Norwegian School of Economics
Bergen, Spring 2018

The Role of ETFs in the Corporate Bond Market

An empirical study of potential impacts of fixed income ETFs on the underlying U.S. corporate bond market

Vetle Rotlid and Yngve Rubi

Supervisor: Nataliya Gerasimova

Master thesis, Master of Science in Economics and Business Administration, Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Preface

This thesis concludes our five years at NHH, and we graduate from this institution with our respective Master of Science in Economics and Business Administration degrees, both majoring in Finance.

The task of writing a thesis has been both challenging and interesting. From the idea formation stage to the statistical analysis, we have had the opportunity to challenge ourselves and expand our knowledge in several areas.

First of all, we want to thank our supervisor Nataliya Gerasimova for constructive advice and feedback throughout all phases of this process. Her help in developing our thesis topic and insights on the academic writing process has been particularly valuable.

The R software has proved to be a very useful tool for us during this semester, and we look back at a steep learning curve of using this programming language for our project. Thus, we want to thank all authors of the R-packages that we have used during this coding adventure. You have made our life a lot easier during the last months in addition to knowledgeable users in various online programming forums.

This journey would not have been possible without the support we have received from our respective families and loved ones. We are also very grateful for all the memories we have with our friends here at NHH, thank you for making these five years a joyful ride.

"The true delight is in the finding out rather than in the knowing"

- Isaac Asimov

Bergen, June 19 2018

Vetle Rotlid

Vetle Rotlid

Yngve Rubi

Yngve Rubi

Abstract

Exchange traded funds (ETFs) have become popular investment vehicles in the U.S. corporate bond market. A market that is characterised by over-the-counter transactions, low liquidity and high trading costs, is with ETFs more accessible to retail investors and arbitrageurs alike. The ongoing trend is raising several questions from both academics and practitioners. A stream of recent publications explores how these new, mostly passive investment vehicles are affecting the liquidity, valuations and other aspects of the underlying markets.

We set out to investigate the effect of ETFs on the commonality of underlying bonds in the U.S. corporate bond market. In our thesis, we examine if different measures of fixed income ETF activity are explanatory factors of commonalities in bond returns, yields, trading volume and illiquidity. Previous research finds that the turnover of ETF shares influences the commonality of individual securities more compared to other ETF activity measures in the equity market. For this reason, we additionally investigate if ETF turnover carries the same relevance in the corporate bond market. In our empirical research we employ naïve OLS, time series and panel regressions to investigate the relationship between ETF activity variables and corporate bond commonality. We include both time and individual fixed effects and various control variables in the models. Additionally, we conduct robustness tests where we add fundamental factors that are potential drivers of bond commonalities in our time series models.

Our empirical findings suggest that there exists a relationship between ETF activity and several commonality measures, indicating that fixed income ETFs may have an influence on the comovement of underlying bonds. In addition, we find turnover to have the most pronounced effect of all the included ETF measures. Implications of ETFs inducing higher commonality could be lower diversification benefits and higher liquidity risk. As fixed income ETFs are experiencing solid growth, further research on their implications is needed.

Contents

List of tables

List of figures

1. Introduction

1.1 Background

The introduction of index funds in the 1970s and exchange traded funds (ETFs) in the 1990s, simplified the process of constructing well diversified portfolios for investors considerably. Passive ownership is reaching higher levels in all asset classes. According to a research report by FTSE Russel (2017), mutual funds and ETFs pursuing passive strategies grew from representing only 12% of managed equity funds in the U.S. markets in 1998 to 46% as of December 2016. Most of the growth in passive ownership during the last 20 years has been driven by ETFs and the trend does not seem to be halting.

The growth in index investing fuelled by ETFs, has initiated a debate over what the possible benefits and risks of the trend are. For individual investors, ETFs offer several benefits due to their low cost, tax efficiency and liquidity. However, there are concerns regarding the effect of increased passive ownership and ETF growth on market behaviour, pricing, liquidity and other factors. Concerns are raised by both practitioners (Martin, 2017) and academics (Wurgler, 2011).

