, we can further infer that the impact of the creation or redemption of ETF shares on daily tracking error is independent from liquidity cost involved in obtaining or selling the basket of underlying securities. The only significant interaction in our augmented model (2.6) is between liquidity cost and portfolio adjustments. This suggests that in addition to its direct effect on ETF tracking ability, the liquidity cost of underlying securities also has an indirect effect through its interaction with market transactions caused by changes in the composition of the ETF's portfolio constituents.

---

[1] Note that with the inclusion of interaction terms, the interpretation of all coefficients changes. In model (2.6), an independent variable (e.g. XLM) represents the unique effect under the assumption that all other variables that might be interacting with it (XLM) are equal to zero.

Table 2.5: Regression of individual determinants of tracking error and model (2.4)

Variable	**Model (2.4)**							
TER	1.0546***							1.0203***
	(0.1542)							(0.1551)
XLM		0.0821**						0.0760**
		(0.0259)						(0.0236)
RelCash			0.0187***					0.0157***
			(0.0044)					(0.0044)
DivYield				0.5310***				0.4873***
				(0.0397)				(0.0374)
DISTR					-0.3127**			-0.3584**
					(0.1252)			(0.1333)
RelNetCR						0.0004***		0.0001*
						(0.0001)		(0.0001)
PortAdj							-0.0474***	-0.0230**
							(0.0110)	(0.0089)
Constant	0.0771***	0.0611***	0.0771***	0.0615***	0.0771***	0.0771***	0.0775***	0.0618***
	(0.0064)	(0.0055)	(0.0071)	(0.0056)	(0.0070)	(0.0071)	(0.0072)	(0.0044)
Observations	14,077	12,444	14,077	12,907	14,077	14,074	14,077	12,442
R^2(adj.)	0.0480	0.0054	0.0042	0.0238	0.1396	0.0015	0.0002	0.2908
# of ETFs	8	8	8	8	8	8	8	8
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Max. VIF	-	-	-	-	-	-	-	1.05

This table reports the results for the fixed effects regression of tracking error (*TE*) on individual centred independent variables as well as on regression (2.4) (in %). The tested independent variables are total annual expense ratio (*TER*), weighted portfolio liquidity cost (*XLM*), cash holdings relative to total AuM (*RelCash*), dividend yield to NAV (*DivYield*), distribution to ETF-shareholders relative to NAV (*DISTR*), net creation/redemption of ETF shares expressed as relative change in current total AuM (*RelNetCR*), and portfolio adjustments (*PortAdj*). Respective standard errors are reported in parentheses. Furthermore, the number of observations, adjusted R-squared, the number of ETFs for the respective regression and the maximum variance inflation factor (VIF) are stated. The statistical significance of the results being different from zero is based on a two-tailed test at the *10%, **5% and ***1% confidence levels.

Table 2.6: Regression of individual determinants of tracking error and model (2.5)

Variable				**Model (2.5)**				
TER	0.0594***							0.0577***
	(0.0047)							(0.0046)
TER^2	0.0850***							0.0859***
	(0.0064)							(0.0063)
XLM		0.0188***						0.0178***
		(0.0044)						(0.0041)
XLM^2		-0.0225***						-0.0207***
		(0.0049)						(0.0043)
RelCash			0.0177***					0.0135***
			(0.0042)					(0.0033)
$RelCash^2$			-0.0076					0.0028
			(0.0088)					(0.0081)
DivYield				0.0396***				0.0372***
				(0.0031)				(0.0033)
$DivYield^2$				0.0032				0.0021
				(0.0069)				(0.0059)
DISTR					-0.1016***			-0.1165***
					(0.0263)			(0.0275)
$DISTR^2$					-0.0666***			-0.0678***
					(0.0119)			(0.0105)
RelNetCR						0.0104***		0.0042**
						(0.0026)		(0.0015)
$RelNetCR^2$						-0.0023		0.0011
						(0.0025)		(0.0013)
PortAdj							-0.0474***	-0.0195**
							(0.0110)	(0.0077)
Constant	0.0771***	0.0611***	0.0771***	0.0615***	0.0771***	0.0771***	0.0775***	0.0584***
	(0.0052)	(0.0054)	(0.0071)	(0.0056)	(0.0070)	(0.0071)	(0.0072)	(0.0031)
Observations	14,077	12,444	14,077	12,907	14,077	14,074	14,077	12,442
R^2 (adj.)	0.1242	0.0122	0.0047	0.0239	0.1995	0.0015	0.0002	0.4716
# of ETFs	8	8	8	8	8	8	8	8
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Max. VIF	1.00	1.00	1.00	1.00	1.00	1.00	-	1.08

This table reports the results for the fixed effects regression of tracking error (*TE*) on individual centred independent variables, including squared terms, as well as on regression (2.5) (in %). The tested independent variables are total annual expense ratio (*TER*), weighted portfolio liquidity cost (*XLM*), cash holdings relative to total AuM (*RelCash*), dividend yield to NAV (*DivYield*), distribution to ETF-shareholders relative to NAV (*DISTR*), net creation/redemption of ETF shares expressed as relative change in current total AuM (*RelNetCR*), and portfolio adjustments (*PortAdj*). Respective standard errors are reported in parentheses. Furthermore, the number of observations, adjusted R-squared, the number of ETFs for the respective regression, and the maximum variance inflation factor (VIF) are stated. The statistical significance of the results being different from zero is based on a two-tailed test at the *10%, **5% and ***1% confidence levels.

Table 2.7: Regression of models (2.5) and (2.6)

Variables	Model (2.5)	Model (2.6)
TER	0.0577***	0.0576***
	(0.0046)	(0.0046)
TER^2	0.0859***	0.0859***
	(0.0063)	(0.0063)
XLM	0.0178***	0.0179***
	(0.0041)	(0.0041)
XLM^2	-0.0207***	-0.0208***
	(0.0043)	(0.0043)
RelCash	0.0135***	0.0139***
	(0.0033)	(0.0034)
$RelCash^2$	0.0028	0.0034
	(0.0081)	(0.0083)
DivYield	0.0372***	0.0373***
	(0.0033)	(0.0033)
$DivYield^2$	0.0021	0.0022
	(0.0059)	(0.0059)
DISTR	-0.1165***	-0.1165***
	(0.0275)	(0.0275)
$DISTR^2$	-0.0678***	-0.0678***
	(0.0105)	(0.0105)
RelNetCR	0.0042**	0.0055**
	(0.0015)	(0.0017)
$RelNetCR^2$	0.0011	0.0021
	(0.0013)	(0.0014)
PortAdj	-0.0195**	-0.0179**
	(0.0077)	(0.0072)
XLMxRelCash		-0.0306
		(0.0304)
XLMxRelNetCR		-0.0011
		(0.0007)
XLMxPortAdj		-0.0851**
		(0.0281)
Constant	0.0584***	0.0585***
	(0.0031)	(0.0031)
Observations	12,442	12,442
R^2 (adj.)	0.4716	0.4718
# of ETFs	8	8
Fixed effects	Yes	Yes
Max. VIF	1.08	1.88

This table reports the results for fixed effects regressions (2.5) and (2.6) of tracking error (*TE*) on centred independent variables (in %), namely total annual expense ratio (*TER*), weighted portfolio liquidity cost (*XLM*), cash holdings relative to total AuM (*RelCash*), dividend yield to NAV (*DivYield*), distribution to ETF shareholders relative to NAV (*DISTR*), net creation/redemption of ETF shares expressed as relative change in current total AuM (*RelNetCR*), portfolio adjustments (*PortAdj*) and the interaction terms of XLM with *RelCash* (*XLMxRelCash*), *RelNetCR* (*XLMxRelNetCR*), and portfolio adjustments (*XLMxPortAdj*). Respective standard errors are reported in parentheses. Furthermore, the number of observations, adjusted R-squared, the number of ETFs for the respective regression and the maximum variance inflation factor (VIF) are stated. The statistical significance of the results being different from zero is based on a two-tailed test at the *10%, **5% and ***1% confidence levels.

2.3.3 DISCUSSION OF FINDINGS

The finding of all tested key variables exhibiting significant effects on tracking error in ETFs corroborates most of our postulated hypotheses. The liquidity cost of underlying securities in particular has a measurable and positive effect on the tracking ability of ETFs. The findings further confirm our notion that cash holdings, dividend yield and daily creation/redemption of ETF-shares have a significant and positive effect on tracking error. Cash distributions to ETF-investors appear to substantially reduce return deviations from a price-index-benchmark, as do portfolio adjustments, albeit to a much smaller degree.

The empirical evidence on the effects of liquidity of the underlying portfolio on ETF tracking error presented in this paper can be considered an extension of the works by Rompotis (2012) and Meinhardt et al. (2012), who confirm a positive impact for German ETFs. While these works address the relation between liquidity and tracking ability on an aggregate level of ETF liquidity, we take a more bottom-up approach by taking liquidity of individual stocks into consideration. Theoretically, the only time that the liquidity cost of underlying securities is bound to have an effect is at the occurrence of a market transaction. Then we should expect liquidity costs to play a role only in events triggering portfolio adjustments. Controlling for cash holdings, portfolio adjustments and creation/redemption separately, we still find *XLM* to have a strongly significant and independent effect on ETF tracking error. A possible explanation is that relatively small internally initiated market transactions that are not fully captured by the rather large-transaction-oriented factors *RelNetCR* or *PortAdj* cause liquidity-related transaction costs. These can occur, for instance through constant rebalancing by fund-management in order to better match or optimise index weights over time. In these cases, XLM represents the liquidity cost borne by the ETF for its attempts to optimise the weights of the underlying portfolio.

This, in turn, suggests that although daily creation/redemption usually takes place as an in-kind-transaction, ETFs are not fully protected from the effect of the liquidity cost of their underlying securities. Whilst the magnitude of the positive relation between liquidity cost and tracking error in absolute figures is indeed small, we still have to challenge Kostovetsky's (2003) claim that liquidity is not at all a determining factor of tracking error. To put the absolute magnitude of the figures into some perspective, it should be re-emphasised that this study addresses tracking error at the daily level.

For our set of German DAX index family ETFs, we find cash flows into or out of the fund to have a substantial effect on tracking error. In terms of dividend yield, our results are in line with the findings of Elton et al. (2002) and Blitz et al. (2012), who show for the US and European markets respectively, that the main cause of tracking error besides total expense ratio is forfeited return due to delayed reinvestment of cash dividends. Fund managers have been aware of this effect for quite some time and apply various strategies to reduce or at least control the effect, provided of course that such measures are possible given the regulatory constraints. The notion of forfeited returns on delayed investment especially holds true for distributing ETFs. Being unable to reinvest the proceeds from dividend payments, these ETFs have to rely on a few fixed distribution dates to reduce accumulated cash holdings. Again, fund management is usually aware of this source of tracking error and it might indeed be worthwhile to determine whether it is the optimal strategy to solely rely on a few distribution dates or whether there are still better ways to smooth the effects of idle cash on tracking error.

Frino and Gallagher (2002) use cash flows in their model to control for cost-inducing cash transactions of passive funds. However, to our knowledge, daily holdings of cash relative to total assets under management as potential explanatory variable of ETF tracking error have not been tested yet. This having been said, our analysis does indeed verify the positive and significant impact on tracking ability.

The process of creation and redemption of shares is fundamental to the entire concept of ETFs, as it ensures continuous arbitrage trading and hence NAVs which are very close to index prices. However, reviewing the available literature, creation/redemption does not appear to be a topic of focus for research on tracking error. Both Gallagher and Segara (2006) and Gastineau (2004) argue that ETFs should not be affected in their tracking ability by creation/redemption of shares due to the in-kind-delivery and fees charged to the AP. Still, our findings confirm a positive independent effect on tracking error. With the insignificance of the interaction term between *RelNetCR* and *XLM*, it becomes apparent that the process of creating and redeeming ETF shares has an effect on tracking ability beyond the dimension of liquidity cost of the underlying basket. Thus, while we can agree with Gallagher and Segara (2006) and Gastineau (2004) that the cost of creation/redemption is successfully transferred to the AP, we have to challenge their conclusion that this implies that creation/redemption has no impact on ETF tracking ability at all. One possible explanation for the effect has to do with the imperfect replication of index weights: Since creation/redemption of shares takes place in pre-defined units, it is almost impossible for an AP to perfectly allocate the corresponding NAV value among the index constituents due to indivisibility of individual stock shares. That is, there will most probably be a remainder in cash or in stocks that does not perfectly match the actual index weights. As a result, the ETF will exhibit tracking error due to differences between the benchmark and the underlying basket. Yet this effect should be rather small for ETFs with large assets under management. Another explanation could be daily charging or attribution of fees from or to the fund: Although they are effectively paid on a monthly or quarterly basis, fees to or from the fund like securities-lending fees or management fees are commonly calculated and attributed to ETF-NAV on a daily basis. A substantial change in AuM from one day to another due to creation/redemption, would also significantly affect the base for calculating such fees and hence daily attribution of fees to NAV.

With portfolio adjustments having a negative effect on an ETF's tracking error, we have to reject the hypothesis that they reduce tracking ability through transaction costs. Instead, we observe days on which ETFs change the composition of their underlying portfolio by swapping constituents (mostly due to officially announced index adjustments) to exhibit less pronounced tracking error even after controlling for the interaction with liquidity cost. This suggests that portfolio adjustments have an impact on tracking error beyond the dimension of liquidity-related transaction costs, allowing the portfolio to better track the benchmark index. Once more, we attribute this effect to the imperfect replication of index weights in the ETF portfolio: Similar to cash distributions to investors, portfolio adjustment events appear to be a chance to dispose of unwanted weight deviations that have accumulated over time.

Since XLM is only available for constituents of the DAX index universe, our ETF-sample is constrained to the German equity market. We still consider the sample to be representative for the German market for two reasons: First, with 23% of the total ETF trading volume (approximately 40% including OTC trading) as of April 2014, the sample represents a substantial part of overall ETF trading in Germany. Second, in terms of trading volume and market capitalisation, the replicated indices cover by far the largest part of the German equity market. Even for the identified funds, data availability is somewhat poor, with some funds exhibiting data gaps in their respective time series and some funds falling out of the sample altogether. However, the issue of data gaps is mainly driven by an individual fund, which exhibits a longer period of missing data. Therefore, we believe that in the panel cross-section, the effect should be smoothed and should not substantially affect results.

We understand that the high number of squared terms and interaction terms in our models bears the risk of high multicollinearity. While acknowledging this risk, we still consider it essential to leave the variables in the model and to control for all these factors separately. We partly circumvent the problem by centring all independent variables and by orthogonalisation of all second-order terms. As a result, the measured maximum variance inflation factors do not exceed 1.88 for any of our models.

2.4 CONCLUSION

In this study, we try to identify the determining factors of daily tracking error in ETFs in the German DAX index universe. We particularly focus on the liquidity of individual underlying securities as a potential factor affecting ETF tracking ability.

As postulated, we find daily tracking error to be dependent on liquidity of stocks in the underlying portfolio, management fees, cash holdings, dividend yield, cash distributions from an ETF to its investors, portfolio adjustments and the process of creation/redemption of ETF shares. Liquidity affects tracking error both directly and in interaction with portfolio adjustments, implying that liquidity cost plays a significant role in various events that trigger a market transaction. Still, cash inflows and outflows in the form of dividends and distributions respectively, appear to be the factors with the most substantial effect on tracking error in our model.

Research on some of the tested variables seems to have just commenced. This is especially true for liquidity of individual underlying stocks or the process of creation/redemption of ETF shares. In particular, our findings on the latter clearly challenge the current notion of ETF tracking ability being immune against effects from creation/redemption due to in-kind transactions.

It is our aim to contribute to a broader discussion of potential tracking error determinants and to provide some new insights to their dependence on market conditions. Due to the lack of availability of XLM data, research beyond the DAX index universe appears to be unmanageable at this time. Yet, there are other liquidity proxies that are similar to Deutsche Börse's XLM, such as the cost of round-trip trade (CRT) introduced by Irvine, Benston and Kandel (2000), which could help to further elaborate on true liquidity effects on ETF tracking error in other markets. Furthermore, since there is still no consensus in the literature on the impact of economic regimes on tracking error and its determinants, it might also prove valuable to apply our model to sample periods with more extreme return patterns, such as the global financial crisis.

3. IMPACT OF DESIGNATED MARKET MAKERS ON ETF LIQUIDITY IN SECONDARY MARKETS – A CASE FOR ALGORITHM BASED TRADING

ABSTRACT

We analyse the effect of employing additional DMM on the liquidity cost of ETFs listed on Deutsche Börse's XETRA-platform. In particular, we try to determine whether market makers, who predominantly rely on algorithmic- or high-frequency trading (HFT) are able to systematically provide better liquidity than other types of market makers. Applying a difference-in-difference model with propensity score matching based control groups, we show that the hiring of any additional market maker significantly reduces liquidity costs of an ETF. We mainly argue that competition between market makers causes liquidity costs to drop. We also find significant and substantial additional liquidity gains from mandating an external HFT market maker over a non-HFT market maker. Yet, given that the outperformance is only significant two or three months after treatment, it appears that the HFT market maker needs some time to fully exploit its systemic advantage.

Keywords: ETF, Exchange-Traded Fund, Market Making, Liquidity Cost, Market Liquidity, XETRA Liquidity Measure, High-Frequency Trading, Algorithmic Trading

JEL Classification: G12, G14, G15, G23

Authors: Christoph Kaserer, Friedrich Osterhoff

First Author: Friedrich Osterhoff

Current Status: Working Paper (submitted for publication)

3.1 INTRODUCTION

Liquidity is a central aspect of market quality for any security. This is especially true for exchange-traded funds (ETFs) where liquidity, together with expense ratios and tracking error, is the key differentiator on which most institutional investors base their investment decisions (Greenwich Associates, 2014). With liquidity playing such a decisive role, an increasing number of financial markets rely on designated market makers (DMMs) to enhance liquidity beyond endogenous levels (cf. Anand et al., 2009). Since market making often represents a considerable share of income in an ETF's business model, investment banks or trading desks that are closely affiliated with the ETF provider often assume the DMM role. However, recently, market making has become less attractive to ETF providers as a source of income. First, with overall spreads having declined substantially for years, profit margins have increasingly come under pressure (see Appendix C). Second, due to technological advances, market makers are facing further increases in the already high cost of a competitive trading infrastructure. Third, while algorithmic and high-frequency trading (HFT) are already an integral part of market making and a potential source of a competitive edge, the public debate on their merits and on adequate regulation is still ongoing; consequently, the future of this specific business model is somewhat uncertain. Facing these challenges, several ETF providers have decided to completely or partly outsource the market making to banks or specialists.

Many papers examine the effects of market making and of algorithmic and high-frequency trading on market liquidity and they have shown that: (i) DMMs do improve market liquidity (e.g. Nimalendran & Petrella, 2003; Anand & Weaver, 2006; Venkataraman & Waisburd, 2007; Anand et al., 2009; Perotti & Rindi, 2010; Skjeltorp & Ødegaard, 2010; Menkveld & Wang, 2013; Anand & Venkataraman, 2013) and (ii) with the advent of algorithmic and high-frequency trading, market quality has improved significantly (e.g. Hendershott et al., 2011; Hendershott & Moulton, 2011; Riordan & Storkenmaier, 2012; Hasbrouck & Saar, 2013; Hendershott & Riordan, 2013; Malinova et al., 2013; Brogaard, et al., 2013; Boehmer et al., 2014; Brogaard et al., 2014;

Jarnecic & Snape, 2014). However, research on the nexus of secondary market liquidity and ETFs is scarce at best. This is partly due to the fact that the discussion of ETF liquidity is dominated by the widely accepted view that the key (if not the only relevant) driver is the liquidity of the underlying basket (e.g. Ryan & Follet, 2001; Kittsley & Edrosolan, 2008; Agrrawal & Clark, 2009; Calamia et al., 2013; 2014; Roncalli & Zheng, 2014). In addition, ETFs are often viewed as being too similar to stocks to make specific research on ETF liquidity worthwhile. However, Borkovec and Serbin (2013) show that ETFs exhibit liquidity and cost characteristics that are qualitatively different from common stocks with comparable volume, volatility, spread and price levels.

In this period of increased outsourcing activity among ETF providers, this paper seeks to extend the research on ETF liquidity and contribute to the understanding of the relationship between market making and market liquidity. More specifically, we try to determine whether hiring an additional market maker has a measurable effect on an ETF's liquidity cost and whether a designated liquidity provider that relies predominantly on algorithmic or high-frequency trading is systematically better at generating liquidity than other types of market makers. We are among the first to elaborate on these issues with a clear focus on ETFs. Compared to studies that examine the impact of market making on liquidity at the level of individual stocks, our research design has an advantage in that, by addressing the effect at the level of aggregated portfolios, we are able to mitigate any potential uncontrolled liquidity effects of individual stocks and, hence, reduce the risk of endogeneity.

Most other studies solely rely on ETF bid-ask spreads to approximate market liquidity. In doing so, they fail to address market depth as a decisive factor in liquidity. In contrast, we use a more holistic proxy for market liquidity in our model: Deutsche Börse's volume-weighted spread XETRA Liquidity Measure (XLM). XLM measures the order-size-dependent liquidity costs of a round-trip for ETFs, taking the entire depth of the limit order book into account (cf. Krogmann, 2011; Stange & Kaserer, 2011;

Hendershott & Riordan, 2013; Rösch & Kaserer, 2013). Applying XLM should allow for a more elaborate view of the liquidity costs of the observed equity ETFs.

The results of our difference-in-difference treatment models corroborate the notion that contracting an additional market maker immediately and substantially reduces liquidity costs, regardless of whether the new market maker applies algorithmic or high-frequency trading. We attribute this effect to the growing competition between individual DMMs, which leads to decreases in liquidity costs. Our findings suggest that merely relying on one DMM bears the risk for an ETF that trading may not be executed at the lowest possible levels of liquidity costs; this potentially tarnishes its attractiveness to investors. We also find some evidence that HFT market makers are better at providing liquidity; yet, given that the outperformance only becomes significant two or three months after hiring the new DMM, it appears that the HFT market maker needs some time to fully exploit its potential.

The remainder of this paper is structured as follows: In Section 3.2, we provide an overview of the relevant literature. In Section 3.3, we describe the empirical components of the study, such as the data and the methodology, and discuss our findings. In Section 3.4, we provide concluding remarks and present an outlook on potential fields of future research.

3.2 LITERATURE REVIEW

Given our focus on how DMMs affect the liquidity cost of ETFs in a technologically advancing market environment, we review the relevant empirical literature on ETF liquidity and its determinants in general (3.2.1), the impact of designated market makers on market liquidity (3.2.2) and the effect of increased algorithmic and high-frequency trading on market quality (3.2.3).

3.2.1 ETF LIQUIDITY

A wide variety of studies examines ETF liquidity and its key determinants. Overall, there is a broad consensus that the prevailing factor driving the liquidity of an ETF is the liquidity of its underlying basket. Ryan and Follet (2001), Kittsley and Edrosolan (2008), Agrrawal and Clark (2009), Calamia et al. (2013; 2014), and Roncalli and Zheng (2014), among others, all confirm the view that ETF liquidity effectively depends on the liquidity of its respective benchmark. However, recent findings suggest that, despite being an overriding factor, the liquidity of the underlying index is not the only element explaining liquidity on the aggregated ETF level. Agrrawal and Clark (2009) develop a ranking algorithm for secondary market liquidity and show that, on average, the most liquid ETFs typically have lower bid-ask spreads, larger assets under management, lower expense ratios and higher average trading volumes; they also tend to have at least a five-year trading history. For European ETFs, Calamia et al. (2013) identify fund size and market fragmentation as determinants of liquidity. In their study of European ETFs, Roncalli and Zheng (2014) challenge the relation between AuM and liquidity and they show that the market is dominated by institutional investors with long holding periods (including seed money). They also provide evidence that liquidity varies considerably between ETF providers (i.e. ETFs replicating the same benchmark exhibit significant differences in liquidity). They argue that fragmentation, the cross-listing of ETFs and the low transparency of the European ETF market are the most important factors causing this effect. Kittsley and Edrosolan (2008) find that secondary market liquidity is also determined by risk, market activity and trading volume. The effect of the latter is further supported by Calamia et al. (2014), who show that, in their sample, underlying volatility and trading volume affect ETF liquidity. Their findings suggest that market makers will trade ETFs like stocks as long as the trading volume is sufficiently high to manage the inventory with low risk. Consequently, they argue that market makers will only account for the illiquidity of the underlying stock basket in their quoted spread when trading volumes are too low. Sanchez and Wei (2010) examine the bid-ask spread, spread information component and the holding period of 77 ETFs and find that, compared to less-diversified sector funds, broad-based ETFs exhibit lower spreads. In their sample, the overall liquidity of ETFs is not unambiguously better in comparison to their underlying stocks; although ETFs are traded more frequently, they still have greater total spreads.