Index-linked or passive investment strategies focus on methodologies by which the portfolio weight usually is decided by a company's market capitalization, or in the case of bonds on the market value of outstanding debt. Such strategies disregard company characteristics as valuation and idiosyncratic risk, which are more important to an active investor. As broad baskets of securities are either sold or bought by investors, higher commonality in price and volume movement among the basket securities can arise. According to the theoretical predictions from modern portfolio theory, consequences of increased correlation among securities, could be lower diversification benefits and higher portfolio volatility (Markowitz, 1952). In addition, a lower dispersion in security returns could reduce the opportunity set for active investors and make stock picking more difficult (Gorman, Sapra, & Welgand, 2010).

While the effects of passive investing in the equity markets are put under more scrutiny, the research conducted on the effects in other asset classes is not as extensive. Research by Sullivan & Xiong (2012), and Bolla, Koller & Wittig (2016) suggest that an increase in index trading leads to stronger comovement of securities in an equity index. Tang & Xiong (2012) investigates correlations in the commodity markets and found that concurrent with the growth of index investing non-energy commodity futures have become more correlated with the oil price.

To our knowledge, similar research has not been conducted in the American corporate bond market. Corporate bonds are traded in an over-the-counter (OTC) market that compared to the equity markets are quite illiquid (Lettau & Madhavan, 2018) and less accessible to retail investors (Schacht, 2016). The emergence of bond ETFs has made it possible to take positions in liquid assets that gives exposure to the market. The illiquid nature of the market could make the underlying securities more sensitive to ETF related trading activity. We want to contribute to the growing research of the effects of index-linked investing by investigating if the growth of ETF activity in the corporate bond markets have effects on commonalities of underlying securities.

1.2 Problem statement and thesis structure

In this thesis, we want to explore how exchange traded fund flows and trading activity influence the return, volume and liquidity comovement of the securities in the underlying corporate bond market. Similar research in the U.S. stock market suggest that ETFs affect commonality of underlying securities, is this also the case in the U.S. corporate bond markets? We perform the analysis by looking at monthly measures of commonality in fixed income securities and measures of ETF activity. Findings are controlled for the effects of fundamental factors that could possibly explain variation in comovement of securities in the underlying market.

Hypothesis 1: Measures of bond ETF activity can explain parts of the variation in the commonality of securities in the underlying market.

We suspect the dependence between activity in ETFs and movement in underlying to be induced by arbitrage activity as previous research suggests (e.g. Sushko & Turner (2018), Da & Shive (2018)). Measures such as ETF turnover, which could be a close proxy to arbitrage activity may therefore explain more of the commonality in the underlying market than other ETF measures.

Hypothesis 2: Turnover of ETF shares have more influence on commonality measures of securities in the underlying market compared to other ETF activity measures.

To test our hypotheses, we create measures for corporate bond commonality and ETF activity. We investigate the link between the measures by applying statistical methods that have been used in similar research such as naïve OLS, time series and panel regressions with fixed effects. In section 5.5, we link our hypotheses to the empirical results from the different models.

Thesis structure

Our thesis is structured into six chapters. In the first chapter, we establish the background for our research topic in addition to provide a problem statement and our hypothesis. In chapter two, we provide a summary of previous research on related topics and background information on how ETFs work and the fixed income ETF market. In addition, we provide information on other theories and topics that are relevant for our analysis, such as comovement, index replication and fundamental factors that could influence bond correlations. The third chapter is the methodology part where we describe our main empirical methods and variables. In chapter four, we describe our data sources, the construction of subsamples and provide summary statistics for our full bond sample. We describe and discuss the results from our empirical research in chapter five, while the conclusion is found in chapter six. In addition, we include a full list of references and the appendix at the end.