3.2.2 MARKET MAKING

While the asynchronous arrival of buyers and sellers on a market results in trading uncertainty, Demsetz (1968) argues that the risk connected to it can be mitigated by the regular presence of market makers or dealers on a market (as cit. in Venkataraman & Waisburd, 2007). By simultaneously posting limit orders on both sides of the electronic limit order book, these market makers act as counterparty for any market participant who wants to trade immediately and in doing so, they provide liquidity (Jones, 2013). The concept of regular liquidity providers is far from new; their role in electronic limit order markets has already been extensively appraised both in theoretical analyses (e.g. Anand & Subrahmanyam, 2008; Bessembinder et al., 2011; Bessembinder et al., 2013), as well as empirical studies (Skjeltorp & Ødegaard, 2010). Yet, while the literature on market making in general is abundant, most models emphasise endogenous liquidity provision. That is, they focus on the behaviour of dealers and traders in the absence of any obligation to supply liquidity (Bessembinder et al., 2013). However, several exchanges have a system of designated market makers[14] where market makers are contractually obliged to ensure certain levels of liquidity in return for compensation.[15]

Overall, the empirical findings on the effect of DMMs on market quality clearly corroborate the idea that they significantly improve liquidity. Anand and Weaver

---

14 For a comprehensive overview of stock exchanges working with designated liquidity providers, see Charitou and Panayides (2009).

15 Under the assumption that they can buy at the bid price and sell at the ask price, a key component of market maker compensation is the bid-ask spread. However, given that DMMs might be contractually obliged to take a side in a transaction, which they would refuse to do under free market conditions, they are often further compensated for their liquidity provision through a fixed sum and/or trading cost rebates.

(2006) use intraday options data from the Chicago Board Options Exchange (CBOE) to analyse the impact of superimposing a specialist system on an existing multiple market maker system, and they show that following the adoption of a specialist system, quoted, current and effective spreads decrease. They further find that after the switch, the market share of the CBOE increases significantly, suggesting that specialists use spreads to attract order-flow. Nimalendran and Petrella (2003) study the specialist system at the Italian Stock Exchange (ISE) and find evidence that a hybrid system of voluntary market-driven market makers and DMMs offers superior performance with regard to market quality, especially execution costs and market depth. They also find that thinly-traded stocks benefit more from the adoption of a hybrid trading system, both in terms of greater liquidity and lower adverse selection costs. To some degree, the suitability of their methodology and the robustness of their results is challenged by Battalio (2003), who argues that there are substantial structural differences between ISE specialists and specialists on the New York Stock Exchange (NYSE). On the other hand, Perotti and Rindi (2010) corroborate the findings of Nimalendran and Petrella (2003) by analysing a set of small-cap and mid-cap stocks in the ISE, thereby showing that the presence of a DMM reduces spread and price volatility, the probability of informed trading and the adverse selection component of the spread.

The idea that small-cap and inactive stocks in particular are affected by contracting DMMs is further supported by Venkataraman and Waisburd (2007), who suggest that especially younger, smaller, and less volatile firms tend to contract a designated dealer; those that do exhibit better market quality and their stocks experience an average cumulative abnormal return of 4.9% around the announcement of dealer introduction, which is positively correlated with improvements in liquidity. Menkveld and Wang (2013) use the roll-out of Euronext's limit order market system to the Amsterdam exchange in 2001 to run an event study based on small-cap stocks, and they find that contracting a market maker improves liquidity levels, reduces liquidity risk and generates an average abnormal return of 3.5%. These results are consistent with the study conducted by Anand et al. (2009) that examines the effect of DMMs at the Stockholm Stock Exchange and provides evidence that firms that contract market makers experience significant improvements in market quality and price discovery and decreased costs of capital. Moreover, they find that contractual or exogenous liquidity attracts significant additional endogenous liquidity. Skjeltorp and Ødegaard (2010) present similar results for the Oslo Stock Exchange where the secondary market liquidity of stocks improved following the appointment of DMMs, especially in cases where smaller firms relied on the services of DMMs. Anand and Venkataraman's (2013) study even suggests that, due to the lack of endogenous market makers for small stocks, DMMs are the only reliable counterparty available in this segment of the market.

Similar to our study, Hengelbrock (2008) examines the effect of taking on additional DMMs in Deutsche Börse's XETRA electronic limit order book, the so-called designated sponsors. He provides evidence that while bid-ask spreads narrow when market making is performed by one or two designated sponsors, additional specialists do not necessarily result in higher liquidity. His results differ both across market segments and across different types of market makers, with the spreads being lower for ETFs that contract brokers instead of banks. He also provides evidence that the observed spread decline is not necessarily primarily caused by a decrease in adverse selection costs; rather it is the result of inter-dealer competition and risk sharing. While most empirical studies on the effects of market making on market quality focus on stocks, De Winne et al. (2013) use order book data for a CAC 40 ETF and its respective index constituents to show that DMMs significantly contribute to the liquidity of the ETF market.

In summary, the findings from the presented studies suggest that a purely endogenous liquidity provision may not be optimal for all securities and that, in terms of market quality, there are potential benefits to be gained from contracting one or even more DMMs to ensure certain levels of market liquidity.

3.2.3 Algorithmic Trading in Market Making

In basic terms, market makers aim to earn the bid-ask spread by buying at the bid price and selling at the ask price. Given that they bear the risk of trading with and losing money to informed counterparties during these transactions, they have an incentive to make sure that their orders incorporate as much current information as possible, as quickly as possible (Jones, 2013). Based on this rationale, Biais and Foucault (2014) name three reasons why speed is a key factor in market making: first, a more immediate reaction to transient increases in market illiquidity; second, speed mitigates the market makers' exposure to the risk of being picked off by an informed counterparty through faster reaction to new information; and, third, it results in a more efficient management of inventory risk. With markets recently experiencing a significant automation and acceleration trend, market making has seen a shift away from the purely human protagonist towards more automated processes.[16] Whereas algorithmic or high-frequency trading strategies are not necessarily novel or more sophisticated, they can usually be executed at considerably lower costs, allowing for gains that may be passed on to the market and investors in the form of narrower bid-ask spreads and smaller commissions (cf. Jones, 2013). In addition to the cost argument, Hendershott and Riordan (2013) show that, for constituents of the German DAX index, where algorithms account for just over half of the market and nonmarketable limit order volumes, algorithmic traders monitor market liquidity more actively than human traders, consuming liquidity when the bid-ask quotes are narrow and supplying liquidity when they are wide. Algorithmic traders also react more quickly to events, even more so when the bid-ask spreads are wide.

Several studies try to determine the overall effect of automation and increased speed on market quality. An analysis of NASDAQ order book data by Hasbrouck and Saar (2013) suggests that increased low-latency activity leads to decreased bid-ask spreads, increased displayed depth in the limit order book and lower short-term volatility, both

---

16 For a general overview of the literature on algorithmic and high-frequency trading, see Gomber et al. (2011) and Biais and Foucault (2014).

during normal and exceptionally volatile periods. Hendershott et al. (2011) examine the increase in algorithmic trading after the automation of quote dissemination at the NYSE in 2003, and they show that, for large stocks in particular, it narrows spreads and reduces adverse selection as well as trade-related price discovery. Their results are further corroborated by Boehmer et al. (2014), who, using the same proxy for algorithmic trading, also find a negative correlation between algorithmic trading and bid-ask spreads, especially for large, high-priced stocks with low volatility. They even find that liquidity in the smallest tercile of stocks declines in light of greater intensity of algorithmic trading.[17] Additionally, Riordan and Storkenmaier's (2012) study suggests that the greatly reduced system latency of the XETRA trading platform results in smaller quoted and effective spreads. Yet, contrary to Hendershott et al. (2011) and Boehmer et al. (2014) they find the effect to be mainly concentrated in small- and medium-sized stocks. Hendershott and Moulton (2011) analyse the introduction of the Hybrid Market at the NYSE in 2006, which increased automation and greatly reduced execution time for market orders. They find that the change reduces noise in prices, making them more efficient. However, they also show that it raises the cost of immediacy (i.e. bid-ask spreads) because of increased adverse selection. With regard to high-frequency trading, Brogaard et al. (2013) provide evidence that an increase in HFT activity following a technology upgrade at the London Stock Exchange does not have any measurable effect on institutional traders' execution costs. Brogaard et al. (2014) analyse 120 randomly selected NASDAQ stocks and NYSE stocks for 2008 and 2009 and find that, overall, HFT facilitates price efficiency by trading in the direction of permanent price changes and in the opposite direction of transitory pricing errors, both on average and on the highest volatility days. Jarnecic and Snape (2014) examine the order submission strategies and supply of liquidity by HFT participants in the limit

[17] Biais and Foucault (2014) advise some caution when interpreting the findings of both Hendershott et al. (2011) and Boehmer et al. (2014). First, any proxy for HFT based on the normalised number of messages is likely to reflect the activity of both HFT and algorithmic trading operating at lower frequencies. Hence the studies' findings may simply reflect the effect of algorithmic trading operating at a relatively slow speed rather than the effect of HFT. Second, it could be a crowding out effect due to more intense competition from HFT reducing the likelihood of profitable execution for slow investors.

order book and provide evidence that confirms the notion that high-frequency trading improves market liquidity even in fast or volatile markets; still, issues remain regarding the effect of HFT participants on market depth. Finally, Malinova et al. (2013) show reductions in HFT activity in the Canadian stock market to be associated with an increase in spreads and worsened execution quality for retail investors.

Overall, with a few exceptions, the relevant literature on algorithmic and high-frequency trading clearly suggests that almost every time a change in market structure has resulted in lower latency and more algorithmic and high-frequency trading, market quality in general, and liquidity, in particular, have improved significantly.

3.2.4 Current State of Research and Potential Research Gaps

The literature on determinants of ETF liquidity is abundant, as is empirical evidence on the effect of DMMs and algorithmic trading on market liquidity. Still, studies at the nexus of market making and liquidity, on the one hand, and ETFs, on the other hand, are scarce at best. This is partly because equity ETFs are often treated as being too similar to stocks to make individual research worthwhile. However, Borkovec and Serbin (2013) show that ETFs exhibit qualitatively different liquidity and cost characteristics than common stocks with comparable volume, volatility, spread and price levels.

Our research intends to bridge this gap in the literature in several ways. First, by examining the effect of additional DMMs on the liquidity cost of ETFs, we contribute to the research on determinants of ETF liquidity and offer perspectives beyond basket liquidity. This should be of particular relevance to ETF providers who have to decide whether the investment for hiring more than one designated sponsor is rewarded with higher market quality. Furthermore, by addressing the effect of market making on liquidity at the aggregated portfolio level of ETFs, we are able to mitigate any potential uncontrolled liquidity effects of individual stocks on the results, thereby reducing the risk of endogeneity. While most research still uses bid-ask spreads as a proxy for liquidity, Barclay et al. (1999), among others, propose more advanced execution-cost-based measures that consider the bid-ask spread and the depth of the market.[18] By employing Deutsche Börse's unique volume-weighted XETRA Liquidity Measure (XLM) as a proxy for liquidity, we also avoid the pitfall of neglecting market depth. XLM is calculated using price impact information as a measure of the cost of immediate demand for liquidity by investors placing an order (cf. Krogmann, 2011; Hendershott & Riordan, 2013). By considering the entire depth of the limit order book, the proxy measures the order-size-dependent liquidity costs of a round-trip of a given order volume and condenses all of the market impact information for each individual stock into a single figure. Finally, building on Hengelbrock's (2008) idea of potential variation in liquidity impact between different types of DMMs, we contribute to the literature on algorithmic trading by testing whether market makers that explicitly and predominantly rely on high-frequency trading provide systematically better levels of liquidity to an ETF than those that do not explicitly and predominantly apply algorithmic or high-frequency trading.

3.3 EMPIRICAL PART

The overall design of our study is geared towards answering: (i) whether mandating an additional designated sponsor significantly improves the secondary market liquidity cost of an ETF and (ii) whether market makers who explicitly and predominantly rely on algorithmic and high-frequency trading have a stronger impact on ETF liquidity than non-HFT market makers. Before we describe our data and methodology, we first provide a broad overview of the institutional background of XETRA and its system of designated sponsors.

---

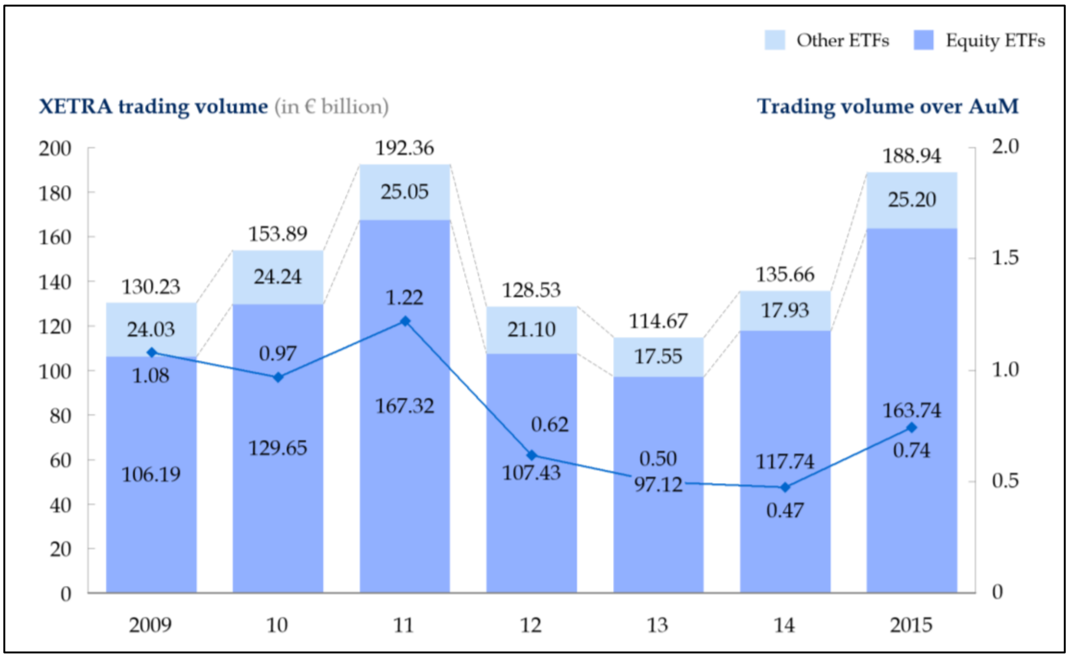
18 See Roncalli and Zheng (2014) for an overview of liquidity proxies, including their merits and flaws.

3.3.1 INSTITUTIONAL BACKGROUND

DEUTSCHE BÖRSE'S XETRA TRADING SYSTEM

XETRA is a fully electronic trading system with a centralised limit order book operated by Deutsche Börse for equity, ETFs, exchange-traded products, mutual funds, certificates, options, bonds and rights issues. Continuous trading takes place from 9:00 am until 5:30 pm, framed by an opening and closing call auction starting at 8:50 am and 5:30 pm respectively. Generally, matching of buy and sell orders takes place according to a price- and time-priority rule, and visible orders always take priority over hidden orders. Trading is anonymous for all participants, and trades are processed through a central counterparty (CCP).

Figure 3.1: XETRA order book ETF trading volume and turnover 2009–2015



This chart shows the annual ETF trading volume on XETRA in billions of euros (LHS) for equity and other ETFs for 2009–2015. It also reports the turnover as trading volume over AuM for all XETRA-traded ETFs (RHS). Trading volume over AuM is based on a full year analysis with year-end AuM. (Source: Deutsche Börse)

XETRA has seen a surge in ETF trading in recent years, with a rapid increase in the number and variety of instruments and the amount of AuM. Despite a significant drop in trading volume, as well as volume relative to AuM in 2012 (see Figure 3.1), as of

2014 XETRA still remains the second largest market by order book volume in Europe with a market share of more than 25% in European ETF trading (Deutsche Bank Research, 2014), making it a key reference market for exchange-based trading of German-listed exchange-traded funds.

THE ROLE OF DESIGNATED SPONSORS ON XETRA

Listing an ETF on XETRA requires the mandating of at least one designated liquidity provider, a so called designated sponsor. As for any DMM, the main task of a designated sponsor on XETRA is to enhance liquidity by quoting binding bid- and ask-prices to ensure that investors find counterparties for a transaction at almost any time (Hengelbrock, 2008). They also participate in auctions and quotes during volatility interruptions and/or fast markets. Depending on the asset class, designated sponsors are expected to satisfy minimum requirements for several quality criteria as stipulated by the Frankfurt Stock Exchange. For ETFs, designated sponsors have to quote at least 80% of the effective trading time during continuous trading, and they have to have a planned and opening auctions participation rate of at least 80% (see Appendix D). Provided that designated sponsors fulfil both criteria at a 90% rate, they are reimbursed in full for any transaction fees accumulated in the course of market making operations. In addition, the ETF issuer determines a minimum volume for each ETF (i.e. the smallest tradable number of shares on the buy and sell side) and a maximum bid-ask spread. Apart from the reimbursement of transaction fees, designated sponsors are compensated for their services in several ways. They usually receive an annual fee from the ETF provider, which may consist of a fixed sum as well as a variable component linked to the fulfilment of contractually agreed-upon quality criteria. Cross-selling aspects may also play a role, as most designated sponsors provide additional services, such as acting as authorised participants and handling the creation and redemption of new shares (cf. Hengelbrock, 2008).

3.3.2 DATA AND METHODOLOGY

DATA

We identify all passively managed equity ETFs listed on XETRA between December 2010 and July 2015. We exclude actively managed ETFs because sudden changes in liquidity of an active ETF might be due to the potentially frequent changes in its underlying portfolio rather than exogenous effects, such as hiring an additional market maker. Since active ETFs represent a small fraction of the market in terms of number, AuM and trading volume, the adjustment should not bias the sample. For each identified ETF, we gather the following information from Deutsche Börse's monthly turnover-statistics and ETF master-data: issuer, method of replication, asset class, listing date, benchmark, AuM, XETRA Liquidity Measure (XLM) in basis points for a hypothetical round-trip trade of € 100,000, trading volume in the XETRA order book and OTC volume settled via Clearstream's CASCADE platform, as well as the number and identity of all active designated sponsor(s).

The observed period encompasses a total of 951 equity ETFs replicating 480 different indices. Of those, 522 ETFs replicating 295 indices satisfy the data requirements for our analysis. Due to model specifications, each ETF in the final sample has to have a minimum of 25 months of data history: 12 months prior to the month of (potential) treatment, the month of (potential) treatment, and 12 months afterwards. Overall, this gives us a total of 8,390 valid observations – 7,470 and 7,462 control observations for the analysis of 172 observed instances of mandating non-HFT and 280 instances of mandating HFT market makers for individual ETFs respectively.[19]

[19] The overlap of observations is due to the fact that several observations fulfil the requirements for both the non-HFT and HFT control groups.

METHODOLOGY

We conduct an event study where we estimate the treatment effect of hiring the new market maker on fund liquidity cost. While estimating treatment effects by means of regression is quite common among researchers, Li (2013) correctly points out that this method bears the risk of an omitted variable bias and, thus, endogeneity. Similar to Jovanovic and Menkveld (2011) and Yao and Ye (2014), among others, we employ a difference-in-difference approach to determine the treatment effect of contracting an additional market maker. The key assumption behind the difference-in-difference approach is that, in the absence of treatment, the unobserved differences between the treatment and control groups are the same over time. This postulation can be expressed as:

$$E\left(Y_{0,Post} - Y_{0,Pre} \mid T = 1\right) = E\left(Y_{0,Post} - Y_{0,Pre} \mid T = 0\right) \tag{3.1}$$

where $Y_{0,Post}$ and $Y_{0,Pre}$ are the respective liquidity cost levels of untreated ETFs after and before the time of treatment, *T* is the treatment (i.e. hiring of an additional market maker) with the value 1 for the treated group and the value 0 for the control group, and *E(.)* are the expected unobserved differences for the treatment and control groups. The expected change in liquidity of the untreated ETFs in the control group over time is expected to be equal to the change in liquidity of the group of treated ETFs, had they not received treatment. Based on this assumption, we can calculate the average treatment effect on the treated ETFs (ATT) as:

$$ATT = E\left(Y_{1,Post} - Y_{1,Pre} \mid T = 1\right) - E\left(Y_{0,Post} - Y_{0,Pre} \mid T = 0\right) \tag{3.2}$$

where ATT is the difference between the expected change in liquidity for the treated ETFs and the expected change in liquidity for the untreated ETFs in the control group.

The difference should then be solely attributable to the treatment (i.e. the introduction of an additional market maker).

For each observation we calculate the change of liquidity cost in terms of XLM basis points over the time between 1 month prior to the potential treatment and 1, 2, 3, 6 and 12 months after the potential treatment respectively. Put differently, if t is the month in which the hire (T) takes place, then Y_{Pre} is the liquidity cost at $t-1$ and Y_{Post} is the liquidity cost of that same ETF at $t+1$, $t+2$, $t+3$, $t+6$ or $t+12$ respectively. The information in treatment month t is intentionally neglected in order to avoid the possible frictions caused by the implementation of a new market maker to affect the final estimation results. We then compare the change of liquidity cost for each of the treatment instances with the average change of liquidity in an individual control group consisting of 5, 10, 15, 20, 25, 30 or 50 matched untreated ETF observations, and we calculate the difference-in-difference to determine the actual treatment effect.

For the assumption expressed in equation (3.1) to hold, it is necessary to apply a robust method when defining the control groups for each observed treatment observation. In order to minimise the risk of selection bias in the estimation of treatment effects, similarity is the key concern when defining suitable control groups. To address that, we use Rosenbaum and Rubin's (1983) proposed method of propensity score matching (PSM). A propensity score can be defined as the conditional probability that a unit in the sample receives the treatment, given a set of observed variables X. It is a balancing score, which means that, given the balancing/propensity score, the conditional distribution of the observed covariates in vector X is the same for the treated group and the untreated control group (Rosenbaum & Rubin, 1983). Formally, the balancing score $b(X)$ satisfies:

$$X \perp T \mid \mathrm{b}(X) \tag{3.3}$$

Different from matching directly on the X covariates, PSM reduces all of the information that potentially has an effect on the probability of hiring an additional market maker to one dimension by condensing it into a single continuous covariate, which then serves as the matching variable for the treatment and control groups. The propensity of ETF i to receive treatment can then be expressed as:

$$Pr(T_i = 1 \mid \mathrm{X}_i) \tag{3.4}$$

If all of the information relevant to treatment participation and outcomes is observable, the propensity score should produce valid matches for estimating the impact of a treatment. Then, the individual observations can be compared on the basis of their propensity scores rather than matching them on individual factor-values in vector X. In our analysis, an ETF's propensity to receive the treatment is estimated for each monthly observation via a logit-regression on observed characteristics that are assumed to have an impact on the treatment decision. Then, each treatment observation is compared with the control groups of 5 to 50 matched untreated observations (T=0) that are the nearest neighbours to the treated observation (T=1) in terms of propensity to treatment. In our logit-model, the factors assumed to be relevant for the probability of hiring an additional market maker are the six months moving averages of the fund's prior liquidity cost and the volatility of liquidity cost over time, the monthly trading volume[20] relative to AuM, the number of designated sponsors already in place, and the market share in terms of AuM relative to competitors replicating the same underlying index. Liquidity and trading volume are also used in Anand et al.'s (2009) analysis of market maker impact on the Stockholm Stock Exchange. Although we effectively look at the impact of designated sponsors on the on-exchange liquidity on XETRA, we still incorporate OTC trading volume in our model. Giulianini (2012) estimates that roughly 70% of all ETF trading in Europe is

20 Both order book trading via XETRA and all OTC trading, settled via Clearstream's CASCADE system.

OTC. While it is difficult to determine a precise number, especially for the US markets, it is possible to retrieve figures for OTC trades in Germany that are settled through CASCADE. Although this figure is only an approximation of the total OTC trading volume, we still deem it viable to control for the visible part of OTC trades, especially given its potential share in total trading volume. We also control for method of replication, fund currency and quarterly time fixed effects, as well as for abnormal changes in liquidity cost over the period of the last 12 months prior to treatment, in order to adjust for cases where the ETFs experienced extreme liquidity movements in the recent past that might lead to a bias in the selection of the control group.

We execute this procedure twice to independently calculate the ETFs' propensities to hire an additional non-HFT market maker and a HFT market maker respectively, for each observation. This ensures that the respective treatments (*T=1*) of hiring a non-HFT or HFT market maker are compared with suitable control groups (*T=0*) in which propensity of treatment refers to the market maker type in question. After estimating the respective average treatment effects on the HFT treated ETFs, we compare the results via two-sample t-tests (adjusted for dissimilar variances in the samples) to determine whether HFT market makers have a significantly greater impact on liquidity cost. Even if there were systematic differences between the non-HFT and HFT treated ETFs, we could still directly compare the ATTs from both types of treatment, since our difference-in-difference approach with matched control groups should cancel out any non-treatment-related effects. It may appear to be more straightforward to determine the systematic effect of HFT market making on liquidity cost by calculating the difference-in-difference between the two forms of treatment, that is, those events in which HFT market makers are contracted and those in which non-HFT market makers are hired. However, this procedure would only generate a very small sample with 280 treatment observations and 172 observations out of which control groups of suitable sizes would have to be built. Moreover, we feel that while it is possible to define the variables that determine the decision of ultimately hiring an additional market maker, it is quite difficult to model the ETF provider's choice of selecting a HFT market maker over a non-HFT market maker. This decision might partly be based on non-controllable factors, such as reputation, prior business-relations, competitive pricing and negotiations.

While the literature presents various approaches to define and differentiate between algorithmic and high-frequency trading, the studies usually describe high-frequency trading as a subset of algorithmic trading (e.g. Gomber et al., 2011; Jones, 2013; Brogaard et al., 2013; Biais & Foucault, 2014; Brogaard et al., 2014). Biais and Foucault (2014) consider speed to be the key differentiator between the two; while all algorithmic traders use algorithms and computer programs to implement their trading strategies – for example brokers may rely on algorithms to split larger orders into smaller optimal lots – not all of them rely on speed. Hendershott and Riordan (2013) take a more generalist stance and assert that any proprietary trading algorithm can usually be referred to as high-frequency trading. Brogaard et al. (2014) acknowledge the difficulties in clearly identifying high-frequency trading, especially for larger and integrated market participants who engage in HFT and, simultaneously, act as brokers for customers and engage in proprietary lower-frequency trading strategies. For our analysis, we cannot identify individual trades on XETRA as HFT or non-HFT. We rely on publicly available information to determine whether a certain designated sponsor's business model can be described as predominantly engaging in high-frequency trading activities.

In both PSM-processes, observations are excluded if the respective fund does not exist for at least another 12 months. In doing so, we try to ensure that the potential anomalies of funds that are about to cease to exist shortly after treatment do not bias our estimators. In order to avoid overlapping effects, the observations in the respective control groups have to be free from any change in their market maker setup for at least three months prior to and after the treatment date t. ETFs already employing a HFT market maker at the time of observation or at any point up to three months prior to or after t are excluded from the treatment group and control group for the non-HFT PSM-model.

Abadie and Imbens (2012) derive the large sample distribution of PSM estimators and prove that the estimation of the propensity score affects the large sample distribution of the PSM estimators. For the ATT estimator, they show that ignoring the estimation error in the propensity score may lead to confidence intervals that are either too large or too small and derive adjustments to the large sample variances of the PSM estimators. Based on their findings, we calculate robust standard errors for our estimation of the treatment effects.

With the methodology and underlying terminology being defined, we describe and formalise our hypotheses on the effect of additional market makers on ETF liquidity. We postulate that mandating an additional designated sponsor significantly reduces the liquidity cost of an ETF in terms of XLM basis points. In our hypothesis, the additional market maker always has a significant impact on liquidity cost, regardless of the number of designated sponsors employed prior to treatment. Since we separately test for the impact of non-HFT and HFT market makers, we divide this postulation into two hypotheses:

Hypothesis 1: *Mandating an additional non-HFT market maker significantly reduces liquidity cost.*

This can be expressed as:

$$ATT_{non-HFT} = E\left(Y_{1,Post} - Y_{1,Pre} \mid T = 1\right) - E\left(Y_{0,Post} - Y_{0,Pre} \mid T = 0\right) < 0 \tag{3.5}$$

Hypothesis 2: *Mandating an additional HFT market maker significantly reduces liquidity cost.*

This can be expressed as:

$$ATT_{HFT} = E\left(Y_{1,Post} - Y_{1,Pre} \mid T = 1\right) - E\left(Y_{0,Post} - Y_{0,Pre} \mid T = 0\right) < 0 \tag{3.6}$$

We further hypothesise that, relative to other types of market makers, designated sponsors, whose business model mainly relies on algorithmic and high-frequency trading, generate more substantial reductions in liquidity cost for an ETF in terms of XLM basis points. Thus, we develop the following hypothesis:

Hypothesis 3: *The designated sponsors that mainly rely on high-frequency techniques have a significantly greater effect on liquidity cost than non-HFT market makers.*

This can be expressed as:

$$(ATT_{HFT} - ATT_{non-HFT}) < 0 \tag{3.7}$$

3.3.3 EMPIRICAL RESULTS

We present our empirical findings in three steps. First, we report descriptive statistics for the key characteristics of our ETF sample. We then test the postulated hypotheses by means of difference-in-difference models and two-sample t-tests.

DESCRIPTIVE STATISTICS

Out of the 522 ETFs in our sample, 163 are physically replicating while the remainder are swap-based ETFs. Table 3.1 provides further information on the monthly observations in the sample. At the time of observation, an average ETF in the sample has been listed for almost 40 months, has just over € 81 million in AuM, and its liquidity cost for a € 100,000 round-trip transaction is approximately 51 basis points. Yet, with the observed XLM ranging from 4.56 to 432.97 basis points, liquidity costs are clearly widely dispersed. An ETF has an average of four competitors on the same index. However, looking at the mean and median for market share, it becomes apparent that assets under management are usually concentrated among one or two ETFs. While some ETFs rely on up to six designated market makers, in more than half of all the observations the ETF merely employed the minimum required number of one designated sponsor.

Table 3.1: Descriptive statistics for ETF characteristics in the sample

	Mean	Median	Std. dev.	Min.	Max.	Skew.	Kurtosis
Age	39.89	36.77	21.68	0.90	155.97	1.67	7.82
AuM	81.02	27.63	202.79	0.06	4,475.35	10.75	172.51
Liquidity Cost	51.01	37.44	44.25	4.56	432.97	2.70	14.07
Trading Volume	13.02	1.97	44.13	0.00	1,690.37	14.54	382.39
Turnover	0.26	0.06	5.16	0.00	462.27	88.84	7,954.50
No. of Competitors	4.10	2.00	4.52	1.00	24.00	2.17	7.45
Market Share	46.88	33.76	40.08	0.03	100.00	0.30	1.39
No. of DMM	1.36	1.00	0.64	1.00	6.00	2.10	8.49

This table reports the descriptive statistics for 8,390 observations used in the matching processes. It provides mean, median, standard deviation, minimum, maximum as well as skewness and kurtosis for the key characteristics of the monthly observations in the sample, namely fund age in months, assets under management (AuM) in millions of euros, liquidity cost measured in XLM basis points for a hypothetical trading volume of € 100,000, monthly trading volume (including OTC trading, settled via Clearstream) in million euros, turnover measured as monthly trading volume (including OTC trading, settled via Clearstream) relative to AuM, number of competitors replicating the same index, market share in replicated index relative to competitors (in %) and number of designated sponsors.

AVERAGE TREATMENT EFFECT OF AN ADDITIONAL DMM ON TREATED ETFS

The outcomes for estimating the effect of hiring an additional designated sponsor on ETF liquidity cost are described in Table 3.2 and Table 3.3 for non-HFT and HFT market makers respectively. Both tables report the difference-in-difference treatment effects compared with the control groups of 5 to 50 nearest neighbour observations. Overall, the data in both tables suggest that the introduction of an additional designated sponsor has an immediate, significant and economically relevant effect on an ETF’s liquidity cost on secondary markets. Regardless of the size of the control group, the effect remains significant over the entire observation period ending 12 months after the treatment. Therefore, Hypotheses 1 and 2 are clearly corroborated by our results. With the exception of one outlier in each of the two sample sets – *t-1* in the

M=10 model for non-HFT market makers and *t-3* in the M=15 model for HFT market makers – all of the models produce insignificant results for the treatment effect at any point in time before the new designated sponsor is contracted.

Contracting a designated sponsor relying on non-HFT methods significantly reduces an ETF's liquidity cost (see Table 3.2). Already one month after hiring the new DMM, liquidity cost is reduced by approximately 7.52 XLM basis points[21] on average for a hypothetical € 100,000 round-trip transaction. While there is some variation in the effect depending on the time of measurement, with distinctly smaller absolute effects being measured two months after treatment, the results still suggest relatively stable and sustainable liquidity gains over time of up to 8.26 XLM basis points after treatment. All the results for the time after treatment are highly significant (only a few exceptions are not significant at 1% confidence levels).

The results for DMMs relying primarily on algorithmic and high-frequency trading are quite similar to those for non-HFT market makers (see Table 3.3). Contracting a HFT market maker significantly reduces the liquidity cost for the ETF on XETRA. Regardless of the time of observation and the size of the control group, all the results are highly significant at 1% confidence levels. In absolute terms, one month after hiring an additional market maker the average treatment effect on the treated is approximately 8.52 XLM basis points for a hypothetical € 100,000 round-trip transaction. Different from the results for non-HFT designated sponsors, we observe increasing liquidity gains in the case of HFT market makers until six months into the treatment, after which it stabilizes at around 13.16 XLM basis points.

21 If not stated otherwise, all figures for XLM basis points in the following sections are averages over the varying control group sizes M=5, 10, 15, 20, 25, 30 and 50.

Table 3.2: ATT of hiring an additional non-HFT DMM

Period	$ATT_{non\text{-}HFT}$ M=5	$ATT_{non\text{-}HFT}$ M=10	$ATT_{non\text{-}HFT}$ M=15	$ATT_{non\text{-}HFT}$ M=20	$ATT_{non\text{-}HFT}$ M=25	$ATT_{non\text{-}HFT}$ M=30	$ATT_{non\text{-}HFT}$ M=50
t-12	-0.169	0.016	0.039	-0.035	0.038	-0.177	-0.311
	(3.194)	(3.130)	(3.136)	(3.139)	(3.143)	(3.166)	(3.171)
t-6	-0.644	-0.484	-0.378	-0.389	-0.327	-0.172	-0.006
	(0.782)	(2.221)	(2.172)	(2.159)	(2.117)	(2.115)	(2.108)
t-3	-0.788	-0.935	-0.797	-0.689	-0.707	-0.654	-0.546
	(1.500)	(1.281)	(1.492)	(1.610)	(1.643)	(1.662)	(1.689)
t-2	-1.653	-1.820	-1.425	-1.162	-0.876	-0.755	-0.581
	(1.317)	(2.452)	(2.404)	(2.413)	(2.392)	(2.388)	(2.358)
t-1	-0.791	-0.818***	-0.385	-0.444	-0.236	-0.316	-0.208
	(1.390)	(0.149)	(2.040)	(2.027)	(2.013)	(2.009)	(1.990)
t+1	-7.973***	-7.102***	-7.506***	-7.466***	-7.670***	-7.569***	-7.344***
	(1.588)	(1.163)	(0.690)	(1.945)	(1.946)	(1.940)	(1.947)
t+2	-5.093***	-4.716***	-5.036***	-5.050***	-5.123**	-5.163**	-5.103**
	(1.609)	(1.299)	(0.939)	(0.202)	(2.086)	(2.076)	(2.084)
t+3	-6.947***	-6.431***	-6.357***	-6.303***	-6.341***	-6.308***	-6.228***
	(1.691)	(1.366)	(1.046)	(0.605)	(2.121)	(2.111)	(2.097)
t+6	-8.821***	-8.383***	-8.247***	-8.048***	-8.009***	-8.102***	-8.201***
	(1.666)	(1.455)	(1.205)	(0.783)	(2.121)	(2.136)	(2.119)
t+12	-7.946***	-7.478***	-7.484***	-7.704***	-7.658***	-7.678***	-7.845***
	(1.922)	(1.588)	(1.185)	(0.345)	(2.435)	(2.426)	(2.421)

This table reports the average treatment effect of contracting an additional non-HFT designated sponsor on the treated ETFs (ATT) as calculated with the difference-in-difference model described in Section 3.3.2. It is measured as a change in liquidity cost in XLM basis points for a hypothetical round-trip transaction of € 100,000. The average treatment effect on the treated is calculated 1, 2, 3, 6 and 12 months before and after the time of treatment (*t*). Matching is done via a logit-regression based propensity score estimator as described in Section 3.3.2, and the results are reported for individual control groups for each of the 172 treatment observations with varying sizes (M) of 5, 10, 15, 20, 25, 30 or 50 nearest neighbour control group observation matches. Abadie and Imbens' (2012) robust standard errors are reported in parentheses. Statistical significance of the result being different from zero based on a two-tailed test at *10%, **5% and ***1% confidence levels.

Table 3.3: ATT of hiring an additional HFT DMM

Period	ATT_{HFT} M=5	ATT_{HFT} M=10	ATT_{HFT} M=15	ATT_{HFT} M=20	ATT_{HFT} M=25	ATT_{HFT} M=30	ATT_{HFT} M=50
t-12	-0.153	-1.394	-0.509	-0.327	-0.419	-0.305	-0.350
	(2.160)	(1.483)	(1.197)	(3.520)	(3.489)	(3.486)	(3.451)
t-6	0.477	-0.211	0.730	0.930	0.711	0.806	0.462
	(1.467)	(0.921)	(2.975)	(2.920)	(2.884)	(2.873)	(2.835)
t-3	-0.419	-1.237	-0.808*	-0.663	-0.598	-0.403	-0.181
	(1.687)	(1.060)	(0.445)	(2.314)	(0.378)	(2.314)	(2.320)
t-2	0.176	-1.294	-0.575	-0.545	-0.292	-0.091	-0.082
	(1.501)	(1.491)	(1.493)	(1.440)	(1.429)	(1.429)	(1.520)
t-1	0.973	-0.208	0.068	0.152	0.173	0.169	0.136
	(1.305)	(1.264)	(1.152)	(1.076)	(1.111)	(1.106)	(1.122)
t+1	-8.828***	-7.886***	-8.083***	-8.602***	-8.633***	-8.698***	-8.919***
	(1.962)	(1.714)	(1.655)	(1.524)	(1.438)	(1.323)	(1.345)
t+2	-8.592***	-7.580***	-7.842***	-8.431***	-8.434***	-8.448***	-7.820***
	(2.138)	(1.901)	(1.904)	(1.774)	(1.699)	(1.645)	(1.557)
t+3	-11.781***	-10.719***	-10.820***	-11.544***	-11.526***	-11.352***	-10.810***
	(2.146)	(1.913)	(1.898)	(1.794)	(1.746)	(1.672)	(1.583)
t+6	-14.340***	-13.844***	-13.676***	-14.350***	-14.483***	-14.432***	-13.772***
	(2.232)	(2.019)	(1.961)	(1.796)	(1.776)	(1.653)	(1.551)
t+12	-14.641***	-13.188***	-13.020***	-13.138***	-13.160***	-12.969***	-11.895***
	(2.584)	(2.318)	(2.235)	(2.044)	(1.995)	(1.849)	(1.661)

This table reports the average treatment effect of contracting an additional HFT designated sponsor on the treated ETFs (ATT) as calculated with the difference-in-difference model described in Section 3.3.2. It is measured as a change in the liquidity cost in XLM basis points for a hypothetical round-trip transaction of € 100,000. The average treatment effect on the treated is calculated 1, 2, 3, 6 and 12 months before and after the time of treatment (*t*). Matching is done via a logit-regression based propensity score estimator as described in Section 3.3.2, and the results are reported for individual control groups for each of the 280 treatment observations with varying sizes (M) of 5, 10, 15, 20, 25, 30 or 50 nearest neighbour control group observation matches. Abadie and Imbens' (2012) robust standard errors are reported in parentheses. Statistical significance of the result being different from zero based on a two-tailed test at *10%, **5% and ***1% confidence levels.

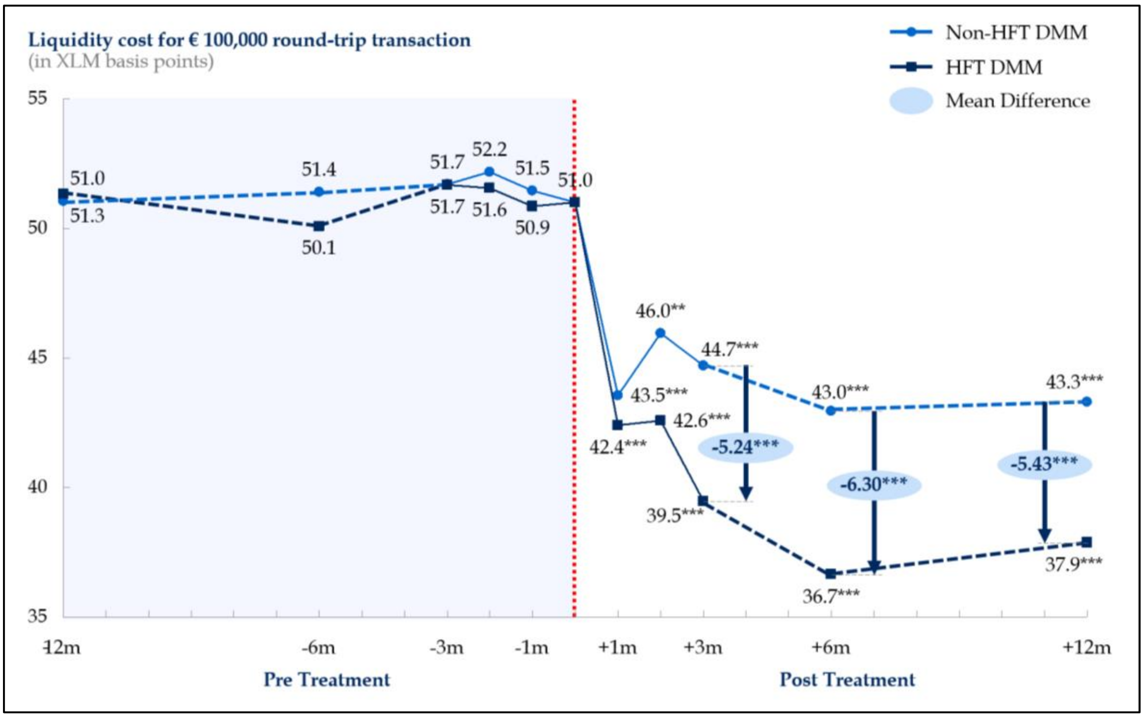
Table 3.4 reports the mean difference between the average treatment effects on the treated ETFs of hiring an additional non-HFT and HFT market maker respectively. Two types of standard errors are shown for each mean difference. First, the common standard errors based on the average treatment effects for the two observation sets of 172 non-HFT treatments and 280 HFT treatments are shown. However, as previously described, Abadie and Imbens (2012) prove that the estimation of the propensity score affects the large sample distribution of the PSM estimators, and they show that ignoring the estimation error in the propensity score may lead to the confidence intervals being either too large or too small. This fact is not accounted for in the standard two-sample t-test, which we use to determine the statistical significance of the mean differences. Therefore, we adjust the standard errors to account for the issues raised by Abadie and Imbens (2012), and we report them in italics underneath the default standard errors. Overall, the negativity of all the mean differences between the non-HFT and HFT ATT-results indicates a general outperformance of HFT over non-HFT market making. However, the statistical evidence for a systematic outperformance of HFT market makers over non-HFT competitors over time is somewhat weaker. One month after treatment there is no significant difference in liquidity reduction between the two groups. A more pronounced and statistically significant difference in the effect on liquidity cost can be found two, three, six and 12 months after treatment. Yet, the absolute mean differences and the statistical significances slightly decline again towards t+12. In summary, in our sample we observe a superior liquidity generation of HFT market makers over non-HFT market makers, at least for certain points in time after treatment; thus, Hypothesis 3 as expressed in equation (3.7) cannot be ultimately rejected.

Table 3.4: Mean difference between ATT from hiring an additional HFT DMM and non-HFT DMM

Period	Δ ATT M=5	Δ ATT M=10	Δ ATT M=15	Δ ATT M=20	Δ ATT M=25	Δ ATT M=30	Δ ATT M=50
t-12	0.016	-1.410	-0.548	-0.291	-0.457	-0.128	-0.039
	(4.599)	(4.519)	(4.487)	(4.481)	(4.466)	(4.477)	(4.426)
	(3.856)	*(3.464)*	*(3.357)*	*(4.716)*	*(4.696)*	*(4.709)*	*(4.687)*
t-6	1.122	0.273	1.109	1.319	1.039	0.978	0.468
	(3.539)	(3.451)	(3.410)	(3.390)	(3.354)	(3.353)	(3.329)
	(1.662)	*(2.404)*	*(3.683)*	*(3.631)*	*(3.578)*	*(3.567)*	*(3.533)*
t-3	0.369	-0.303	-0.011	0.026	0.109	0.251	0.365
	(3.003)	(2.858)	(2.804)	(2.813)	(2.802)	(2.798)	(2.775)
	(2.258)	*(1.663)*	*(1.557)*	*(2.819)*	*(1.686)*	*(2.849)*	*(2.870)*
t-2	1.828	0.526	0.850	0.617	0.584	0.664	0.499
	(2.813)	(2.753)	(2.699)	(2.706)	(2.678)	(2.682)	(2.643)
	(1.997)	*(2.870)*	*(2.830)*	*(2.810)*	*(2.786)*	*(2.783)*	*(2.806)*
t-1	1.764	0.610	0.453	0.596	0.409	0.484	0.344
	(2.306)	(2.285)	(2.251)	(2.237)	(2.224)	(2.220)	(2.206)
	(1.907)	*(1.273)*	*(2.343)*	*(2.295)*	*(2.299)*	*(2.294)*	*(2.284)*
t+1	-0.856	-0.784	-0.578	-1.137	-0.964	-1.129	-1.574
	(2.984)	(2.887)	(2.854)	(2.855)	(2.853)	(2.856)	(2.862)
	(2.525)	*(2.072)*	*(1.793)*	*(2.471)*	*(2.419)*	*(2.349)*	*(2.366)*
t+2	-3.499*	-2.864	-2.806*	-3.381**	-3.310	-3.285	-2.717
	(3.185)	(3.066)	(3.056)	(3.040)	(3.027)	(3.027)	(3.023)
	(2.676)	*(2.302)*	*(2.123)*	*(1.785)*	*(2.690)*	*(2.649)*	*(2.601)*
t+3	-4.834**	-4.288**	-4.463**	-5.241***	-5.184**	-5.044**	-4.582**
	(3.281)	(3.139)	(3.104)	(3.114)	(3.098)	(3.098)	(3.086)
	(2.732)	*(2.351)*	*(2.167)*	*(1.893)*	*(2.748)*	*(2.693)*	*(2.628)*
t+6	-5.519**	-5.461**	-5.429***	-6.302***	-6.474***	-6.330***	-5.571**
	(3.434)	(3.356)	(3.319)	(3.292)	(3.289)	(3.296)	(3.290)
	(2.785)	*(2.489)*	*(2.302)*	*(1.959)*	*(2.766)*	*(2.701)*	*(2.626)*
t+12	-6.696**	-5.710**	-5.536**	-5.434***	-5.502**	-5.290**	-4.050*
	(3.711)	(3.605)	(3.573)	(3.572)	(3.554)	(3.545)	(3.515)
	(3.220)	*(2.810)*	*(2.530)*	*(2.072)*	*(3.148)*	*(3.050)*	*(2.936)*

This table reports the mean differences between the average treatment effects (Δ ATT) of contracting an additional non-HFT designated sponsor and of contracting an additional HFT designated sponsor, as presented in Table 3.2 and Table 3.3 respectively. This is measured as the absolute difference in the change of liquidity cost measured in XLM basis points for a hypothetical round-trip transaction of € 100,000. The mean differences between the average treatment effects are calculated 1, 2, 3, 6 and 12 months before and after the time of treatment (*t*) and for different control group sizes (M). Standard errors are reported in parentheses. In addition, Abadie and Imbens' (2012) robust standard errors is reported underneath in parentheses in italics. Statistical significance of the result being different from zero based on a one-tailed two-sample t-test that accounts for different variances in both samples at *10%, **5% and ***1% confidence levels.