2. Theory

The following chapter is divided into four parts. In the first part, we review previous research on the influence of ETFs and other passive investment vehicles on the corporate bond and stock market. In part 2.2, we describe how ETFs work, their dynamics and the history of the fixed income ETF market. The last two sections provide theory on asset comovement and a description of how ETF sponsors use sampling to replicate the market index.

2.1 Literature review

In this part, we present some of the most relevant previous literature on ETFs and their effect on underlying securities in both the bond and equity markets.

Sultan (2015) looks at the relationship between bond ETFs and the liquidity of underlying securities. Findings from the research indicate that ETF ownership has a positive impact on the liquidity of U.S. corporate bonds when only bonds that are already bought and held by an ETF are considered. Nam (2017) examines how the liquidity of underlying securities change when a basket security is introduced. By performing empirical tests on the corporate bond markets before and after the introduction of ETFs, she finds that the resulting liquidity improvements are more pronounced for highly arbitraged, low-volume, high yield, long-term and 144a bonds. Hence, the less accessible the market was before the introduction of ETFs, the more the liquidity of underlying securities improved (Nam, 2017).

Dannhauser (2017) investigates if bond ETFs have valuation effects on underlying bonds. She finds that ETFs have a positive valuation effect on bond index constituents, while she finds ETFs to have an insignificant or negative impact on liquidity, which diverges from the findings from Sultan (2015). Concerning valuation, she shows that an increase in ETF ownership decreases the yield spread of bonds leading to a higher valuation, and she even finds these effects to be permanent rather than temporary. Further, Dannhauser discovers an inverse relationship between ETF activity and liquidity traders' proportion of volume. She argues against improved liquidity being a factor behind the proven lower yield effects, as the ETF impact on liquidity is insignificant for high yield bonds and negative for investment grade bonds. However, she does not reject improved overall liquidity given that investors now have access to invest in more liquid ETF shares.

Clark & Mauck (2014) look at the growth in the fixed income ETF market. They find that U.S. fixed income ETF volume is positively correlated with the VIX index, supporting a notion that uncertainty in the financial markets may lead to increased interest for exchange traded funds. Additionally, they indicate that the increase in ETF trading volume is mainly driven by interest and demand from institutional investors.

Sullivan & Xiong (2012) investigates the relationship between equity ETFs and systematic risk, and find a positive relationship between passive investing and a rise in equity market risk measured by market beta. They find that pairwise correlations and cross-correlations between return volatility and volume volatility have increased significantly since 1997 and show that the diversification benefits have decreased for all styles of equity portfolios. These findings are in large supported by Bolla et al. (2016), but they also look at regional differences given that some markets are more mature with regards to passive investing. They find that large-cap companies in less developed markets display high ETF impact on correlations, which is consistent with the findings in more developed markets. However, in developed markets they also find spillover effects to small-cap firms.

Da & Shive (2018) make use of U.S. equity ETF holdings to document a link between return comovement in stocks and different ETF activity measures. They find that a one standard deviation increase in ETF turnover is associated with a 1% increase in average correlation among the stocks in its portfolio. Interestingly, they find a stronger effect among small stocks with a low turnover. In addition, they propose that ETF activity is related to overshooting and price reversals, which can be a symptom of excess comovement.

2.2 ETFs and dynamics

Since the aim of this thesis is to investigate how fixed income ETFs potentially influence the corporate bond market, it is essential to understand how the products work and in which ways they are linked to the underlying market.

2.2.1 What is an ETF?

Exchange traded funds (ETFs) are basket securities, which means that ETF investors get exposure to baskets of different assets such as stocks, bonds or commodities (Tucker, 2016). ETFs generally follow a passive investment strategy that makes them similar to index tracking mutual funds. Unlike mutual funds, ETFs trade intraday on an exchange and can offer the same convenience and trading ease as listed stocks (Wigglesworth, 2018). For instance, when an investor buys a share of SPY, which is the world's largest ETF she will get exposure to the S&P500 index that includes 500 listed US companies. SPY as most ETFs follows a passive investment strategy, this implies that the purpose of the instrument is to replicate the underlying index in a cost efficient manner and not to outperform it (Tuchman, 2013). The main similarities and differences between ETFs, open- and closed-end mutual funds are summarised in table 2.1.