Figure 3.2: Impact of non-HFT and HFT market makers on liquidity over time



This figure illustrates the effect of hiring an additional non-HFT or HFT designated sponsor on the liquidity of a hypothetical exchange-traded fund (with the liquidity cost being equal to the sample-mean) over time (for control group M=20). The exchange-traded fund's liquidity cost is measured in XLM basis points for a hypothetical round-trip transaction of € 100,000. The time of treatment (i.e. contracting of the additional designated market maker) is visualised with a vertical dotted line. The shaded area on the left-hand side represents the 12 months before treatment, the area on the right-hand side of the vertical treatment line represents the 12 months after treatment. Mean difference between $ATT_{non\text{-}HFT}$ and ATT_{HFT} is shown in shaded bubbles. The statistical significance of ATT results and the mean difference results being different from zero are based on a two-tailed t-test and a one-tailed two-sample t-test respectively, at *10%, **5% and ***1% confidence levels.

We further illustrate the impact of hiring additional non-HFT or HFT market makers over time in Figure 3.2, where we apply the results for the difference-in-difference model using a control group of 20 nearest neighbour matches to a hypothetical ETF with an at-treatment liquidity cost equal to the sample average of 51.01 XLM basis points. The shaded area on the left-hand side represents the time before treatment. As one would expect, no significant effect on liquidity is measurable during that period, neither with respect to hiring an additional market maker nor with respect to the relative impact of different types of DMMs. One month after treatment, however, a steep liquidity cost reduction is observed for both kinds of market makers. Although varying in absolute terms, the liquidity cost is significantly lower than the pre-treatment levels at any point in time after treatment. In both cases, the liquidity gains appear to stabilise approximately six months after hiring the DMM and they do not appear to change substantially after that.

3.3.4 DISCUSSION OF FINDINGS

IMPLICATIONS OF FINDINGS

The results presented in Section 4.3.3 suggest that, regardless of whether we look at non-HFT or HFT market makers, mandating an additional designated sponsor for equity ETFs is beneficial to market quality, as it significantly reduces the liquidity cost. The empirical evidence that we present is in line with the vast majority of the literature on market making presented in this paper. It confirms Hengelbrock's (2008) findings in particular; that study also provides insights into the effect of additional market makers on XETRA and the impact of different types of market making on liquidity.

The observed treatment effects in our models are both highly significant and in most cases economically relevant. Using the mean treatment effect 12 months after hiring the DMM and the average annual XETRA trading volume for all treated equity ETFs over the sample period, we estimate the annual liquidity cost reductions per ETF from adding one non-HFT or HFT market maker to be € 70,000 and € 238,000 respectively. Applying these figures to the aggregated average annual trading volumes of all ETFs, the total annual liquidity cost reduction that could hypothetically be passed on to the market and investors from the observed 452 treatment instances comes to approximately € 90 million. While these estimates are hypothetical, they still corroborate the view that additional DMMs generate economically substantial liquidity gains. Since hiring a DMM is costly, the finding of economically substantial liquidity gains is of particular importance. The ETF provider has to decide whether the additional liquidity gain and the resulting higher attractiveness of its ETF to the market are worth the financial obligations that come with hiring an additional designated sponsor. Given that most ETFs on XETRA currently rely on the minimum required number of one DMM, there appears to be a considerable potential for liquidity cost reductions.

The decrease in liquidity after hiring an additional market maker might seem intuitive at first. Yet, given the fact that we look at the impact of *additional* market makers, our results suggest that the designated sponsors in place prior to treatment do not reduce the liquidity cost to the lowest possible levels. Otherwise, the additional market maker should not have a significant effect on the ETF's liquidity cost structure. In line with Hengelbrock's (2008) findings, we postulate that, by adding another market maker to the pool, the ETF provider increases competition between individual DMMs, which then leads to decreased spreads and increased liquidity. Faced with decreasing spreads on which to earn a profit, liquidity providers (including the voluntary market makers who cannot rely on a fixed compensation) should be incentivised to further increase trading volume. Based on Anand et al.'s (2009) reasoning that "liquidity begets liquidity" (p. 1447), one should also mention the potential indirect effects associated with hiring an additional designated sponsor, which could cause liquidity to decrease even further. For instance, a new market maker might open up new sales channels, enabling the ETF to access previously unattended investor segments. In turn, these additional fund-flows could have an indirect effect on the liquidity cost via increased trading volume. However, for the observed sample we were unable to find any evidence for significant changes in AuM or trading volume one year after hiring the additional DMM (see Appendix E). Therefore, we conclude that the liquidity cost reduction is mainly generated by the previously postulated inter-DMM competition effect. The finding of unchanged AuM also suggests that, despite the improved liquidity that can be attributed to contracting the DMM, investors do not appear to honour the higher market quality, and ETF providers should not count on benefiting from increased funds and fees, at least not in the first year.

The results presented in Table 3.4 confirm that market makers relying on algorithmic or high-frequency techniques generate significantly higher liquidity cost reductions than non-HFT market makers, at least after a certain amount of time. Three months after hiring the new designated liquidity provider, a HFT DMM outperforms its non-HFT counterpart by some 4.80 XLM basis points[22] on average. The outperformance increases until six months after treatment and stabilises at approximately 5.46 XLM basis points. Given that the mean difference in liquidity gains is only significant after two or three months into treatment, it appears that the HFT market maker needs some time to materialise its systemic advantage over non-HFT market makers. Still, our results confirm the existing empirical evidence on the superiority of HFT market making presented in the reviewed literature.

ROBUSTNESS OF MODEL AND MATCHING QUALITY

Our model does not report significant treatment effects on ETF liquidity at any time before the hiring of the additional market maker, with one exception in one model for non-HFT market makers two months prior to treatment. We consider this outlier to be caused by one or a few extreme liquidity cost values in the control group. Given that when the size of the control group changes the significance vanishes again, the overall picture still supports the idea that the model design does not wrongfully attribute liquidity effects to the hiring of an additional DMM.

In order to determine the quality of our matching procedure and the balance between the treatment group and the matched control groups, we use various indicators, which are presented in Table 3.5 and Table 3.6. First, we calculate the pseudo-R^2 from the propensity scores on all matching covariates on the matched samples. In general, the pseudo-R^2 is intended to indicate how well the covariates explain the probability of treatment. We re-estimate the propensity score only on the treatment group and the matched control group observations. Since the aim of the matching procedure is to remove any systematic difference in the distribution of covariates between the treatment group and the matched control group, we expect the re-estimated pseudo-R^2 of the treatment and the matched control group observations to be fairly low

---

22 Average over all seven models at *t+3*.

(Caliendo & Kopeinig, 2005). As can be inferred from the data in Table 3.5 and Table 3.6, the pseudo-R^2 values are very small (less than 1% for all our models), regardless of the type of market making or the chosen size of the control group.

Table 3.5: Matching quality for non-HFT DMM sample

Control Group	Pseudo-R^2	Max. Bias	Mean Bias	Median Bias	Rubin's B	Rubin's R
M = 5	0.007	3.612	2.828	2.814	20.107	0.637
M = 10	0.005	3.809	2.340	2.426	16.459	0.870
M = 15	0.003	3.010	1.913	1.702	13.833	0.936
M = 20	0.003	3.440	1.835	1.605	12.442	0.910
M = 25	0.002	2.219	1.422	1.350	10.469	0.958
M = 30	0.002	3.311	1.494	0.976	10.711	0.841
M = 50	0.002	3.354	1.406	1.256	10.800	0.558

This table reports indicators for the quality of the matching algorithm of the non-HFT difference-in-difference models for various control group sizes (M). These are the pseudo-R^2, the maximum standardised percentage bias found in individual key matching covariates used in the models, the mean and median standardised percentage bias over all of the matching covariates used in the models, Rubin's B and Rubin's R. For a matching algorithm to be sufficiently balanced, pseudo-R^2 should be low, the standardised percentage bias should be below 5%, Rubin's B should be below 25 and Rubin's R should be between 0.5 and 2.0.

Table 3.6: Matching quality for HFT DMM sample

Control Group	Pseudo-R^2	Max. Bias	Mean Bias	Median Bias	Rubin's B	Rubin's R
M = 5	0.004	4.933	1.981	1.300	15.648	1.013
M = 10	0.004	6.357	2.017	1.233	14.176	0.782
M = 15	0.003	6.638	1.603	1.167	13.132	0.720
M = 20	0.003	5.555	1.891	1.867	13.444	0.658
M = 25	0.003	5.684	1.804	1.318	13.467	0.706
M = 30	0.003	6.017	1.853	1.512	13.681	0.706
M = 50	0.003	5.488	1.841	1.216	13.234	0.889

This table reports indicators for the quality of the matching algorithm of the HFT difference-in-difference models for various control group sizes (M). These are the pseudo-R^2, the maximum standardised percentage bias found in individual key matching covariates used in the models, the mean and median standardised percentage bias over all matching covariates used in the models, Rubin's B and Rubin's R. For a matching algorithm to be sufficiently balanced, pseudo-R^2 should be low, the standardised percentage bias should be below 5%, Rubin's B should be below 25 and Rubin's R should be between 0.5 and 2.0.

According to Rosenbaum and Rubin (1985), the standardised percentage bias is another indicator used to assess the distance in marginal distributions of the matching covariates. It is the percentage difference of the sample means in the treatment group and the matched control group as a percentage of the square root of the average of the sample variances in both groups. Most empirical studies view a bias below 3% or 5% to be sufficient to consider the control group as being bias-free (cf. Caliendo & Kopeinig, 2005; Solivas et al., 2007). With mean and median biases for our matching algorithms all clearly below 3%, the matched control groups in our non-HFT- and HFT-models do not appear to exhibit any substantial bias (see Table 3.5 and Table 3.6). For non-HFT-matching-models, the observed maximum values of 2.22% to 3.81% standardised bias, in most cases, are driven by the dummy variable for the style of replication. For the HFT-matching-models the maximum bias of individual covariates is slightly higher ranging from 4.93% to 6.64%. This is mainly driven by the differences in AuM. Given that they are only moderately higher than the suggested 3% to 5% threshold, and given that the overall mean/median covariate-bias of all models is evidently below 3%, we consider the matching of our control groups to be sufficiently robust.

We also report Rubin's (2001) B and R for the non-HFT and the HFT difference-in-difference models in Table 3.5 and Table 3.6 respectively. Rubin's B is the absolute standardised difference of the means of the linear index of the propensity score in the treatment group and the matched control group, and it should be less than 25. Rubin's R is the ratio of the variances of the treatment group and corresponding matched control groups. For the two groups to be considered sufficiently balanced, Rubin's R should, ideally, lie between 0.5 and 2.0. Both of Rubin's (2001) indicators are within the proposed boundaries. Taking all these indicators together, and considering that the few outliers are somewhat weakened by other indicators supporting the good balance of these models, our treatment observations and their respective matched control groups appear to be sufficiently balanced.

In addition to the balance between the treatment group and the matched control groups, we also examine the distance between the propensity score of a treatment observation and the propensity scores of its respective matched control group observations. Put differently, we determine whether the treatment and matched control observations have similar probabilities to hire an additional DMM. The results presented in Table 3.7 and Table 3.8 support the overall view of a good matching quality of our difference-in-difference models. While the maximum distances measured suggest that there are some extreme outliers in all models, the mean matching distance is always clearly below 1%, with the one exception in the model M=50 for HFT market makers, where it is 1.3%. Furthermore, even when the size of the control groups increases, the 95th and 99th percentile figures are sufficiently close to the respective treatment propensity scores.

Overall, our robustness and quality checks of the model, the matching balance and distance in particular, provide evidence that our models are robust enough to make valid inferences. In addition to the robustness checks, we also address the issue of incorrectly estimating the standard errors raised by Abadie and Imbens (2012), and we estimate the robust standard errors accordingly.

Table 3.7: Matching distance for non-HFT DMM sample

Control Group	Mean Distance	Median Distance	Distance 95th Percentile	Distance 99th Percentile	Maximum Distance
M = 5	0.001	0.001	0.003	0.007	0.132
M = 10	0.002	0.001	0.005	0.012	0.147
M = 15	0.002	0.002	0.007	0.017	0.160
M = 20	0.003	0.002	0.010	0.022	0.165
M = 25	0.004	0.003	0.012	0.026	0.172
M = 30	0.005	0.003	0.014	0.030	0.182
M = 50	0.008	0.006	0.020	0.046	0.231

This table reports the matching distance in terms of the propensity score between the treatment observations of hiring an additional non-HFT market maker and their respective matched control groups (M = 5, 10, 15 20, 25, 30, 50). It reports the mean, median, and maximum distance as well as the 95th and 99th percentile of observed distances.

Table 3.8: Matching distance for HFT DMM sample

Control Group	Mean Distance	Median Distance	Distance 95th Percentile	Distance 99th Percentile	Maximum Distance
M = 5	0.001	0.001	0.003	0.008	0.156
M = 10	0.002	0.002	0.006	0.014	0.184
M = 15	0.004	0.003	0.010	0.019	0.226
M = 20	0.005	0.004	0.011	0.024	0.245
M = 25	0.006	0.005	0.013	0.028	0.263
M = 30	0.007	0.006	0.017	0.032	0.294
M = 50	0.013	0.011	0.030	0.045	0.344

This table reports the matching distance in terms of the propensity score between the treatment observations of hiring an additional HFT market maker and their respective matched control groups (M = 5, 10, 15 20, 25, 30, 50). It reports the mean, median and maximum distance as well as the 95th and 99th percentile of observed distances.

Because we only identify those market makers as algorithmic or high-frequency traders who predominantly and explicitly rely on those techniques, by design our sample of HFT market makers excludes hybrid HFT market makers, such as high-frequency trading desks from broker-dealers and large integrated banks. The categorisation of some important HFT players as non-HFT might limit the inferences that can be made from our sample, as it affects the estimate of their differences (cf. Biais & Foucault, 2014; Brogaard et al., 2014; Yao & Ye, 2014). Since we detect statistically significant and more pronounced results for the pure HFT group than for the potentially biased non-HFT sample, the biased non-HFT sample would lead to overestimating the true treatment effect of non-HFT market makers. This in turn would mean that the outperformance of HFT might be larger after all.

3.4 CONCLUSION

Recently, a shift has occurred in ETF market making away from the predominant arrangement of solely relying on in-house resources for market making towards outsourcing this capability to external specialists. Yet, the literature on the actual effect that market makers have on market quality is rather scarce, at least when it comes to ETFs. We address this gap by determining whether hiring an additional designated sponsor significantly affects the liquidity cost of an ETF on Deutsche Börse's XETRA-trading-platform. Moreover, we try to determine whether DMMs that explicitly and primarily rely on algorithmic and high-frequency trading techniques are systematically better in providing liquidity to an ETF than other types of market makers.

In summary, we find the treatment effects of hiring an additional non-HFT or HFT market maker to be highly significant and economically substantial with observed liquidity cost reductions adding up to an average € 70,000 to € 238,000 per year, depending on the type of market maker. We argue that competition between market makers is the primary reason for the decrease in the liquidity cost. Furthermore, we provide evidence that HFT designated sponsors cause significantly greater reductions in liquidity cost than respective non-HFT market makers.

With these findings, our paper contributes to the understanding of how DMMs drive fund liquidity costs and the extent to which the differences in their setups might affect their ability to generate liquidity. We have discussed the potential impact of indirect effects on liquidity over time. However, determining the magnitude and significance of those indirect effects in the aftermath of hiring an additional market maker might be a meaningful venue for future research.

4. ETF FLOWS AND UNDERLYING STOCK RETURNS: THE TRUE COST OF NAV-BASED TRADING

ABSTRACT

In this paper, we examine whether the creation or redemption of ETF shares has a measurable and significant effect on the underlying stocks' returns in the closing auction. Our findings show that ETF flow-related stock transactions significantly affect stock prices. We provide empirical evidence showing that creations/redemptions of ETFs that replicate indices in the German DAX index family have a highly significant and economically viable effect on abnormal returns of underlying stocks in the closing auction an effect that is particularly pronounced in small stocks and on bullish trading days. We argue that, given the sizable additional annual earnings of up to € 53,000 per stock, the AP has a motivation to exploit this inefficiency, by means of active price manipulation during the closing auction. Hence, dealing in ETF shares on the primary market, for example through NAV-based orders, might entail hidden costs for investors that, until now, have not been recognised in the literature and perhaps not even by investment professionals.

***Keywords*:** ETF, Exchange-Traded Fund, Fund Flows, Trading Volume, Asset Basket, Asset Pricing

***JEL Classification*:** G12, G14, G15, G23

***Authors*:** Friedrich Osterhoff, Maximilian Overkott

First Author: Friedrich Osterhoff

***Current Status*:** Working Paper

4.1 INTRODUCTION

Exchange-traded funds (ETFs) are widely acknowledged to be one of the most successful financial innovations of the past few decades, a perception that is supported by the substantial growth of assets under management (AuM) over the past several years, amounting to US$ 2.81 trillion globally as of December 2015 (Deutsche Bank Research, 2016), and the growing share of ETF-related transactions in total trading volume. With ETFs playing such an increasingly vital role in capital markets, a thorough understanding of how they might affect underlying securities, and, in that way also markets as a whole, is of significant interest to both investment professionals and financial researchers.

The effect that ETFs have on the quality of underlying securities is not entirely new to theoretical or empirical financial research. In that field, existing literature on ETFs often focusses on a fund's inception and its effect on stock liquidity (e.g. Hedge & McDermott, 2004; Yu, 2005; Richie & Madura, 2007; De Winne et al., 2013), volatility (e.g. Lin & Chiang, 2005; Ben-David et al., 2014), or long-term valuation (e.g. Madura & Ngo, 2008; Bae et al., 2013). Other studies also confirm the effect of fund flows on stock prices (e.g. Edelen & Warner, 2001; Yu, 2005; Coval & Stafford, 2007; Jotikasthira et al., 2012). A growing yet still relatively small fraction of studies acknowledge the structural differences between ETFs and mutual funds, the in-kind creation/redemption of ETF shares in particular; thus, there is a need for research that clearly and specifically focuses on ETFs in order to more fully understand how their fund flows might affect underlying securities (e.g. Kalaycıoğlu, 2004; Staer, 2014).

This present study of ETFs replicating German equity indices in Deutsche Börse's DAX index family and their relation with and impact on their underlying stocks aims to extend the research that recognises ETFs not as a sub-class of mutual funds but as a separate class of assets. We also seek to shed new light on the relationship between ETF flow-related trading volume and stock returns. In contrast to previous studies that describe the effect of ETFs on their respective underlying stocks at an aggregated index- or market-level, we emphasise the need to identify the different levels of influence depending on the size and liquidity of a stock. Toward that end, we investigate ETF flow effects on returns at the most granular level possible, namely that of individual stocks. The variable in our approach is the abnormal stock return in the closing auction in the German stock market. By using a controlled environment, such as the closing auction, it is less likely that general market movements or news would drive abnormal returns in our sample.

Our findings show that ETF flow-related stock transactions significantly affect stock prices. We provide empirical evidence showing that creations/redemptions of ETFs replicating indices in the DAX index family have a highly significant and economically viable effect on abnormal returns of underlying stocks in the closing auction. The effect is particularly pronounced in small stocks and on trading days that are generally bullish. We argue that the authorised participant (AP), might be able to exploit this inefficiency, and that, given the sizable additional earnings (according to our models), active price manipulation might be an attractive option for him/her. Dealing ETF shares, for example through net asset value (NAV)-based orders, might entail hidden costs that, until now, have not been recognised in the literature and perhaps not even by investment professionals.

The fact that these effects are measurable and robust for one of the world's largest and presumably most efficient stock markets increases concerns about the even greater impact they could have in less-developed markets and the hidden costs for investors, which could be sizeable.

The remainder of this paper is structured as follows. Section 4.2 provides an overview of the relevant literature with a focus on the correlation between fund flows and stock prices. Section 4.3 presents the empirical components of the study. Section 4.3.1 provides information on the institutional background. Section 4.3.2 describes the data and research design, Section 4.3.3 presents the results and discusses the implications of our findings and the robustness of our models. In Section 4.4, we provide our concluding remarks.

4.2 LITERATURE REVIEW

4.2.1 EFFECTS OF ETFS ON STOCK LIQUIDITY, VOLATILITY AND VALUATION

The potential effects that ETFs have on their underlying securities have been examined extensively by various studies (cf. Hedge & McDermott, 2004; Yu, 2005; Van Ness et al., 2005; Richie & Madura, 2007; De Winne et al., 2013).[23] Several papers focus on the overall effects of ETF inception on stock characteristics; for example, Hedge and McDermott (2004) find that the stock liquidity in the Dow Jones Industrial Average and NASDAQ 100 indices significantly improves after the introduction of ETFs replicating these indices. Adapting the methodology, Richie and Madura (2007) come to similar conclusions for the NASDAQ 100 index. Moreover, they show that liquidity improvement is more pronounced for index constituents with lower weights. Yu's (2005) analysis of ETF introductions in the US market also corroborates the belief that the market liquidity of component stocks improves after ETFs start trading. Confirming the aforementioned US findings for France, De Winne et al. (2013) examine how an ETF replicating the French CAC 40 index affects the liquidity of the underlying stocks. They find that index stock spreads decline in comparison to non-index stocks after the introduction of ETFs. Ben-David et al. (2014) find a positive relation between ETF ownership and the volatility of the underlying securities in the US market with a one standard deviation change in ETF ownership being associated with a 19% increase in intraday volatility. They argue that the effect is related to arbitrage activity between ETFs and their underlying stocks. Their results support the view that ETFs attract an additional layer of demand shocks into the prices of the underlying securities through arbitrage. They note that the evidence suggests that ETF ownership in stocks increases the noise in the price of those stocks. Lin and Chiang (2005) also find a significant increase in stock volatility, regardless of firm size. They examine the effect that the introduction of the TTT ETF had on the volatility of Taiwan 50 index component stocks. For a sample of international ETFs listed on the American Stock Exchange,

---

[23] For a concise overview of the academic literature on ETFs, see Charupat and Miu (2013).

Madura and Ngo (2008) find positive and significant valuation effects on the dominant component stocks in response to the inception of a new ETF, and they attribute the variation in the valuation effects to the size of the ETF as well as stock-specific characteristics, such as liquidity. These stock-specific characteristics also appear to lead to a more pronounced increase in stock trading volume following the ETF inception. Bae et al. (2013) investigate the effect of ETFs on underlying stock characteristics; however, contrary to Madura and Ngo (2008), they find a negative effect on valuation but a positive effect on systematic volatility, the size of observed short holdings and liquidity, particularly for small stocks.

4.2.2 FLOW-INDUCED TRADING OF STOCKS (AROUND THE CLOSE)

Fund-related transactions, rather than the presence of a new fund, have been the focus of another growing stream of research that examines the correlation between fund flows and market returns in particular. Yu (2005) analyses how the trading of ETFs affects stock prices by observing the informational efficiency of the underlying securities, and he shows that ETF markets provide a substantial amount of the information that is incorporated into the efficient price of an underlying stock. His results suggest that ETF trades have a permanent impact on component stock prices. Coval and Stafford (2007) and Jotikasthira et al. (2012) investigate the effect of asset fire sales and purchases on equity returns, and they find that fund flows and flow-related block trades have a significant impact on stock prices. Using daily equity data for the US markets, Edelen and Warner (2001) show that fund flows and institutional trading positively affect market returns. Kalaycıoğlu (2004) investigates the relationship between returns and ETF fund flows, and, using a sample of five ETFs covering major US equity indices for different time frequencies, he finds a significantly negative correlation on a monthly basis, yet positive relations on a weekly and daily basis. He concludes that the obtained evidence has different implications on the tested correlation in different frequencies, and that the results do not, outright, support the price pressure hypothesis. In addition, Staer (2014) shows a strong positive relation between daily ETF flows and the returns on the underlying stocks, suggesting a price pressure effect related to the flow activity. In line with these results, Comerton-Forde and Putniņš (2011) and Bacidore et al. (2012) acknowledge fund-related NAV-trading as a potential source of abnormal returns around market close. The latter argue that one possible motive is to capture some of the price impact of the traders' own transactions, not necessarily for profit maximisation but, rather, as a way to minimise losses.

Empirical evidence suggests that stock prices can be affected by individual trades even during very short isolated trading periods during a day, such as call auctions. A considerable number of studies examine premeditated price manipulation as one source of abnormal stock returns over short periods of time. Aggarwal and Wu (2006), Comerton-Forde and Rydge (2006) and Comerton-Forde and Putniņš (2011; 2014) all corroborate the occurrence of stock price manipulation, particularly around the close for the US and Canadian markets. Aggarwal and Wu (2006) argue that the manipulation is primarily concentrated in small stocks. Hillion and Suominen (2004) note that the effect can be reduced by the introduction of a call auction at close. With investors being particularly active during the last five minutes of trading, the adoption of such an auction at the Paris Bourse led to a reduction in market manipulation and the establishment of a more efficient market structure at close. Comerton-Forde et al. (2007) and Chang et al. (2008) investigate the adoption of a call market method to open and close the market in Singapore. Consistent with prior research, they conclude that day-end price manipulations are reduced. Pinfold and He (2012) confirm these insights for the New Zealand stock exchange. Küçükkocaoğlu (2008) and Kadıoğlu et al. (2015) examine the possibility of stock price manipulation around the close at the Istanbul Stock Exchange, and both studies suggest that close end price manipulation through big buyers and big sellers is possible, even though the implementation of closing call auction sessions resulted in a significant reduction in closing price manipulation. Testing for closing price manipulation in the Finnish stock market, Felixson and Pelli (1999) provide evidence that block trades and spread trades explain part, but not all, of the existing abnormal returns. For a sample of Taiwanese stocks, Huang and Chan

(2014) show that large individual investors are pushing closing prices upwards, even after the introduction of a closing call auction. Suen and Wan (2013) use intraday data from the Hang Seng index and provide evidence for abnormally large orders and price changes during the last five seconds of the auction sessions. In contrast to most other studies, their results suggest that while prices, on average, are more efficient in a closing auction procedure than in a random closing procedure, they are also more prone to manipulation attempts.

Summarising the strand of research concerned with flow-induced price manipulation, it seems evident that the last minutes of trading are, on average, the most important, and the possibility of manipulating prices in either direction exists. Yet, this chance is perceptibly reduced by the adoption of a call auction at market close. However, as we will show later on, flow-induced price pressure can be strong enough to have a significant effect on end-day returns as well as during the final auction.

4.2.3 CURRENT STATE OF RESEARCH AND POTENTIAL RESEARCH GAPS

To date, the reviewed literature suggests that although ETFs are derivative products of securities baskets, they can have a significant and substantial effect on their underlying portfolio component stocks. The inception of a new ETF seems to result in long-lasting changes in stock characteristics, such as liquidity, volatility and valuation. It also appears that flow-related transactions by institutional investors have the potential to affect stock prices, especially during the last few minutes of trading – even with a call auction system in place that is meant to reduce active manipulation.

The previous section made it clear that the literature has extensively addressed the relationship between mutual fund flows and underlying asset returns. However, with the exceptions of Kalaycıoğlu (2004) and Staer (2014), the relationship between ETF flows and their underlying stocks has received far less attention. Staer (2014) describes the structural differences between mutual funds and ETFs, the in-kind creation/redemption process in particular, which make it impossible to assume that mutual fund cash flows and ETF (in-kind) flows have the same effect on stock prices. Thus, we contribute to a growing body of research that recognises ETFs not as a sub-class of mutual funds but as a separate asset class. In this context, our paper also contributes to the existing research gap by explicitly investigating the effects that ETF flows have on single stock returns. This fine granularity is necessary in order to identify the different levels of influence that ETFs have depending on the size and liquidity of a stock. A more general analysis of aggregated index returns, as conducted by Staer (2014), has no chance of providing such a detailed picture. Our approach uses the closing auction abnormal return as the outcome variable because in a more controlled environment general market movements or news are unlikely to be the drivers of abnormal returns. In this way, we also contribute to the research on price effects of trading around the close. While most other papers focus on prosecuted cases of manipulation to determine the effect on stock prices (e.g. Comerton-Forde & Putniņš, 2011), we consider all forms of last minute trading, thereby allowing for a broader and less biased view of the effects of trading around the close.

4.3 EMPIRICAL PART

In this paper, we attempt to determine whether the creation/redemption of ETF shares has a measurable and significant effect on returns of the underlying stocks in the closing auction. We postulate that there is a positive (negative) relationship between creations (redemptions) and abnormal returns. That is, net creations of ETF shares should lead to abnormal positive returns in the respective underlying stocks, whereas net redemptions should result in abnormal negative returns. Before we describe our data and methodology, we first provide a broad overview of the institutional background of XETRA, its closing call auction system and the mechanisms of NAV-based trading.

4.3.1 Institutional Background

The XETRA Electronic Trading System

In November 1997, Deutsche Börse implemented XETRA, its fully electronic trading system with a centralised limit order for equity, ETFs, exchange-traded products, mutual funds, certificates, options, bonds and rights issues. Matching of buy and sell orders takes place according to a price-time priority rule in which visible orders have priority over hidden orders. Trading is anonymous for all participants, and trades are processed through a central counterparty (CCP). Continuous trading takes place from 9:00 am until 5:30 pm and is framed by an opening and closing call auction starting at 8:50 am and 5:30 pm respectively. The closing auction for equities can be divided into a call phase, a price determination phase and an order book balancing phase (Deutsche Börse, 2014). During the call phase, market participants can either enter orders and quotes or modify and delete their existing orders and quotes. In order to avoid price manipulation, the call phase has a random end after a minimum period of time. In the next phase, the auction price is determined based on the principle of the most executable volume with regard to the order book situation at the end of the call phase.

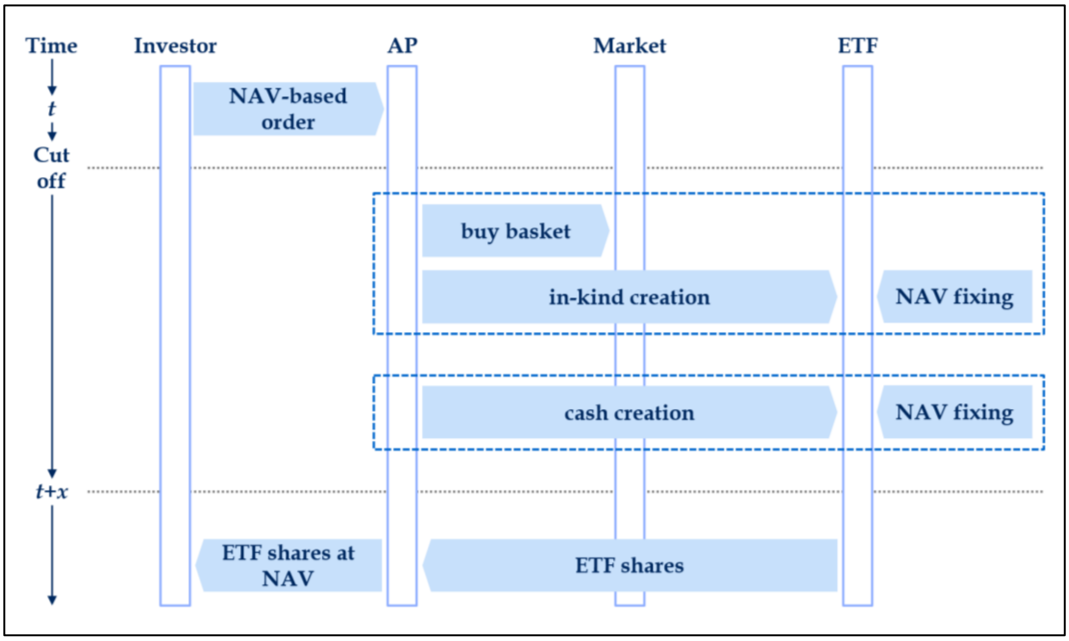
NAV-Based Trading[24]

An important difference between mutual funds and ETFs is that the latter can be traded by market participants throughout the trading day. Generally, investors have three different options for trading shares: they can trade through an exchange, directly OTC counterparty or through NAV-based trading. However, intraday trading of ETF shares via an exchange entails various transaction costs, such as investment or trading costs (e.g. fees, commissions, spreads), which are subject to change throughout the trading day. In contrast, NAV-based trades offer a seemingly more transparent way of trading ETF shares. The investor trades the shares directly with the AP at net asset

[24] This section is loosely based on information presented in a paper authored by Kim (2014).

value, that is, at the closing price of that day. Any such deal has to take place before a cut-off time determined by the ETF provider. For NAV-based trades, the AP utilises the creation/redemption process to generate or redeem the necessary lot of shares (see Figure 4.1 for details about ETF share creation).

Figure 4.1: NAV-based order process and the resulting creation of ETF shares



This chart lays out the NAV-based order process and the resulting creation of ETF shares. The investor places an NAV-based order with the AP before the cut-off at trading day *t*. The AP then buys the basket of underlying shares in the market based on the size of the original order and starts the creation process to transfer the basket to the ETF in return for shares. The exact exchange ratio is determined by the NAV at the end of trading day *t*. Once the AP receives the ETF shares at *t+x* (commonly *t+1* or *t+2* depending on the settlement cycle) they are forwarded to the investor at NAV. (Source: Kim, 2014).

The AP has two options to fulfil the order: a creation/redemption in cash or in-kind. For a cash creation/redemption, the AP delivers cash to the ETF management in return for a corresponding number of ETF shares. The management then allocates the money amongst its holdings, thereby investing the money in the respective underlying stocks. For the in-kind creation/redemption, the AP exchanges the shares for a corresponding basket of underlying securities. Since APs will receive (pay) the NAV from (to) the investor, they have an incentive to pay (receive) closing prices for the basket securities in the cash market. The final price for the transaction can only be determined when the

NAV is published, which usually takes place one day after the deal has been arranged. At least in the case of an in-kind creation/redemption, APs may theoretically try to actively exploit the fact that their actions on the trading floor affect the NAV and manipulate the closing price to make a profit by buying (or selling) the index constituents throughout the trading day at a price that is relatively lower (or higher) than the closing price. For example, by placing an unlimited order in the closing auction, the AP can try to drive the price up (or down) in the last minutes of trading.

4.3.2 DATA AND METHODOLOGY

Our research design is based on observing the effect of share creations/redemptions on stock returns in a controlled and isolated environment. In our study, this controlled environment is the closing auction that concludes each stock's trading day. One reason why we only observe this short period of time, and not the entire trading day, is that we expect abnormal returns to be less biased by other longer-lasting factors that might affect asset pricing throughout the day. Another advantage of examining an isolated period of time is that it enables us to avoid dealing with the issue of reverse causality that other studies are facing (cf. Kalaycıoğlu, 2004). When the closing auction starts at 5:30 pm, all cut-off times for NAV-based ETF trades have long passed, and, so too, has any opportunity for a new creation/redemption for that particular day. By focussing on the last five minutes of trading, we can safely neglect the possibility that abnormal stock returns could cause a creation/redemption of ETF shares; hence, we can assume that the former affects the latter and not the other way around. Furthermore, Cushing and Madhavan (2001) also show that the last five minutes of the trading day explain a disproportionate fraction of the variation in daily returns, which is another reason to scrutinise this short period separately.

DATA

For this study we use both daily data and intraday data of ETFs that replicate indices of the DAX index family and their respective underlying constituent stocks for the period January 2011 to December 2014. Eventually, we use the daily fund flows of 14 physically replicating ETFs[25] that mimic seven different DAX family indices, which we identify via the Morningstar Fund Database, to model the daily flow-related trading volumes for 137 German stocks in the sample period. Daily net creation/redemption volume for each of the 14 ETFs is compiled from Bloomberg. For each underlying stock we obtain the following data from Deutsche Börse: the daily trading volume in euros, free float market capitalisation in euros, the ETF's respective weight in the replicated indices and its liquidity cost, as expressed by Deutsche Börse's XETRA Liquidity Measure (XLM).[26] We collect intraday tick data for each of the DAX-universe component stocks and for the MSCI EMU ex Germany index from the Karlsruhe Capital Market Database (KKMDB) and Thomson Reuters respectively. This provides us with a total of 62,054 observations for our main regression models.

METHODOLOGY

In a first step, we calculate the closing auction return $r_{i,t}^{C}$ for day t for each stock i as the relative difference between the stock's closing price ($PX\_Close_{i,t}$) and the last price of continuous trading ($PX\_LastCont_{i,t}$), that is, the last observed intraday tick price before the closing auction commences at 5:30 pm. Hence, $r_{i,t}^{C}$ can be expressed as follows:

$$r_{i,t}^{C} = \frac{PX\_Close_{i,t}}{PX\_LastCont_{i,t}} - 1 \tag{4.1}$$

25 Please see Appendix F for a list of all ETFs in the sample.

26 XLM measures the order-size dependent liquidity costs of a round-trip for hypothetical transaction volumes of XETRA listed stocks, taking the entire depth of the limit order book into account (cf. Hachmeister, 2007; Krogmann, 2011; Stange & Kaserer, 2011; Hendershott & Riordan, 2013; Rösch & Kaserer, 2013).

We then calculate daily abnormal returns for the closing auction period by applying the market model proposed by MacKinlay (1997), as follows:

$$r_{i,t,h} = \beta_{i,BMn,t} \cdot r_{BMn,t,h} + \varepsilon_{i,t,h} \tag{4.2}$$

where $r_{i,t,h}$ is the return of stock i on trading day t during trading hour h and $r_{BMn,t,h}$ is the return of the MSCI EMU ex Germany index for the corresponding period of time during trading day t. The MSCI EMU ex Germany index covers the entire Eurozone equity market with the exception of Germany. By choosing an index that is broad enough to entail all relevant information in the European stock markets but does not comprise any of the sample stocks, we ensure that the flow-related transactions in the sample do not affect the market return or our market model. $\beta_{i,BMn,t}$ is estimated by means of rolling regressions of hourly returns over the last 20 trading days with heteroskedastic and autocorrelation-consistent (HAC) variance estimates. We exclude observations with insignificant estimates from our sample. The resulting distribution of estimates is truncated at the 0.5th percentile and the 99.5th percentile.[27] In the next step, abnormal returns during the closing auction $r_{i,t}^{C*}$ are calculated as follows:

$$r_{i,t}^{C*} = r_{i,t}^{C} - \left(\beta_{i,BMn,t} \cdot r_{BMn,t}^{C}\right) \tag{4.3}$$

We then perform panel regressions, which account for the stock and time fixed effects, in order to control for inherent characteristics. In addition to net creation/redemption,

---

[27] Overall, a total of 2,889 valid observations are removed from our sample by means of exclusion due to insignificance or percentile truncation. Please see Appendix G for an overview of the final distribution of betas in our sample.

we identify the effects that trading volume, free float market capitalisation and stock liquidity cost have on abnormal returns:[28]

$$r_{i,t}^{C*} = (\alpha + v_i + \tau_t) + \beta_1 \cdot CR_{i,t} + \beta_2 \cdot Volume_{i,t} + \beta_3 \cdot FreeFloat_{i,t-1} + \beta_4 \cdot XLM_{i,t} + \varepsilon_{i,t} \tag{4.4}$$

Here v_i and τ_t are the stock and trading day fixed effects respectively. $CR_{i,t}$ is the standardised daily net creation/redemption-induced transaction volume of constituent stock i relative to the free float market cap of that stock. The variable is calculated as follows:

$$CR_{i,t} = \frac{\sum_{k=1}^{n}(CR_{k,t} \cdot Weight_{k,i,t})}{FreeFloat_{i,t-1}} \tag{4.5}$$

where $CR_{k,t}$ is the sum of the net creation/redemption volumes in millions of euros from all sample ETFs replicating the index k that contain stock i as constituent, $Weight_{k,i,t}$ is stock i's weight in index k on day t, $FreeFloat_{i,t-1}$ is the free float market capitalisation of stock i in millions of euros at the beginning of trading at day t and $CR_{i,t}$ is the sum of the weighted creation/redemption volumes from the n indices replicated by at least one of the sample ETFs that contain stock i, divided by $FreeFloat_{i,t-1}$. The variable represents the share of total free float equity that is bound in creation/redemption-induced transactions on a particular trading day. We adjust our creation/redemption dataset in order to account for reporting inconsistencies on *t+1* as discussed by Rakowski and Wang (2009) as well as for the *t*+2 settlement cycle on German stock exchanges. We do so by aggregating all creations and redemptions

[28] In order to ensure the valid statistical inference of our model, we apply HAC robust standard errors that are clustered at the stock level.

in the period between t and t+2.[29] For a better understanding, we standardise the variable $CR_{i,t}$. Staer (2014) correctly notes that APs could already hold the underlying securities; thus, an AP could exchange those securities for the ETF shares without having to buy them on the cash market. However, in order to avoid unnecessary market risk on their books APs usually transact flow-related orders on the market (Abner, 2010 as cit. in Staer, 2014). Therefore, we assume that flow-related transactions are normally executed through market transactions.

Furthermore, we introduce $Volume_{i,t}$, $FreeFloat_{i,t-1}$, and $XLM_{i,t}$ as control variables to our regression model. $Volume_{i,t}$ is the natural logarithm of the overall trading volume for stock i in millions of euros and it accounts for the correlation between overall turnover and trading volume with abnormal returns. $FreeFloat_{i,t-1}$ is the natural logarithm of the free float market capitalisation of stock i available at the beginning of day t. It represents the theoretical total equity volume that APs can use up for their flow-related transactions. The larger the free float of a stock, the less likely it should be that a single transaction can cause price pressure, thereby resulting in a measurable price movement. $XLM_{i,t}$ is the hypothetical round-trip liquidity cost needed to execute the flow-related transaction in stock i (rounded to the nearest thousand euros) expressed in XLM basis points. For that we use daily XLM liquidity cost data on each stock for different hypothetical transaction volumes, and we linearly interpolate and extrapolate to arrive at robust liquidity cost estimates for the actual flow-related transaction volume. Contrary to liquidity proxies, such as the bid-ask spread, XLM takes the entire depth of the order book into account. In this way, it provides a much better picture of the true liquidity cost of executing a transaction with the volume dictated by the underlying creation or redemption of ETF shares.

In order to determine whether stock market capitalisation or economic regimes have an impact on the magnitude of the flow-related effect on stock returns, we introduce dummy variables and interaction terms with $CR_{i,t}$ for small-cap stocks and large-cap

[29] We relax this correcting factor and only consider creations/redemptions on *t* and *t+1* later on in Section 4.3.4 where we discuss model robustness and show that the results are not significantly altered.

stocks and boom and bust phases respectively. For the boom (or bust) dummies, we identify 6,834 (6,347) trading day observations by taking all observations in the sample below and above the 10^{th} percentile and 90^{th} percentile of the daily German Composite DAX index (CDAX) returns respectively:

$$r_{i,t}^{C*} = (\alpha + v_i + \tau_t) + \beta_1 \cdot CR_{i,t} + \beta_2 \cdot Volume_{i,t} + \beta_3 \cdot FreeFloat_{i,t-1} + \beta_4 \cdot XLM_{i,t} + \beta_5 \cdot Boom_t + \beta_6 \cdot CRxBoom_{i,t} + \beta_7 \cdot Bust_t + \beta_8 \cdot CRxBust_{i,t} + \varepsilon_{i,t} \quad (4.6)$$

For the stock-size dummies, we identify the 10% largest and the 10% smallest stock observations, using daily free float market capitalisation as a measure of size:

$$r_{i,t}^{C*} = (\alpha + v_i + \tau_t) + \beta_1 \cdot CR_{i,t} + \beta_2 \cdot Volume_{i,t} + \beta_3 \cdot FreeFloat_{i,t-1} + \beta_4 \cdot XLM_{i,t} + \beta_5 \cdot Big_{i,t} + \beta_6 \cdot CRxBig_{i,t} + \beta_7 \cdot Small_{i,t} + \beta_8 \cdot CRxSmall_{i,t} + \varepsilon_{i,t} \quad (4.7)$$

Moreover, we test all these dummies and interaction terms in an aggregated model, which can be expressed as follows:

$$r_{i,t}^{C*} = (\alpha + v_i + \tau_t) + \beta_1 \cdot CR_{i,t} + \beta_2 \cdot Volume_{i,t} + \beta_3 \cdot FreeFloat_{i,t-1} + \beta_4 \cdot XLM_{i,t} + \beta_5 \cdot Boom_t + \beta_6 \cdot CRxBoom_{i,t} + \beta_7 \cdot Bust_t + \beta_8 \cdot CRxBust_{i,t} + \beta_9 \cdot Big_{i,t} + \beta_{10} \cdot CRxBig_{i,t} + \beta_{11} \cdot Small_{i,t} + \beta_{12} \cdot CRxSmall_{i,t} + \varepsilon_{i,t} \quad (4.8)$$

We postulate that small stocks are more prone to the effect of flow-related trading on the closing auction return, as even small transactions can already have an emptying effect on the market, resulting in price pressure. We further assume that the effect is particularly strong in economically smooth or even upward trending periods, when liquidity is high and liquidity cost is not eating up any profits from this kind of NAV-based transaction during the closing auction. With an average net flow-related transaction per stock of € 240,000 during bullish periods compared to a negative per stock transaction of € 99,000 during bearish periods, we find the net creation/redemption activity in our sample to be substantially higher during bullish periods than during bearish or normal return periods.

4.3.3 Empirical Results

Descriptive statistics

Table 4.1 provides the descriptive statistics on the key variables for the 137 stocks in the full sample. For the observed period from January 2011 to December 2014, the average and median daily abnormal return during the closing auction was approximately 0.7 basis points. The relatively high standard deviation, as well as the high maximum and minimum values,[30] suggest that the daily abnormal returns observed in our sample vary considerably. Creations and redemptions relative to free float market capitalisation appear to be almost balanced over the sample period, with the mean being slightly negative. In our sample there is an overweight of redemptions (i.e. flow-related sell-transactions) that leads to the average of negative 0.1 basis points creation/redemption relative to free float. In absolute terms however, a netted 2.7 basis points of a stock's free float market capitalisation per day are being traded, on average, due to either creations or redemptions of ETF shares. Again, the maximum and minimum values illustrate that large flow-related transactions can comprise more than 1.2% of the total available free float market capitalisation of single stocks on a single trading day. The stock observations in our sample have a median daily trading volume

---

30 The negative minimum abnormal return in the closing auction can be attributed to the stock of Gagfah SA on 5 August 2011, when it lost almost 12% of its value over the day. The positive maximum abnormal return in the closing auction can be attributed to the stock of SGL Carbon SE on 28 February 2011, when the stock price rose by more than 6% over the course of the day.

of € 12.91 million and a median free float of approximately € 2.21 billion. With an observed minimum free float of € 10.15 million and a maximum of € 89.53 billion in our sample, we can assume that our analysis covers a broad spectrum of the German stock market.