Table 2.1 Key differences in fund structures, source: Novick et al. (2017)[1]

Feature	Traditional Open-End Mutual Funds	Close-End Funds	ETFs
Exchange-traded	No	Yes	Yes
Visibility into holdings (transparency)	Typically monthly or quarterly	Typically monthly or quarterly	Typically daily
Shares outstanding	Number of shares can change at end-ofday based on purchases and redemption	Supply of shares is fixed	Number of shares can change at end-of-day based on creations and redemptions
Pricing	All transactions are at the fund´s end-of-day NAV	Initial public offering (IPO): IPO price After IPO: market determined	Primary market: NAV Secondary market: market determined
Liquidity	End of day only (Primary processes)	Intraday: subject to market liquidity (Secondary market)	Intraday: Secondary market End of day: Primary processes

2.2.2 ETF history and market overview

The first exchange traded fund was introduced to the market on January 22 in 1993 by State Street Global Investors. This ETF called S&P 500 Trust ETF (ticker: SPY) was designed to track the S&P 500 index (Simpson, 2018). To this day, the SPY ETF remains the largest ETF in the market with $256 billion in assets under management as of May 12. Even though this was the first ETF on the market, the idea and concept of passive investing is older. Both Wells Fargo and American National Bank launched index mutual funds in 1973, while mutual fund guru John Bogle, later known for founding the Vanguard Group, launched First Index

[1] Note: Reprinted from *A primer on ETF trading activity and the role of authorized participants,* by Novick et al.(2017), retrieved from: https://www.blackrock.com/corporate/literature/whitepaper/viewpoint-etf-primary-trading-role-of-authorized-participants-march-2017.pdf, published by Blackrock

Investment Trust tracking the S&P 500 in 1975 (Simpson, 2018). In March 1996, Blackrock's ETF provider iShares launched the first international ETFs (Bojinov, 2015), while State Street's ETF provider SPDR launched the first sector ETFs in December 1998.

It took almost a decade before ETFs investing in other asset classes than equity was introduced. This happened as the first fixed income ETFs were launched in July 2002 by iShares with three funds investing in US treasuries of different maturities and one fund going into corporate bonds (ETF Database, 2018). The first commodity ETF came more than two years later with the GLD ETF launched by State Street Global Advisors, investing in physical gold bullions stored in secure vaults. The first ETNs (exchange traded notes) were launched in June 2006 with the goal of offering investors exposure to commodity futures contracts (Bojinov, 2015). The investment firm and asset manager Charles Schwab debuted with the first commission free ETFs in November 2009.

Since their introduction, the ETFs' assets under management have experienced substantial growth. One and a half decade after the launch of the first ETF total assets under management surpassed $1 trillion in December 2010 (Bojinov, 2015). The fact that it took only additional four years to hit the $2 trillion mark is illustrative to how popular these investment vehicles have become. Table 2.2 provides an overview of all ETP (exchange traded products) assets, based on a report from the world's largest ETF provider BlackRock. According to BlackRock (2017), in December 2017 the assets of all exchange traded-products (both ETFs and ETNs) globally totalled $4.8 trillion. U.S. assets made up more than 70% of this, with European, Asia Pacific and Canadian assets representing 16%, 9% and 2.5%, respectively. Looking at the different asset classes, equity is by far the largest class representing 80% of the assets. Fixed income represents 16%, while commodities and other assets make up 4% of all ETP assets globally.

Table 2.2 ETP assets by listing region, source: BlackRock (2017)

(US$bn)	Equity	Fixed income	Commodity & others	Total
U.S.	2,771	554	93	3,418
Europe	526	182	74	783
Canada	81	33	3	117
Asia Pacific	389	10	26	425
Latin America	6	0	0	6
Middle East & Africa	7	0	2	9
Total	3,781	780	198	4,758

To put the U.S. assets of exchange traded-products in perspective, the total of outstanding debt and equity market capitalisation was $71,944 billion at the end of 2017, implying that ETP assets make up 4.8% (Brandon, Sung, & Podziemska, 2018). The market capitalisation of U.S. equity was $32,120 billion, hence American ETFs' assets under management accounted for 8.6%. These figures illustrate that ETFs are growing into important players in the capital markets.