Table 4.1: Descriptive statistics for the tested variables in the full sample

	Mean	Median	Std. dev.	Min.	Max.	Skew.	Kurtosis
Abnormal Return	0.007	0.007	0.234	-5.419	4.599	0.370	22.502
CR	-0.001	-0.001	0.056	-1.117	1.208	1.142	68.150
Volume	50.450	12.911	82.296	0.011	1951.649	3.646	30.820
Free Float	9.466	2.209	15.831	0.010	89.533	2.401	8.464
XLM	75.221	46.714	86.395	2.374	1689.955	3.790	28.095

This table reports the descriptive statistics for the 62,054 daily observations used in the regression (4.4). It provides mean, median, standard deviation, minimum and maximum data, as well as skewness and kurtosis for the key variables in the model, namely abnormal return in the closing auction (in %), the creation/redemption-related transaction volume in each stock based on the stock weight in the respective benchmark index relative to free float market capitalisation (in %) for the period t to t+2 (*CR*), trading volume of each stock in billions of euros (*Volume*), the free float market capitalisation of stocks available at the beginning of observation day t in millions of euros (*Free Float*) and the liquidity cost of the respective flow-related stock-transaction rounded to the nearest thousand euros measured in XLM basis points (*XLM*).

Table 4.2 shows the pairwise correlations between the dependent variables and all of the independent variables. The reported correlations between abnormal closing auction return and creation/redemption, free float and XLM respectively, provide a first indication of the relationships between the explanatory variables and abnormal return in the closing auction in our models. ETF creations and small free float appear to be correlated with higher abnormal returns. While no clear pattern was observed for the relationship between trading volume or liquidity cost and abnormal return, the correlation between free float and abnormal returns suggests that size has a substantial effect on returns in the closing auction. With a comparably high correlation between trading volume against free float market capitalisation, we run the risk of highly collinear covariates. Therefore, we orthogonalise the trading volume variable against free float market capitalisation in all our regressions using a modified Gram-Schmidt procedure (Golub & Van Loan, 1996). While the orthogonalisation procedure does not have any significant effect on the coefficients' statistical significance or loadings in any of our models, it substantially reduces multicollinearity, and, consequently, the variance inflation factors.

Table 4.2: Matrix of pairwise correlations between tested variables

	Abnormal Return	CR	Volume	Free Float	XLM
Abnormal Return	1.0000				
CR	0.0243***	1.0000			
Volume	-0.0052	0.0432***	1.0000		
Free Float	-0.0204***	0.0027	0.7555***	1.0000	
XLM	-0.0005	0.0621***	0.1659***	0.1019***	1.0000

This matrix reports the pairwise correlations for abnormal return in the closing auction (in %), the creation/redemption-related transaction volume in each stock based on the stock weight in the respective benchmark index relative to free float market capitalisation (in %) for the period t to t+2 (*CR*), the trading volume of each stock in millions of euros (*Volume*), the free float market capitalisation of the stocks available at the beginning of observation day *t* in millions of euros (*Free Float*), and the liquidity cost of the respective flow-related stock-transaction rounded to the nearest thousand euros measured in XLM basis points (*XLM*). The pairwise correlations are based on 62,054 observations used in the regression (4.4). The statistical significance of the results being different from zero is based on a two-tailed test at the *10%, **5% and ***1% confidence levels.

Following Cushing and Madhavan's (2001) methodology, we also check whether the last minutes of trading have an abnormally high explanatory power for determining the overall daily return of a stock. To accomplish this, we conduct a regression of a stocks' daily return on the closing auction return of that day. The R^2 for each five-minute trading period throughout the day should be approximately 1/102 or 0.98% if the returns across trading sessions contribute equally to the entire day's return. However, with a measured R^2 of roughly 1.62% for the last five minutes of trading, it appears that the closing auction periods in our sample are exceptionally informative about the process of stock pricing.

EFFECT OF FLOW-RELATED TRADING ON CLOSING AUCTION RETURNS

Table 4.3 presents the main results for panel regressions (4.4), (4.6), (4.7) and (4.8). As shown, we find clear and striking evidence that ETF net creation (redemption) has a positive (negative) impact on abnormal closing auction returns. This finding holds even after controlling for trading volume, free float, transaction liquidity cost, economic regime and stock size. In our basic model (4.4), a creation (redemption) of ETF shares that causes a one standard deviation shock to average flow-related stock transactions, which is equivalent to 5.6 basis points of an average stock's free float (or approximately € 5.29 million)[31], results in a positive (negative) abnormal return of almost one basis point respectively. One might argue that this effect is economically irrelevant. Yet, there are at least four reasons why such a statement is inaccurate. First, our sample consists of stocks trading in indices that are widely followed on the German stock market, one of the world's largest and most efficient capital markets, where one would not expect these flow-related transactions to have any impact at all. Second, with ETF AuM and trading volume still growing, abnormal returns that are the result of ETF-related trading can be expected to further increase in the future. Third, as illustrated in Table 4.1, there are extreme outliers in creation/redemption flows. On trading days with extremely high creation/redemption volumes, which are mainly driven by large institutional investors, the impact is dramatically multiplied. Fourth, the measured effect of almost one basis point abnormal return neglects the variety of ways that stocks can be impacted depending on the overall state of the market and the size of the affected stock. We will show that the mean effect on smaller stocks, especially during extremely bullish market periods, also has a high level of economic relevance.

Trading volume also has a significant impact on abnormal returns at the 10% confidence interval. A 1% increase in daily trading volume corresponds to an abnormal return of just a little more than 0.4 basis points in the last five minutes of

---

[31] Taking the positive skew of our sample free float market capitalisation into account by using median rather than mean, the 5.6 basis points would translate into € 1.23 million.

trading. Despite the highly significant negative correlation between free float market capitalisation and abnormal returns, the regression setup shows no clear tendency of influence. The underlying reason for this is likely to be provided by the orthogonalisation of volume against free float. Liquidity cost necessary for the transaction, as measured in the XLM basis points, also appears to have no measurable effect on stock pricing during the last five minutes of trading.

The data presented in columns two to four of Table 4.3 confirm the expectation that the effect of creation/redemption-related flows is especially pronounced in small stocks and during bullish periods with generally high average returns. A one standard deviation shock in $CR_{i,t}$ in stocks below the 10^{th} size percentile, on average, leads to an abnormal return that is more than six times higher than the abnormal return in stocks with average free float market capitalisation. Abnormal returns due to flow-related transactions also appear to be stronger during generally bullish market periods, with an additional 2.4 basis points per $CR_{i,t}$ standard deviation. Hence, the effect of one $CR_{i,t}$ standard deviation in small stocks during boom phases (6.6 basis points) is more than ten times greater than the average influence. In general, bearish periods appear to have a significant negative effect on abnormal returns, irrespective of whether or not flow-related transactions happen on the same day. Except for the dummy for bust phases, all of the effects that can be captured during the observed time period appear to stem from flow-related causes, and, primarily, from the process of creation/redemption of ETF shares.

Table 4.4 presents the results for regression (4.6) and for two size-subsamples. In the Large-/Mid-Cap subsample we only observe the constituents of the largest two indices of the DAX index family, the blue-chip (DAX) and the mid-cap (MDAX) indices. In the Mid-/Small-Cap subsample, we focus on the mid-cap (MDAX) and technology (TecDAX) indices.[32] Similar to the full sample, we find a highly significant relationship between ETF flow-related transactions and abnormal returns. Compared

---

[32] The indices are not tested separately, in part, due to the very small sample size. The small-cap index (SDAX) is only covered by swap-based ETFs; thus it is not a part of our sample.

to an average stock in the full sample, the flow effect appears to be weaker in the stocks in the Large-/Mid-Cap subsample, yet substantially stronger in the stocks in the Mid-/Small-Cap subsample. Accounting for the subsample overlap – that is, the mid-cap stocks of the MDAX – the weaker (or stronger) than average effect in columns two and three can be attributed to the large-cap and small-cap stocks in the respective subsample. These findings further corroborate the results of regressions (4.7) and (4.8). Moreover, one can see the considerably larger combined effect of creation/redemption flow during boom times on small stocks relative to large stocks.

Table 4.3: Regression of full sample (including dummies and interaction terms)

Variables	Full Sample (4.4)	Econ. Regime (4.6)	Size (4.7)	Size/Regime (4.8)
CR	0.0082*** (0.0018)	0.0068*** (0.0017)	0.0079*** (0.0021)	0.0065*** (0.0020)
Volume	0.0043* (0.0022)	0.0042* (0.0022)	0.0042* (0.0023)	0.0042* (0.0023)
FreeFloat	-0.0011 (0.0035)	-0.0012 (0.0035)	-0.0027 (0.0035)	-0.0028 (0.0035)
XLM	-0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)
Boom		0.0230 (0.0332)		0.0240 (0.0331)
CRxBoom		0.0245*** (0.0056)		0.0236*** (0.0055)
Bust		-0.0860** (0.0339)		-0.0843** (0.0339)
CRxBust		-0.0018 (0.0054)		-0.0021 (0.0055)
Small			-0.0361 (0.0391)	-0.0360 (0.0391)
CRxSmall			0.0381*** (0.0145)	0.0362** (0.0145)
Big			0.0096 (0.0069)	0.0095 (0.0069)
CRxBig			-0.0010 (0.0022)	-0.0004 (0.0022)
Constant	0.0380 (0.0391)	0.0384 (0.0391)	0.0479 (0.0390)	0.0483 (0.0391)
Observations	62,054	62,054	62,054	62,054
R^2 (adjusted)	0.0798	0.0801	0.0801	0.0804
No. of stocks	137	137	137	137
Max. VIF	1.02	1.42	2.40	2.41

This table reports the results for the fixed effects regressions (4.4), (4.6), (4.7) and (4.8) of abnormal stock returns in the closing auction (in %) on the independent variables, namely daily flow-related net trading-volumes relative to free float market cap (*CR*) for the period t to t+2, the logarithm of trading volume in millions of euros (*Volume*) orthogonalised against free float, the free float market capitalisation available at the start of trading day (*FreeFloat*), the transaction liquidity cost measured in XLM basis points (*XLM*), the dummy variables for the most bullish and bearish trading days on CDAX (*Boom/Bust*), the dummy variables for the largest and smallest stocks in terms of free float market capitalisation (*Big/Small*) and the interaction terms of *CR* with *Boom* (*CRxBoom*), *Bust* (*CRxBust*), *Small* (*CRxSmall*), and *Big* (*CRxBig*) respectively. The respective standard errors are reported in parentheses. Furthermore, the number of observations, the adjusted R-squared, the number of underlying stocks for the respective regression and the maximum variance inflation factor (VIF) are stated. The statistical significance of the results being different from zero is based on a two-tailed test at the *10%, **5% and ***1% confidence levels.

Table 4.4: Regression of the full sample and the size-related subsamples

Variables	Full Sample (4.6)	Large-/Mid-Cap	Mid-/Small-Cap
CR	0.0068*** (0.0017)	0.0060*** (0.0016)	0.0098* (0.0054)
Volume	0.0042* (0.0022)	0.0031 (0.0020)	0.0026 (0.0029)
FreeFloat	-0.0012 (0.0035)	-0.0003 (0.0035)	-0.0012 (0.0043)
XLM	-0.0000 (0.0000)	-0.0000 (0.0000)	0.0001 (0.0000)
Boom	0.0230 (0.0332)	0.0243 (0.0332)	0.0860* (0.0454)
CRxBoom	0.0245*** (0.0056)	0.0181*** (0.0056)	0.0313** (0.0128)
Bust	-0.0860** (0.0339)	-0.0704** (0.0317)	-0.0834* (0.0481)
CRxBust	-0.0018 (0.0054)	-0.0042 (0.0057)	-0.0101 (0.0161)
Constant	0.0384 (0.0391)	0.0284 (0.0398)	-0.0075 (0.0467)
Observations	62,054	54,263	37,046
R^2 (adjusted)	0.0801	0.1039	0.0867
No. of stocks	137	97	109
Max. VIF	1.42	1.41	1.64

This table reports the results for the fixed effects regression (4.6) of abnormal stock return in the closing auction (in %) on independent variables, namely the daily flow-related net trading-volumes relative to free float market cap (*CR*) for the period *t* to *t*+2, the logarithm of trading volume in millions of euros (*Volume*) orthogonalised against free float, the free float market capitalisation available at the start of trading day (*FreeFloat*), the transaction liquidity cost measured in XLM basis points (*XLM*), the dummy variables for the most bullish and bearish trading days on CDAX (*Boom/Bust*) and the interaction terms of *CR* with *Boom* (*CRxBoom*) and *Bust* (*CRxBust*) respectively. The regression is performed for both the full sample and two subsamples—the Large-/Mid-Cap subsample, only containing large and mid-cap stocks of the DAX and MDAX indices, and the Mid-/Small-Cap subsample, only containing constituents of the mid-cap (MDAX) and technology (TecDAX) indices. The respective standard errors are reported in parentheses. Furthermore, the number of observations, the adjusted R-squared, the number of underlying stocks for the respective regression and the maximum variance inflation factor (VIF) are stated. The statistical significance of the results being different from zero is based on a two-tailed test at the *10%, **5% and ***1% confidence levels.

4.3.4 Discussion of the Findings

Implications of the Findings

All of the tested models corroborate our hypothesis that stock transactions caused by creations/redemptions of ETF shares lead to abnormal returns in the closing auction on the same day. Our finding of a positive relationship between ETF flow-related transactions and stock returns is in line with the empirical evidence provided by Kalaycıoğlu (2004) and Staer (2014). By looking at individual stocks rather than aggregated indices we are able to determine the effects on the most granular level possible. Even after controlling for stock- and time-fixed effects, as well as overall trading volume, free float and liquidity cost, the flow-related transactions still appear to have a statistically significant and economically viable impact on closing prices. In our basic model, the observed average effect of one standard deviation shock on free float being bound by flow-related transactions on abnormal return in the daily closing auction is 0.82 basis points. Applying the absolute average (median) daily ETF flow-related transaction volume per constituent stock of € 2.58 million (€ 602,000) to our model, the predicted total estimated profit from flow-related abnormal returns is € 212 (€ 49) per stock and day, or just below € 53,000 (€ 12,000) per year. For example, with a DAX ETF creation/redemption affecting up to 30 stocks,[33] hypothetical profits from flow-related abnormal returns could eventually add up to a little less than € 1.59 million (€ 370,000) per year.

The results of our augmented models also support Aggarwal and Wu's (2006) idea of small stock sensitivity to price shocks. The effect of flow-related stock transactions on abnormal returns appears to be considerably more pronounced in small stocks, for even minor orders have the potential to exert price pressure. With an estimated 4.3 basis points abnormal return per one standard deviation flow-related transaction volume, the effect for those stocks below the 10^{th} percentile is more than six times

---

33 The number of stocks eventually affected not only depends on the exact number of constituent stocks but also on whether the ETF uses an optimisation strategy for its portfolio.

greater than it is for an average sized stock in the sample. In addition, bullish trading days appear to further intensify the effect: abnormal returns due to flow-related trading on the 10% most bullish days are almost five times greater than on average trading days.

The fact that creations/redemptions of ETF shares have a substantial effect on abnormal returns in the closing auction implies that the observed equity market is not entirely efficient. This is all the more remarkable as we find that inefficiency to hold in the German market – certainly one of the most efficient markets in the world. Overall, our findings suggest that ETFs and ETF-related market flows have developed into such a strong determining factor for the observed equity market that they can significantly affect the prices of underlying securities. However, once we accept this notion, we can assume that counterparties will try to capitalise on this market inefficiency whenever possible.

The AP is one market participant, situated at the nexus between the investor and the ETF provider, who could actively trade on this observed effect. The AP's role in the creation/redemption process offers a possible explanation for the persistence of the impact of flow-related stock transactions on abnormal returns, even after controlling for the most common determinants. Assuming that APs have no inclination to bear any unwanted market risk in individual stocks throughout the trading day (e.g. due to an investor's NAV-based order of ETF shares), they should be motivated to promptly fill their orders and make the exchange of basket stocks against ETF shares. In the case of NAV-based trades, APs have the possibility to at least partly mitigate this risk. Knowing that the investor has given them a blank cheque by agreeing to the NAV as transaction price, they can optimise the positions in their own book without having to consider any constraints imposed by the investor, such as order limits. By entirely or at least partly placing (and filling) stock orders in the closing auction, they can reduce their own market risk. An AP should even have an active interest in manipulating the prices of underlying stocks around the close. Regardless of whether motive is profit maximisation or loss minimisation, the AP may try to capture some of the price impact of the traders' own transactions by actively manipulating the closing price and making a profit by buying (or selling) the index constituents beforehand at a price that is relatively lower (higher) than the closing price (cf. Bacidore et al., 2012). For instance, APs can affect the closing price by placing unlimited bulk orders in the closing auction, driving the price in the desired direction. If the previously estimated figures from our models hold, then trading on the observed ETF flow effect on abnormal returns in the closing auction could result in substantial potential earnings and might be an attractive endeavour, all the more in the case of NAV-based trading, where the investor has agreed to the closing price upfront. This is especially true if one takes into account that an AP commonly serves several ETFs from one or even multiple providers at the same time. APs should be particularly motivated to capitalise on the effect when their perceived risk is low and/or seemingly controllable (for example in phases with extraordinarily high creation/redemption-activity or market periods with clear and constant economic trends). This might serve to explain why the effect is so pronounced in bullish periods. First, we observe considerably higher creation/redemption activity in bullish periods. Second, in an environment of seemingly constantly increasing prices APs might think that the active price manipulation in the closing auction entails no risk. If a trading day exhibits extraordinarily bullish tendencies, the AP's risk of suddenly falling prices in the last hours might be considered to be relatively low.

ROBUSTNESS OF THE MODEL

Several proxies used in our models are based on assumptions and estimates. However, even after performing robustness checks on the individual variables, and the model as a whole, the overall outcome does not change substantially. The abnormal return during the closing auction is calculated with a market model against the MSCI EMU ex Germany index. This should guarantee that no ETF creation/redemption in the sample can alter the market benchmark's daily performance. By using HAC standard error estimates for the market beta determination and by excluding extreme beta outliers below or above the 0.5th percentile and 99.5th percentile respectively, our beta estimates should be considered sufficiently robust. For our regressions, we use market model betas calculated with hourly returns for the last 20 trading days. We run the basic regression (4.4) again with varying return periods for the market model, and we report the outcomes in Table 4.5. The results confirm that the effects described in the previous section, in particular the effects of flow-related transactions, are robust to changes in the methodology used to determine the abnormal returns.

Table 4.5: Regression of full sample with varying return periods for beta calculation

Variables	60 min	30 min	15 min
CR	0.0082***	0.0083***	0.0081***
	(0.0018)	(0.0019)	(0.0018)
Volume	0.0043*	0.0047**	0.0038*
	(0.0022)	(0.0021)	(0.0022)
FreeFloat	-0.0011	-0.0010	-0.0014
	(0.0035)	(0.0035)	(0.0035)
XLM	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)
Constant	0.0380	0.0176	0.0475
	(0.0391)	(0.0381)	(0.0469)
Observations	62,054	61,479	62,169
R^2 (adjusted)	0.0798	0.0796	0.0752
No. of stocks	137	137	137
Max. VIF	1.02	1.02	1.02

This table reports the results for the fixed effects regression (4.4) of abnormal stock return in the closing auction (in %) on the independent variables, namely the daily flow-related net trading-volumes relative to free float market cap (*CR*) for the period *t* to *t*+2, the logarithm of trading volume in millions of euros (*Volume*) orthogonalised against free float, the free float market capitalisation available at the start of the trading day (*FreeFloat*) and the transaction liquidity cost measured in XLM basis points (*XLM*). The abnormal return differs for each of the three models. For the first (60 min) regression, abnormal return is calculated with a market model using hourly stock returns. For the second (30 min) and third (15 min) regressions, half-hourly returns and returns for each quarter of an hour respectively, are used. The respective standard errors are reported in parentheses. Furthermore, the number of observations, adjusted R-squared, the number of underlying stocks for the respective regression and the maximum variance inflation factor (VIF) are stated. The statistical significance of the results being different from zero is based on a two-tailed test at the *10%, **5% and ***1% confidence levels.

Moreover, the data presented in columns three and four in Table 4.6 indicate that the aggregation of creation/redemption activity between t to t+1 is not the driving force of our results. Running the basic regression (4.4) again with $CR_{i,t}$ only accounting for net creations/redemptions reported for t+1 (column three) or t (column four) respectively, provides results that are not substantially different from those shown in Table 4.3, which were obtained with the t to t+2 proxy. In the fourth column in Table 4.6, it appears that the magnitude of $CR_{i,t}$ is affected by the change in methodology.

Table 4.6: Regression of full sample with varying *CR* aggregation

Variables	t+2 (full)	t+2 (split)	t+1	t
CR	0.0082*** (0.0018)	0.0103*** (0.0027)	0.0092*** (0.0023)	0.0043** (0.0022)
Volume	0.0043* (0.0022)	-0.0083 (0.0061)	0.0069*** (0.0021)	0.0068*** (0.0025)
FreeFloat	-0.0011 (0.0035)	0.0000 (0.0000)	0.0020 (0.0036)	-0.0028 (0.0042)
XLM	-0.0000 (0.0000)	0.0033 (0.0032)	-0.0000 (0.0000)	0.0000 (0.0000)
Constant	0.0380 (0.0391)	0.0909* (0.0547)	0.0819** (0.0403)	0.1010** (0.0491)
Observations	62,054	20,980	53,822	40,046
R^2 (adjusted)	0.0798	0.0880	0.0863	0.0931
No. of stocks	137	137	137	137
Max. VIF	1.02	1.02	1.01	1.00

This table reports the results for the fixed effects regression (4.4) of abnormal stock return in the closing auction (in %) on the independent variables, namely the daily flow-related net trading-volumes relative to free float market cap (*CR*), the logarithm of trading volume in millions of euros (*Volume*) orthogonalised against free float, the free float market capitalisation available at the start of the trading day (*FreeFloat*) and the transaction liquidity cost measured in XLM basis points (*XLM*). *CR* differs for each of the presented models t+2, t+1 and t. For the first regression (t+2), creation/redemption activity reported between *t* and *t+2* is considered; for the second regression (t+1), any creation/redemption between *t* and *t+1* is taken into account; the third regression (t) only considers creations/redemptions reported at *t*. For the first column, the full sample is used; for the second column, only every third day of the full sample is considered. The respective standard errors are reported in parentheses. Furthermore, the number of observations, the adjusted R-squared, the number of underlying stocks for the respective regression and the maximum variance inflation factor (VIF) are stated. The statistical significance of the results being different from zero is based on a two-tailed test at the *10%, **5% and ***1% confidence levels.

The liquidity cost per stock transaction is estimated to the nearest thousand euros by interpolating or extrapolating from existing daily XLM data for each stock for predefined hypothetical transaction volumes. For the extrapolation, we assume a generally linear increase of liquidity cost with increasing volume. However, we also use logarithmic growth rates and an average of linear and logarithmic growth to estimate, and each time we obtain similar results.

Overall, we can conclude that, regardless of the methodology used to calculate the abnormal return or any of the independent variables, the relevant coefficient ($CR_{i,t}$) is always highly significant and substantial. This even holds after controlling for the other potential abnormal return determinants. In addition, in most models the estimates for the constant are insignificant, which suggests that we have not omitted any significant variable in our regressions. With the exception of one model that had an adjusted R^2 of approximately 7.52%, all models exhibit an adjusted R^2 of 8% or more. For a cross-sectional regression, we consider these figures to be high enough to conclude that our main models have a sufficient goodness-of-fit.

Especially the models containing interaction terms bear a certain risk of high multicollinearity between the tested variables. However, because the maximum variance inflation factors (VIFs) did not exceed 2.41 for any of our main models, we can safely assume that the obtained results are robust.

4.4 CONCLUSION

In this study, we try to determine whether the creation or redemption of ETF shares has a measurable and significant effect on the returns of underlying stocks in the closing auction. Overall, the results of our analysis provide empirical evidence in support of our main hypothesis, which is that creations and redemptions of ETF shares have a highly significant and economically viable effect on the abnormal returns of underlying stocks in the closing auction. The effect is particularly pronounced in small (less liquid) stocks and on bullish trading days. Our findings corroborate the notion that ETFs have gained such a decisive role in equities markets that ETF flow-related stock transactions significantly affect stock prices and lead to abnormal returns in the closing auction. Finding this inefficiency in Germany, one of the most efficient stock markets in the world, we can safely assume that it also persists in other less developed, less efficient capital markets around the globe.

We also identify the AP, the key player in the process of ETF share creation and redemption, as a potential beneficiary of this observed market inefficiency. The AP not only has the opportunity but, given the prospect of additional profits, also a motive to manipulate closing prices. The hypothetical earnings potential of trading on the observed inefficiency that we derive from our models suggests that active price manipulation might be an attractive endeavour for an AP, all the more in the case of NAV-based trading where the investor has agreed to the closing price upfront. Consequently, dealing ETF shares through APs, for example through a NAV-based order, might entail hidden costs that, until now, have not been recognised in the literature and perhaps not even by investment professionals.

5. CONCLUSION

The overall goal of the three essays presented in this dissertation is to broaden the understanding of equity ETFs and their ecosystem by elaborating on some of their most relevant interrelations with corresponding markets. Based on a dataset of German-listed equity ETFs and their underlying stocks, each essay concentrates on a specific ETF-market relationship. The first essay's research question relates to ETF performance by elaborating on the effect of stock market liquidity on ETFs' tracking ability. The second essay examines the impact of hiring an additional DMM on ETF liquidity and seeks to determine whether DMMs who primarily rely on algorithmic or high-frequency trading are systematically better in providing liquidity to ETFs. The third essay investigates the effect of creations and redemptions of ETF shares on the underlying stocks' returns.

5.1 MAIN RESULTS AND IMPLICATIONS

5.1.1 DETERMINANTS OF TRACKING ERROR IN EQUITY ETFS – THE ROLE OF MARKET LIQUIDITY

In light of increasing competition and mounting pressure on ETF providers to differentiate their ETFs from rival products, the ability to track the underlying benchmark as closely as possible has become essential. The first essay's focus is on identifying the determining factors of daily tracking error in equity ETFs, with a particular interest in the way that market liquidity of underlying securities has an impact on ETF tracking ability. The analysis in Chapter 2 builds on the work by Rompotis (2012) and Meinhardt et al. (2012), who all confirm a positive impact for German ETFs. Contrary to those studies, the essay presented here takes a more bottom-up approach by taking the liquidity of individual stocks into account.

The results of the analysis show that daily tracking error is dependent on management fees, cash holdings, dividend yield, cash distributions from an ETF to its investors, portfolio adjustments, the process of creation/redemption of ETF shares and the market liquidity of stocks in the underlying portfolio. This last item affects tracking error both directly and in interaction with portfolio adjustments. Even after separately controlling for cash holdings, portfolio adjustments and creation/redemption, market liquidity of stocks still has a strongly significant and independent impact on ETF tracking error. One possible explanation is that relatively small, internally initiated market transactions that are not fully captured by the large transaction-oriented variables for creation and redemption or portfolio adjustment cause liquidity-related transaction costs. These can occur, for instance, through constant rebalancing by fund management in order to optimise or match index weights more closely over time. In these cases, the observed independent liquidity effect represents the liquidity cost borne by the ETF for its attempts to optimise the weights of the underlying portfolio.

The analysis shows that the creation and redemption of ETF shares significantly affects tracking error, which combines with the findings for portfolio adjustments to suggest that liquidity cost plays a significant role in any market event that triggers a change in the underlying portfolio. Furthermore, it suggests that, contrary to the notion postulated by Gallagher and Segara (2006) and Gastineau (2004), ETFs are not fully shielded from the effect of the liquidity cost of their underlying securities in the course of daily in-kind creation and redemption. One explanation proposed in Chapter 2 relates to the imperfect replication of index weights in the ETF portfolio arising from the indivisibility of single shares. This results in a remainder in either cash or in stock that prevents a perfect match with actual index weights. As a result, the ETF will exhibit a tracking error due to differences between the benchmark and the underlying basket. Another explanation proposed is the daily charging or attribution of fees from or to the fund; although effectively being paid (or received) on a monthly or quarterly basis, fees to or from the fund like securities-lending fees or management fees are commonly calculated and attributed to the ETF-NAV on a daily basis. As a result, sudden and substantial changes in the portfolio composition might affect the tracking ability of an ETF.

Overall, the findings of the first essay suggest that while cash inflows and outflows in the form of dividends and distributions respectively appeared to be the factors with the most substantial effect on tracking errors, the market liquidity of underlying stocks and any market events that cause changes in the ETF portfolio also have a significant effect on the tracking ability of equity ETFs. The findings on creations/redemptions clearly challenge the notion of ETF tracking ability being immune to market frictions during these processes.

5.1.2 IMPACT OF DESIGNATED MARKET MAKERS ON ETF LIQUIDITY

The recent trend among ETF providers to outsource completely or partly the market making capabilities of ETFs to external specialists necessitates a thorough understanding of the impact of hiring an additional DMM on ETF liquidity. This having been noted, the literature on market makers' effect on market quality is still scarce when it comes to ETFs – Chapter 3 helps to address this research gap. Using an extensive dataset of fund and liquidity data for equity ETFs listed in Germany, the essay investigates whether hiring an additional DMM significantly affects the liquidity cost of an ETF in the secondary market. Moreover, it analyses whether DMMs that primarily rely on algorithmic and high-frequency trading techniques are systematically better than other types of market makers in providing liquidity to an ETF.

The essay not only finds the effects of hiring an additional market maker to be highly significant but also economically substantial: 12 months into treatment, the estimated annual liquidity cost reductions from hiring a non-HFT or a HFT DMM amount to an average € 70,000 and € 238,000 respectively. According to the models tested in Chapter 3 and based on the aggregated average annual trading volumes of all XETRA-listed equity ETFs, the total annual liquidity cost reduction that could be passed on to the market and investors from the 452 observed instances of hiring an additional DMM is approximately € 90 million.

The notion of market makers reducing liquidity cost is in itself rather intuitive and has been widely corroborated by empirical research. However, since the research question in Chapter 3 relates to the impact of *additional* market makers, the results suggest that those designated sponsors hired prior to treatment do not reduce the liquidity cost to the lowest levels possible. One explanation could be that by adding another market maker to the pool of existing DMM, the ETF provider increases competition between individual market makers, which leads to decreased spreads and increased liquidity. Finding no significant changes in AuM or trading volume one year after hiring the additional DMM, the analysis suggests that despite the improved liquidity that can be attributed to contracting the DMM, investors do not appear to honour the resulting higher market quality immediately, so that ETF providers should not count on benefiting from increased funds and fees, at least in the first year after hire.

The essay also contributes to the understanding of how differences in the setups of various market makers might affect their ability to generate liquidity. It does so by providing evidence that designated sponsors that rely primarily on algorithmic or high-frequency trading in their market making activities cause significantly greater reductions in ETF liquidity costs than respective non-HFT market makers, at least after a certain amount of time. Six months after hiring a designated sponsor, a HFT DMM outperforms its non-HFT counterpart by some 5.46 XLM basis points on average. While these results validate the existing empirical evidence on the superiority of HFT market making, they also suggest that the HFT market maker's systemic advantages require some time to materialise.

Overall liquidity is a key factor in determining an ETF's market quality, and higher liquidity is generally a desirable outcome for any ETF. Given that most ETFs on XETRA currently rely on the minimum required number of one DMM, there appears to be considerable potential for liquidity cost reductions in the market. However, since hiring an additional DMM is itself a costly endeavour, the ETF provider must decide whether the additional liquidity gain is worth the financial obligations that come with the hiring. If providers decide to hire an additional DMM they should be aware that the evidence indicates that the already hired market makers do not on average appear to bring liquidity to its best possible level and that the effect of hiring is driven largely by inter-DMM competition. Moreover, if improved liquidity is the sole motivation behind hiring an additional designated sponsor, then providers should contract a HFT DMM, as they are systematically better at reducing liquidity cost.

5.1.3 Effect of ETF Flow on Underlying Stock Returns

The essay in Chapter 4 examines whether after years of growth, ETFs now play a decisive enough role in capital markets to have a measurable effect on their underlying securities. It does so by analysing the relationship between the creation and redemption of shares of DAX index family ETFs and the abnormal returns of underlying securities in the closing auction at the level of individual stocks, the finest degree of granularity possible.

In line with the empirical evidence provided by Kalaycıoğlu (2004) and Staer (2014), the findings in Chapter 4 corroborate the view that ETFs significantly affect stock prices through flow-related stock transactions and cause abnormal returns for underlying stocks in the closing auction. Even after controlling for stock- and time-fixed effects, overall trading volume, free float and liquidity costs, flow-related transactions still have a statistically significant impact on closing prices of underlying stocks. The effect is particularly pronounced in small (less liquid) stocks, with flow-related abnormal returns being six times greater than in average-sized stocks, and on generally bullish trading days, where the effect is almost five times greater than on average trading days.

The results presented are not only highly significant but also economically viable. According to the models tested, the estimated potential gains from trading on this ETF-flow-induced pricing inefficiency would add up to € 53,000 per stock annually. With creations/redemptions of DAX ETF shares affecting up to 30 stocks, for instance,[34] profits from flow-related abnormal returns could add up to just under € 1.59 million per year.

The analysis identifies APs as the potential beneficiaries of this observed market inefficiency. Their role in the creation and redemption processes offers a possible explanation for the persistence of the impact of flow-related stock transactions on abnormal returns, even after controlling for the most common determinants. APs not only have the opportunity but also the motivation to risk exposure to potential losses, given the prospect of additional profits; the hypothetical earnings potential of trading on the observed inefficiency estimated by the models in Chapter 4 suggests that active price manipulation might be an attractive endeavour for an AP, especially with NAV-based trading in which the investor has agreed upfront to the closing price. The AP should be particularly motivated to capitalise on the effect when the perceived risk is low and seemingly predictable, as during market periods with clear and constant economic trends.

Consequently, the results presented in Chapter 4 suggest that dealing ETF shares on the primary market through APs, through a NAV-based order for example, might entail hidden costs that had not previously been recognised in the literature or perhaps even by investment professionals.

5.2 AVENUES FOR FUTURE RESEARCH

The three essays presented in this dissertation address specific sets of ETF-market interrelations for equity ETFs. While it is the intention to present conclusions from the empirical evidence that are universally valid, the dissertation's clear focus on equity listed ETFs listed on the German market itself already points to some obvious avenues for future research. In order to test the validity of the conclusions presented for other

---

[34] The number of stocks ultimately affected depends not only on the exact number of constituent stocks but also on whether the ETF uses an optimisation strategy for its portfolio.

markets and asset classes, the current focus on Germany and on equity ETFs could be loosened. Replicating the methodology for fixed-income products or for other developed or emerging markets could grant valuable insights and broaden the empirical foundation for the relationships described in this dissertation. For instance, the fact that the effect of flow-related transactions on stock returns described in Chapter 4 persists in Germany, one of the most efficient stock markets in the world, suggests that those results might also hold and be even stronger in other less developed, less efficient capital markets.

Even within the dissertation's strict parameters of German-listed passive equity ETFs, several issues that deserve further scrutiny emerge from the results. The research questions in Chapters 2 and 3 relate to the effect of external market forces on ETF market quality. With market liquidity being at least partially dependent on general market conditions, it could prove worthwhile to re-examine the effect of stock market liquidity on ETF tracking ability during times of financial turmoil. The same is true for the finding that HFT DMMs are systematically better at providing liquidity than non-HFT market makers. Especially in light of the previously noted critique on perceived excesses of algorithmic and high-frequency trading of ETFs, it would be valuable to understand whether the structural advantage of HFT market makers also holds during generally bearish or fast market periods, or whether other forms of market making can adapt more easily to such situations.

Based on Chapter 3's and 4's findings on the systematic differences between HFT and non-HFT market making and on APs being potential beneficiaries of exploiting the effect of ETF creation and redemption on underlying stock returns, the question emerges of whether the underlying type of market making mechanism (HFT vs. non-HFT) has any impact on an AP's propensity to be engaged in active market manipulation.

In light of the evidence on market maker impact on ETF liquidity, one could also re-examine Anand et al.'s (2009) notion of "liquidity [begetting] liquidity" (p. 1447). Whereas initial analysis in Chapter 3 appears to refute or at least challenge their claim, determining the magnitude and statistical significance of indirect effects in the aftermath of hiring an additional market maker might be a meaningful avenue for future research that would help ETF providers in decision-making processes involved in hiring a suitable DMM.

ETFs have undeniably arrived at the centre stage of financial markets and their "explosive growth […] in recent years poses a challenge that isn't going away, and may well become even more acute as new ETFs enter the market" (SEC, 2015). The challenge that these financial products pose originates primarily from their sheer complexity. While research on some of the key building blocks of ETFs has made significant progress, it has barely begun in other important areas, a thorough and holistic understanding of the entire ecosystem remains a distant goal. However, the better reaction to this seemingly boundless complexity is not resignation but rather an urge to conduct future research in a systematic, energetic way.

BIBLIOGRAPHY

Abadie, A., & Imbens, G. W. (2012). Matching on the estimated propensity score. *NBER Working Paper.*

Aber, J. W., Li, D., & Can, L. (2009). Price volatility and tracking ability of ETFs. *Journal of Asset Management, 10*(4), 210-221.

Abner, D. J. (2010). *The ETF handbook: How to value and trade exchange-traded funds* (Vol. 569). Hoboken, NJ: John Wiley & Sons.

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

Aggarwal, R. K., & Wu, G. (2006). Stock market manipulations. *Journal of Business, 79*(4), 1915-1953.

Agrrawal, P., & Clark, J. M. (2009). Determinants of ETF liquidity in the secondary market: A five-factor ranking algorithm. *ETFs and Indexing, 2009*(1), 59-66.

Anand, A., & Subrahmanyam, A. (2008). Information and the intermediary: Are market intermediaries informed traders in electronic markets. *Journal of Financial and Quantitative Analysis, 43*(1), 1-28.

Anand, A., & Venkataraman, K. (2013). Should exchanges impose market maker obligations? *Working Paper.*

Anand, A., & Weaver, D. G. (2006). The value of the specialist: Empirical evidence from the CBOE. *Journal of Financial Markets, 9*(2), 100-118.

Anand, A., Tanggaard, C., & Weaver, D. G. (2009). Paying for market quality. *Journal of Financial and Quantitative Analysis, 44*(6), 1427-1457.

Bacidore, J., Polidore, B., Xu, W., & Yang, C. Y. (2012). Trading around the close. *Journal of Trading, 8*(1), 48-57.

Bae, K.-H., Kang, J.-K., & Wang, J. (2013). The costs of ETF membership: The valuation effect of ETFs on underlying firms. *Working Paper.*

BaFin. (2012). *BaFin quarterly Q1/12.* Bundesanstalt für Finanzdienstleistungsaufsicht, Frankfurt am Main. Retrieved from https://www.bafin.de/SharedDocs/Downloads/EN/Mitteilungsblatt/Quarterly/bq1201.html

Barclay, M. J., Christie, W. G., Harris, J. H., Kandel, E., & Schultz, P. H. (1999). Effects of market reform on the trading costs and depths of Nasdaq stocks. *Journal of Finance, 54*(1), 1-34.

Battalio, R. (2003). Discussion of "Do 'thinly-traded' stocks benefit from specialist intervention?". *Journal of Banking & Finance, 27*(9), 1855-1857.

BCG. (2015). *Global asset management 2015 - sparking growth with go-to-market excellence.* BCG Perspectives, The Boston Consulting Group, Inc., Boston.

Ben-David, I., Franzoni, F., & Moussawi, R. (2014). Do ETFs increase volatility. *Fisher College of Business Working Paper No. 2011-03-20.*

Bessembinder, H., Hao, J., & Lemmon, M. (2011). Why designated market makers? Affirmative obligations and market quality. *Working Paper.*

Bessembinder, H., Hao, J., & Zheng, K. (2013). Market Making Obligations and Firm Value. *Working Paper.*

Biais, B., & Foucault, T. (2014). HFT and market quality. *Bankers, Markets & Investors*(128), 5-19.

Blitz, D., & Huij, J. (2012). Evaluating the performance of global emerging markets equity exchange-traded funds. *Emerging Markets Review, 13*(2), 149-158.

Blitz, D., Huij, J., & Swinkels, L. (2012). The performance of European index funds and exchange-traded funds. *European Financial Management, 18*(4), 649-662.

Blume, M. E., & Edelen, R. M. (2003). S&P 500 indexers, delegation costs and liquidity mechanisms. *The Wharton School Paper No. 04-03.*

Boehmer, E., Fong, K., & Wu, J. (2014). International evidence on algorithmic trading. *AFA 2013 San Diego Meetings Paper.*

Borkovec, M., & Serbin, V. (2013). Create or buy: A comparative analysis of liquidity and transaction costs for selected U.S. ETFs. *Journal of Portfolio Management, 39*(4), 118-131.

Brogaard, J., Hendershott, T., & Riordan, R. (2014). High-frequency trading and price discovery. *Review of Financial Studies, 27*(8), 2267-2306.

Brogaard, J., Hendershott, T., Hunt, S., Latza, T., Pedace, L., & Ysusi, C. (2013). High-frequency trading and the execution costs of institutional investors. *FSA Occasional Paper Series 43.*

Calamia, A., Deville, L., & Riva, F. (2013). Liquidity in European equity ETFs: What really matters. *GREDEG Working Paper.*

Calamia, A., Deville, L., & Riva, F. (2014). The determinants of ETF liquidity: Theory and evidence from European markets. *GREDEG Working Paper.*

Caliendo, M., & Kopeinig, S. (2005). Some practical guidance for the implementation of propensity score matching. *Discussion Paper.* Berlin: DIW - Deutsches Institut für Wirtschaftsforschung.

Chang, R. P., Rhee, S. G., Stone, G. R., & Tang, N. (2008). How does the call market method affect price efficiency? Evidence from the Singapore Stock Market. *Journal of Banking & Finance, 32*(10), 2205-2219.

Charitou, A., & Panayides, M. (2009). Market making in international capital markets: Challenges and benefits of its implementation in emerging markets. *International Journal of Managerial Finance, 5*(1), 50-80.

Charupat, N., & Miu, P. (2013). Recent developments in exchange-traded fund literature - pricing efficiency, tracking ability, and effects on underlying securities. *Managerial Finance, 39*(5), 427-443.

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.

Chu, P. K.-K. (2013). Tracking errors and their determinants: Evidence from Hong Kong exchange traded funds. *Working Paper.*

Comerton-Forde, C., & Putniņš, T. J. (2011). Measuring closing price manipulation. *Journal of Financial Intermediation, 20*(2), 135-158.

Comerton-Forde, C., & Putniņš, T. J. (2014). Stock price manipulation: Prevalence and determinants. *Review of Finance, 18*(1), 23-66.

Comerton-Forde, C., & Rydge, J. (2006). Call auction algorithm design and market manipulation. *Journal of Multinational Financial Management, 16*(2), 184-198.

Comerton-Forde, C., Lau, S. T., & McInish, T. (2007). Opening and closing behavior following the introduction of call auctions in Singapore. *Pacific-Basin Finance Journal, 15*(1), 18-35.

Coval, J., & Stafford, E. (2007). Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics, 86*(2), 479-512.

Cushing, D., & Madhavan, A. (2001). The Hidden Cost of Trading at the Close. *Trading, 2001*(1), 12-19.

Daley, P., Dorencz, P., & Bargerstock, D. (2010). ETF liquidity explained. *Journal of Indexes, 2010*(2), 10-17.

De Winne, R., Gresse, C., & Platten, I. (2013). Liquidity and risk sharing benefits from the introduction of an ETF. *Working Paper.*

Del Guercio, D., & Reuter, J. (2014). Mutual fund performance and the incentive to generate alpha. *Journal of Finance, 69*(4), 1673-1704.

Delcoure, N., & Zhong, M. (2007). On the premiums of iShares. *Journal of Empirical Finance, 14*(2), 168-195.

Demsetz, H. (1968). The cost of transacting. *Quarterly Journal of Economics, 82*(1), 33-53.

Deutsche Bank Research. (2014). *ETF annual review and outlook.* London: Deutsche Bank Securities Inc.

Deutsche Bank Research. (2016). *European monthly ETF market review - January.* London: Deutsche Bank AG.

Deutsche Börse. (2014). *Xetra release 15.0 - market model equities.* Frankfurt am Main: Deutsche Börse AG.

Driscoll, J., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent data. *Review of Economics and Statistics, 80*(4), 549-560.

Edelen, R. M., & Warner, J. B. (2001). Aggregate price effects of institutional trading: A study of mutual fund flow and market returns. *Journal of Financial Economics, 59*(2), 195-220.

EDHEC. (2015). *The EDHEC European ETF survey 2014.* EDHEC-Risk Institute Publication, EDHEC-Risk Institute, Nice.

Elia, M. (2012). Tracking error of traditional and synthetic European exchange-traded funds. *Working Paper.*

Elton, E. J., Gruber, M. J., Comer, G., & Li, K. (2002). Spiders: Where are the bugs? *Journal of Business, 75*(3), 453-472.

Engle, R. F., & Sarkar, D. (2006). Premiums-discounts and exchange-traded funds. *ETF and Indexing, 2006*(1), 35-53.

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

Fama, E. F., & French, K. (2010). Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance, 65*(5), 1915-1947.

Felixson, K., & Pelli, A. (1999). Day end returns – stock price manipulation. *Journal of Multinational Financial Management, 9*(2), 95-127.

Fischer, M., Kaserer, C., & Overkott, M. (2013). ETFs vs. Index-Investmentfonds. *Absolut report*(5), 22-29.

Flood, C. (2015, October 25). 'Time to re-examine the entire exchange traded fund ecosystem'. *Financial Times*. Retrieved from http://www.ft.com

French, K. (2008). Presidential address: The cost of active investing. *Journal of Finance, 63*(4), 1537-1573.

Frino, A., & Gallagher, D. R. (2001). Tracking S&P 500 index funds. *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.

Frino, A., Gallagher, D. R., Neubert, A. S., & Oetomo, T. N. (2004). Index design and implications for index tracking - evidence from S&P 500 index funds. *Journal of Portfolio Management, 30*(2), 89-95.

Gallagher, D. R., & Segara, R. (2006). The performance and trading characteristics of exchange-traded funds. *Journal of Investment Strategy, 1*(2), 49-60.

Gastineau, G. L. (2002). Equity index funds have lost their way - a solution to growing problems. *Journal of Portfolio Management, 28*(2), 55-64.

Gastineau, G. L. (2004). The benchmark index ETF performance problem - a simple solution. *Journal of Portfolio Management, 30*(2), 96-103.

Giulianini, P. (2012). Understanding ETF trading and liquidity in Europe. *ETFs and Indexing, 2012*(1), 81-86.

Golub, G. H., & Van Loan, C. F. (1996). *Matrix computations* (3 ed.). Baltimore: Johns Hopkins University Press.

Gomber, P., Arndt, B., Lutat, M., & Uhle, T. (2011). High-frequency trading. *Working Paper.*

Greenwich Associates. (2014). *ETFs: An evolving toolset for U.S. institutions.* Greenwich Associates, LLC, Stamford.

Gruber, M. J. (1996). Another puzzle: The growth in actively managed mutual funds. *Journal of Finance, 51*(3), 783-810.

Hachmeister, A. (2007). *Informed traders as liquidity providers: Evidence from the German equity markets.* Wiesbaden: Deutscher Universitäts-Verlag.

Hasbrouck, J., & Saar, G. (2013). Low-latency trading. *Journal of Financial Markets, 16*(4), 646-679.

Hedge, S. P., & McDermott, J. B. (2004). The market liquidity of DIAMONDS, Q's, and their underlying stocks. *Journal of Banking & Finance, 28*(5), 1043-1067.

Hendershott, T., & Moulton, P. C. (2011). Automation, speed and stock market quality: The NYSE's Hybrid. *Journal of Financial Markets, 14*(4), 568-604.

Hendershott, T., & Riordan, R. (2013). Algorithmic trading and the market for liquidity. *Journal of Financial and Quantitative Analysis, 48*(4), 1001-1024.

Hendershott, T., Jones, C. M., & Menkveld, A. J. (2011). Does algorithmic trading improve liquidity? *Journal of Finance, 66*(1), 1-33.

Hengelbrock, J. (2008). Designated sponsors and bid-ask spreads on Xetra. *Working Paper.*

Hill, J. M., Nadig, D., & Hougan, M. (2015). A comprehensive guide to exchange-traded funds (ETFs). *CFA Research Foundation Publications, 2015*(3), pp. 1-181.

Hillion, P., & Suominen, M. (2004). The manipulation of closing prices. *Journal of Financial Markets, 7*(4), 351-375.

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *Stata Journal, 7*(3), 281-312.

Huang, Y. C., & Chan, S. H. (2014). The Trading Behavior of Attention Securities with Different Closing Mechanisms: Evidence from Taiwan. *Review of Pacific Basin Financial Markets and Policies, 17*(4), 1450026.

ICI. (2015). *2015 Investment company factbook.* Investment Company Institute.

Irvine, P. J., Benston, G. J., & Kandel, E. (2000). Liquidity beyond the inside spread: Measuring and using information in the limit order book. *Working Paper.*

Jarnecic, E., & Snape, M. (2014). The provision of liquidity by high-frequency participants. *Financial Review, 49*(2), 371-394.

Johnson, B., Bioy, H., & Rose, G. (2012). Snythetic ETFs under the microscope. *ETFs and Indexing, 2012*(1), 94-100.

Johnson, W. F. (2009). Tracking errors of exchange traded funds. *Journal of Asset Management, 10*(4), 253-262.

Jones, C. M. (2013). What do we know about high-frequency trading? *Working Paper.*

Jotikasthira, C., Lundblad, C., & Ramadorai, T. (2012). Asset fire sales and purchases and the international transmission of funding shocks. *Journal of Finance, 67*(6), 2015-2050.

Jovanovic, B., & Menkveld, A. J. (2011). *Middlemen in limit order markets.* Western Finance Association (WFA).

Kadıoğlu, E., Küçükkocaoğlu, G., & Kılıç, S. (2015). Closing price manipulation in Borsa Istanbul and the impact of call auction sessions. *Borsa Istanbul Review, 15*(3), 213-221.

Kalaycıoğlu, S. (2004). Exchange traded fund flows. *Working Paper.*

Kim, M.-Y. (2014). *The true cost of exchange traded fund NAV trading.* Unpublished Master's Thesis, Technische Universität München, München.

Kittsley, D., & Edrosolan, J. (2008). Looking inside liquidity: An ETF trading case study. *ETFs and Indexing, 2008*(1), 32-36.

Kostovetsky, L. (2003). Index mutual funds and exchange-traded funds - a comparison of two methods of passive investment. *Journal of Portfolio Management, 29*(4), 80-92.

Krogmann, M. (2011). Quantifying liquidity risk. In P. Kaur, K. Lubecka, & A. Hornbrook (Eds.), *Buy-side intelligence: the Euromoney guide to securities trading* (pp. 37-41). North Hill: Euromoney Trading Ltd.

Küçükkocaoğlu, G. (2008). Intra-day stock returns and close-end price manipulation in the Istanbul Stock Exchange. *Frontiers in Finance and Economics, 5*(1), 46-84.

Kundisch, D., & Klein, C. (2009). Der Tracking Error von indexnachbildenden Instrumenten auf den DAX - eine empirische Analyse des ETFs DAX EX sowie zehn Indexzertifikaten. *Der Betrieb, 62*(22), 1141 - 1145.

Li, M. (2013). Using the propensity score method to estimate causal effects: A review and practical guide. *Organizational Research Methods, 16*(2), 188-226.

Lin, A., & Chou, A. (2006). The tracking error and premium/discount of Taiwan's first exchange traded fund. *Web Journal of Chinese Management Review, 9*(3), 1-20.

Lin, C.-C., & Chiang, M.-H. (2005). Volatility effect of ETFs on the constituents of the underlying Taiwan 50 Index. *Applied Financial Economics, 15*(18), 1315–1322.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics, 47*(1), 13-37.

MacKinlay, A. C. (1997). Event Studies in Economics and Finance. *Journal of Economic Literature, 35*(1), 13-39.

Madura, J., & Ngo, T. (2008). Impact of ETF inception on the valuation and trading of component stocks. *Applied Financial Economics, 18*(12), 995-1007.

Malinova, K., Park, A., & Riordan, R. (2013). Do retail traders suffer from high frequency traders? *Working Paper.*

Malkiel, B. (1995). Returns from investing in equity mutual funds 1971 to 1991. *Journal of Finance, 50*(2), 549 – 572.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance, 7*(1), 77-91.

McKinsey & Company. (2011). *The second act beginns for ETFs - a disruptive investment vehicle vies for center stage in asset management*. McKinsey & Company, Inc.

Meinhardt, C., Mueller, S., & Schoene, S. (2012). Synthetic ETFs: Will full replication survive? *Working Paper.*

Menkveld, A. J., & Wang, T. (2013). How do designated market makers create value for small-caps? *Journal of Financial Markets, 16*(3), 571-603.

Milonas, N. T., & Rompotis, G. G. (2006). Investigating European ETFs: The case of the Swiss exchange traded funds. *Annual HFAA Conference Paper.*

Nimalendran, M., & Petrella, G. (2003). Do 'thinly-traded' stocks benefit from specialist intervention? *Journal of Banking & Finance, 27*(9), 1823-1854.

OFR. (2013). *Asset management and financial stability.* U.S. Department of Treasury, Office of Financial Research, Washington. Retrieved from https://financialresearch.gov/reports/

Perotti, P., & Rindi, B. (2010). Market makers as information providers: The natural experiment of STAR. *Journal of Empirical Finance, 17*(5), 895-917.

Pinfold, J. F., & He, D. (2012). The impact of introducing a pre-close on the New Zealand share market. *Journal of Financial Regulation and Compliance, 20*(1), 99-110.

Pope, P. F., & Yadav, P. K. (1994). Discovering errors in tracking error. *Journal of Portfolio Management, 20*(2), 27-32.

Qadan, M., & Yagil, J. (2012). On the dynamics of tracking indices by exchange traded funds in the presence of high volatility. *Managerial Finance, 38*(9), 804-832.

Rakowski, D., & Wang, X. (2009). The dynamics of short-term mutual fund flows and returns: A time-series and cross-sectional investigation. *Journal of Banking & Finance, 33*(11), 2102–2109.

Ramaswamy, S. (2011). Market structures and systemic risks of exchange-traded funds. *BIS Working Papers.* Bank for International Settlements.

Richie, N., & Madura, J. (2007). Impact of the QQQ on liquidity and risk of the underlying stocks. *Quarterly Review of Economics and Finance, 47*(3), 411-421.

Riordan, R., & Storkenmaier, A. (2012). Latency, liquidity and price discovery. *Journal of Financial Markets, 15*(4), 416-437.

Roll, R. (1992). A mean/variance analysis of tracking error. *Journal of Portfolio Management, 18*(4), 15-24.

Rompotis, G. G. (2006). Evaluating the performance and the trading characteristics of iShares. *Working Paper.*

Rompotis, G. G. (2011). Predictable patterns in ETFs' return and tracking error. *Studies in Economics and Finance, 28*(1), 14-35.

Rompotis, G. G. (2012). The German exchange traded funds. *IUP Journal of Applied Finance, 18*(4), 62-82.

Roncalli, T., & Zheng, B. (2014). Measuring the liquidity of ETFs: An application to the European market. *Journal of Trading, 9*(3), 79-108.

Rösch, C., & Kaserer, C. (2013). Market liquidity in the financial crisis: The role of liquidity commonality and flight-to-quality. *Journal of Banking and Finance, 37*(7), 2284-2302.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician, 39*(1), 33-38.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology, 2*(3-4), 169-188.

Ryan, T. F., & Follet, J. (2001). Are ETFs liquid securities? *Institutional Investors Guide to Exchange Traded Funds, 2001*(1), 106-110.

Sanchez, B., & Wei, P. (2010). The liquidity of exchange traded funds. *International Review of Applied Financial Issues & Economics, 2*(4), 621-646.

SEC. (2010). *Findings regarding the market events of May 6, 2010 - Report of the staffs of the CFTC and SEC to the joint advisory committee on emerging regulatory issues.* Washington: U.S. Securities & Exchange Commission / U.S. Commodity

Futures Trading Commission. Retrieved from www.sec.gov/news/studies/2010/marketevents-report.pdf

SEC. (2015, October 15). *Public statements: How can the markets best adapt to the rapid growth of ETFs.* Retrieved from U.S. Securities and Exchange Commission: https://www.sec.gov/news/statement/how-can-markets-adapt-to-rapid-growth-etfs.html

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance, 19*(3), 425–442.

Sharpe, W. F. (1991). The arithmetic of active management. *Financial Analysts' Journal, 47*(1), 7-9.

Shin, S., & Soydemir, G. (2010). Exchange-traded funds, persistence in tracking errors and information dissemination. *Journal of Multinational Financial Management, 20*(4), 214-234.

Skjeltorp, J. A., & Ødegaard, B. A. (2010). Why do firms pay for liquidity provision in limit order markets? *Working Paper.* Norges Bank.

Solivas, E. S., Ramirez, G. M., & Manalo, A. N. (2007). The propensity score matching for correcting sample selection bias. *10th National Convention on Statistics (NCS) Conference Paper.*

Staer, A. (2014). Fund flows and underlying returns: The case of ETFs. *Working Paper.*

Stange, S., & Kaserer, C. (2011). The impact of liquidity risk: A fresh look. *International Review of Finance, 11*(3), 269-301.

Stoll, H. R. (2003). Market microstructure. In G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.), *Handbook of the Economics of Finance* (Vols. 1, Part A, pp. 553-604). Elsevier.

Suen, W., & Wan, K.-M. (2013). Sniping to manipulate closing prices in call auctions: Evidence from the Hong Kong Stock Exchange. *Working Paper.*

Svetina, M. (2010). Exchange traded funds: Performance and competition. *Journal of Applied Finance, 20*(2), 130-145.

Van Ness, B. F., Van Ness, R. A., & Warr, R. S. (2005). The impact of the introduction of index securities on the underlying stocks: The case of the Diamonds and the Dow 30. *Advances in Quantitative Analysis of Finance and Accounting, 2*, pp. 105-128.

Venkataraman, K., & Waisburd, A. C. (2007). The value of the designated market maker. *Journal of Financial and Quantitative Analysis, 42*(3), 735-758.

Wong, K. H., & Shum, W. C. (2010). Exchange-traded funds in bullish and bearish markets. *Applied Economics Letters, 17*(16), 1615-1624.

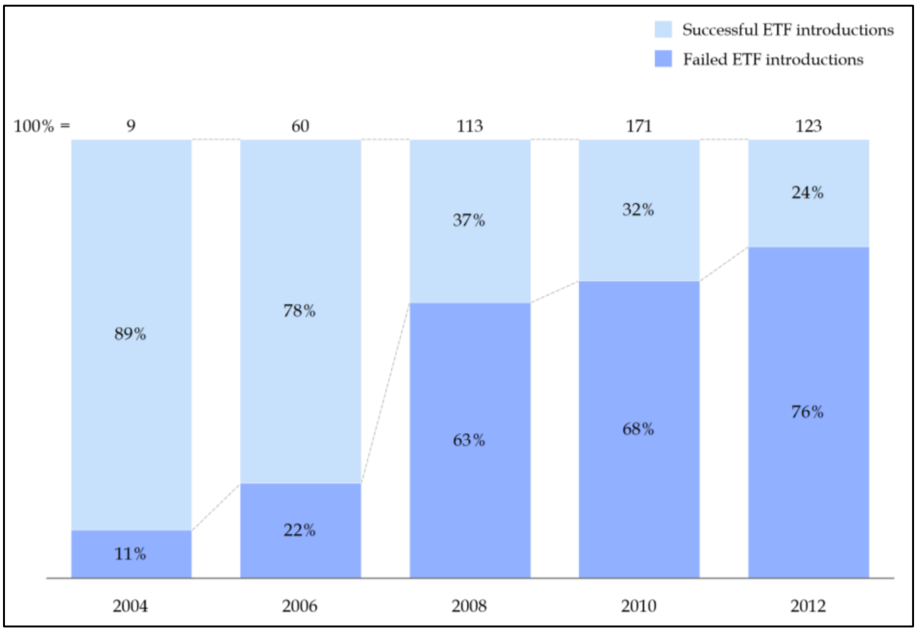
Yao, C., & Ye, M. (2014). Tick size constraints, high-frequency tading, and liquidity. *Working Paper.*

Yu, L. (2005). Basket securities, price formation, and informational efficiency. *Working Paper.*

APPENDIX

APPENDIX A

Appendix A: Success of new ETF introductions by year (Germany)



This table reports the success of global new ETF introductions by vintage. Successful ETF introductions are defined as launches that are able to secure US$ 100 million AuM at any point in their first two years. The total number of product launches per year is also reported at the top of each bar. (Source: Morningstar)

APPENDIX B

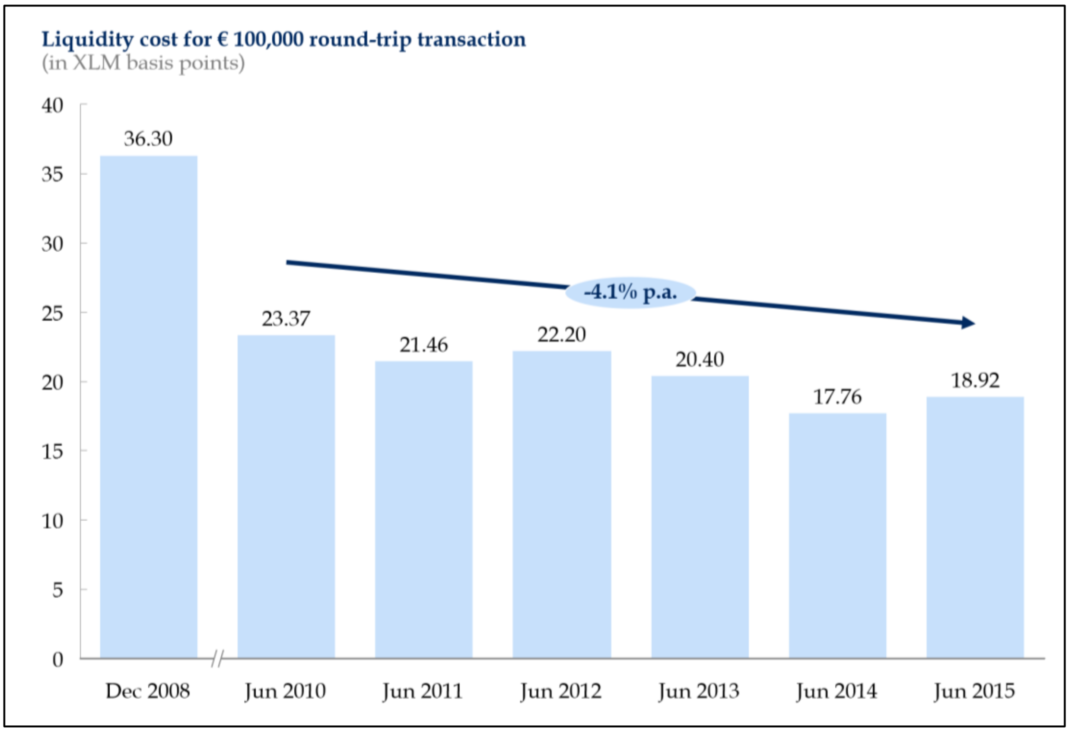
Appendix B: Max. and min. points of impact of independent variables for regression (5)

Variable	**Max./min. point**
TER (min)	-0.3358
XLM (max)	0.4301
DISTR (max)	-0.8589

This table reports the maximum and minimum points of impact of the respective variables on tracking error. The variables with a significant non-linear relationship with tracking error are total annual expense ratio (*TER*), weighted portfolio liquidity cost (*XLM*) and distribution of cash to investors relative to NAV (*DISTR*).

APPENDIX C

Appendix C: AuM weighted average liquidity cost for all ETFs listed on XETRA



This chart reports the change of AuM weighted average liquidity cost (measured in XLM basis points for a hypothetical round-trip transaction of € 100,000) for all ETFs listed on XETRA from 2010 to 2015. For comparison, it also reports the AuM weighted average liquidity cost for 2008. (Source: Deutsche Börse)

APPENDIX D

Appendix D: Minimum requirements for ETFs on XETRA

General Requirements for Quotes	
Minimum volume	Determined by issuer
Maximum spread	Determined by issuer
Requirements in Continuous Trading	
Quotation time	At least 80% (90%) of effective trading time
Requirements in Auctions	
Quote entry	At price determination
Minimum participation rate in planned auctions	80% (90%)
Minimum participation rate in opening auctions only	80% (90%)
Minimum participation rate in volatility interruptions	70% (80%)

This table reports the minimum requirements for ETFs listed on XETRA. The minimum requirements for transaction fee reimbursement are stated in brackets. (Source: Deutsche Börse)

APPENDIX E

Appendix E: ATT of hiring an additional DMM on trading volume and AuM

Matches	Trading Volume			Assets under Management		
	ATT_{HFT} (N=280)	$ATT_{non\text{-}HFT}$ (N=172)	Δ ATT	ATT_{HFT} (N=280)	$ATT_{non\text{-}HFT}$ (N=172)	Δ ATT
M=5	-0.699 (1.662)	-1.027 (1.107)	0.327 (1.997)	4.583 (10.748)	16.283 (13.887)	-11.699 (17.561)
M=10	-0.578 (1.644)	-1.131 (0.991)	0.553 (1.920)	7.794 (10.14)	14.972 (12.285)	-7.177 (15.93)
M=15	-0.567 (1.672)	-1.000 (0.987)	0.433 (1.941)	7.337 (10.575)	13.623 (11.989)	-6.285 (15.987)
M=20	-0.432 (1.698)	-1.275 (0.983)	0.843 (1.962)	6.525 (11.097)	11.829 (11.615)	-5.304 (16.064)
M=25	-0.554 (1.670)	-1.429 (0.986)	0.876 (1.939)	6.259 (11.097)	12.675 (11.502)	-6.415 (15.982)
M=30	-0.511 (1.657)	-1.512 (1.000)	1.001 (1.936)	5.795 (11.082)	12.835 (11.496)	-7.039 (15.968)
M=50	-0.316 (1.637)	-1.444 (0.993)	1.129 (1.914)	7.983 (10.866)	10.919 (10.562)	-2.936 (15.153)

This table reports the average treatment effect of contracting an additional HFT or non-HFT designated sponsor on the treated ETFs (ATT). It is measured as the change in XETRA trading volume and assets under management in millions of euros 12 months after treatment. It also reports the mean differences between the average treatment effects (Δ ATT) of contracting an additional non-HFT designated sponsor and of contracting an additional HFT designated sponsor. Matching is done via a logit-regression based propensity score estimator, as described in Section 3.3.2, and the results are reported for varying control group sizes (M) of 5, 10, 15, 20, 25, 30 or 50 nearest neighbour control group observation matches. Abadie and Imbens' (2012) robust standard errors are reported in parentheses. Statistical significance of the result being different from zero based on a two-tailed test at *10%, **5% and ***1% confidence levels.

APPENDIX F

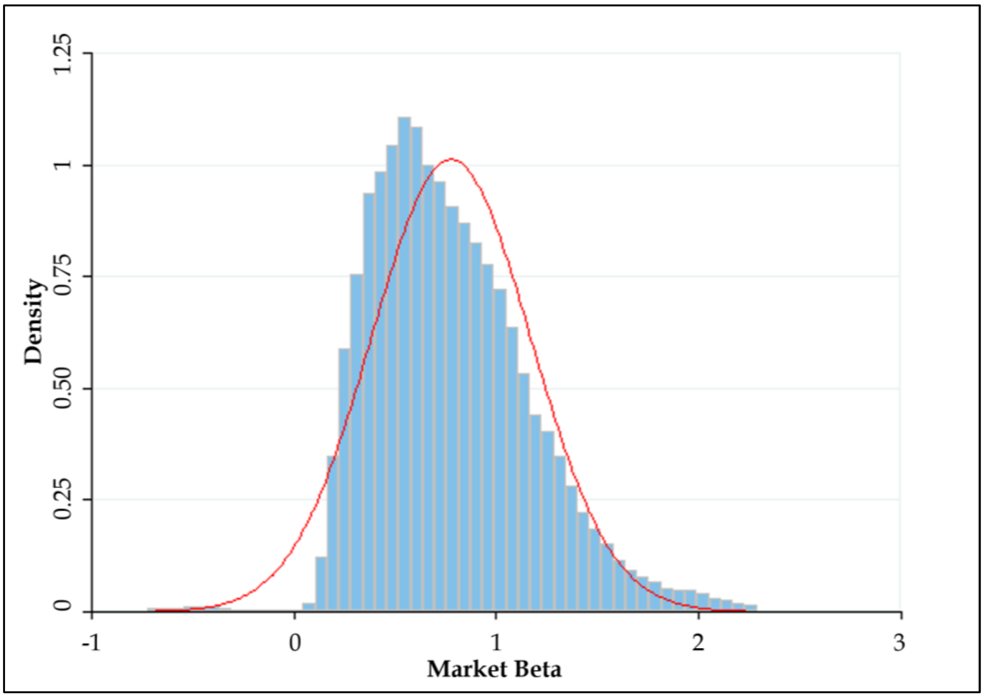
Appendix F: List of ETFs used in the sample and their respective benchmark

Fund name	ISIN	Benchmark
ComStage FR DAX® UCITS ETF	LU0488317024	DAX® (PR)
db x-trackers DAX® UCITS ETF	LU0274211480	DAX® (TR)
db x-trackers DAX® UCITS ETF	LU0838782315	DAX® (PR)
Deka DAX® (ausschüttend) UCITS ETF	DE000ETFL060	DAX® (PR)
Deka DAX® ex Financials 30 UCITS ETF	DE000ETFL433	DAX® ex Financials 30 (PR)
Deka DAX® UCITS ETF	DE000ETFL011	DAX® (TR)
Deka DAXplus® Maximum Dividend UCITS ETF	DE000ETFL235	DAXplus® Maximum Dividend (PR)
Deka MDAX® UCITS ETF	DE000ETFL441	MDAX® (TR)
iShares Core DAX® UCITS ETF	DE0005933931	DAX® (TR)
iShares DivDAX® UCITS ETF	DE0002635273	DivDAX® (TR)
iShares MDAX® UCITS ETF	DE0005933923	MDAX® (TR)
iShares TecDAX® UCITS ETF	DE0005933972	TecDAX® (TR)
Lyxor DAX (DR) UCITS ETF	LU0252633754	DAX® (TR)
Recon Capital DAX Germany ETF	US26923E2072	DAX® (TR)

This table reports the full name (as given in the official prospectus), International Securities Identification Number (ISIN), and respective benchmark index (TR for total return and PR for price return index) for each of the 14 physically replicating ETFs in the DAX index universe.

APPENDIX G

Appendix G: Distribution of sample market beta estimates



The distribution of closing auction return market (MSCI EMU ex Germany) betas estimated for stocks in the final sample (bar chart) after exclusion of insignificant estimates and truncation of extreme outliers at the 0.5th percentile and 99.5th percentile. For comparison, the continuous red line illustrates the normal distribution with similar mean.