The ETF providers

Table 2.3 shows that the top three providers BlackRock, Vanguard and State Street Global Advisors have a combined market share measured by assets under management of 81.7%, close to $2.9 trillion in total. Adding the assets of Invesco PowerShares and Charles Schwab, we find that the top five providers account for 90% of the total AUM in the ETF market.

One of the features of index-linked funds is the ability to provide broad exposure at a low cost for investors. According to Riedl (2018) there are economies of scale in the business of offering passive investment vehicle as the marginal cost of replicating an index is low. This could explain the high market share of a few ETF providers. However, whether this oligopoly is a healthy property of ETF market is up for debate.

Table 2.3 Largest US ETF providers as of May 10 2018, source: ETF.com (2018)

ETF provider	AUM (US$bn)	AUM share
BlackRock	1,393.0	39.5 %
Vanguard	880.6	24.9 %
State Street Global Advisors	611.3	17.3 %
Invesco PowerShares	182.4	5.2 %
Charles Schwab	109.0	3.1 %
First Trust	64.6	1.8 %
WisdomTree	43.9	1.2 %
VanEck	35.6	1.0 %
Proshares	30.3	0.9 %
Northern Trust	16.9	0.5 %
U.S. total	*3,530.1*	
Top 3 providers	2,884.9	81.7 %
Top 5 providers	3,176.3	90.0 %
Top 10 providers	3,367.6	95.4 %

2.2.3 Fixed income ETFs

Fixed income ETFs has been one of the fastest growing categories of ETFs in recent years reaching $780 billion of total assets under management (AUM) globally in December 2017 (BlackRock, 2017). One of the most popular categories of fixed income ETFs are corporate bond ETFs[2]. Lettau & Madhavan (2018) accredits the strong growth in this particular ETF category to four major factors:

1) A number of corporate bonds are primarily traded in over-the-counter markets, which are both illiquid and lack transparency. Bond ETFs, however, trade intraday on liquid electronic exchanges and several have much lower bid-ask spreads than the underlying securities. This leads to lower costs and risks for potential investors (Hendershott & Madhavan, 2015).

2) In contrast to individual bonds, fixed income ETFs are highly transparent. Bid and offer quotes in ETFs are freely available.

3) It is generally cheaper and easier to construct diversified portfolios with bond ETFs compared to individual bonds due to lower transaction costs and the ETFs being well-diversified basket securities.

4) Several investors are interested in keeping the maturity of their bond portfolios constant. This requires constant trading if executed with individual bonds while an ETF can be designed to automatically maintain a fixed maturity.

In the last years, institutional investors have flocked into bond ETFs and the number of block trades (trades over 10 000 shares) doubled in the period 2010-2016 reaching 25% of the value traded in bond ETFs (Banerji, 2017). Institutional investors like pension funds, insurance companies and hedge funds are important players in the corporate bond market and the entrance of these institutions in the ETF market, could fuel further growth in this asset class in the future. According to a bond study among institutional investors by Greenwich

[2] Corporate bond ETFs are investment vehicles that give investors exposure to the whole corporate bond market or a special segment, e.g. high yield. Aggregate bond ETFs invest in the broad market, i.e. both government and corporate bonds (ETF.com, 2018).

Associates (2016), 33% of the survey respondents stated that they planned to increase their usage of bond ETFs in the future.

Even though the corporate bond ETFs have seen an increasing popularity among investors, it is far from the only type of fixed income ETFs. Figure 2.1 illustrates the American fixed income ETF market by the nine different categories and their market share in terms of number of ETFs as of 2017 (Bloomberg, 2018).

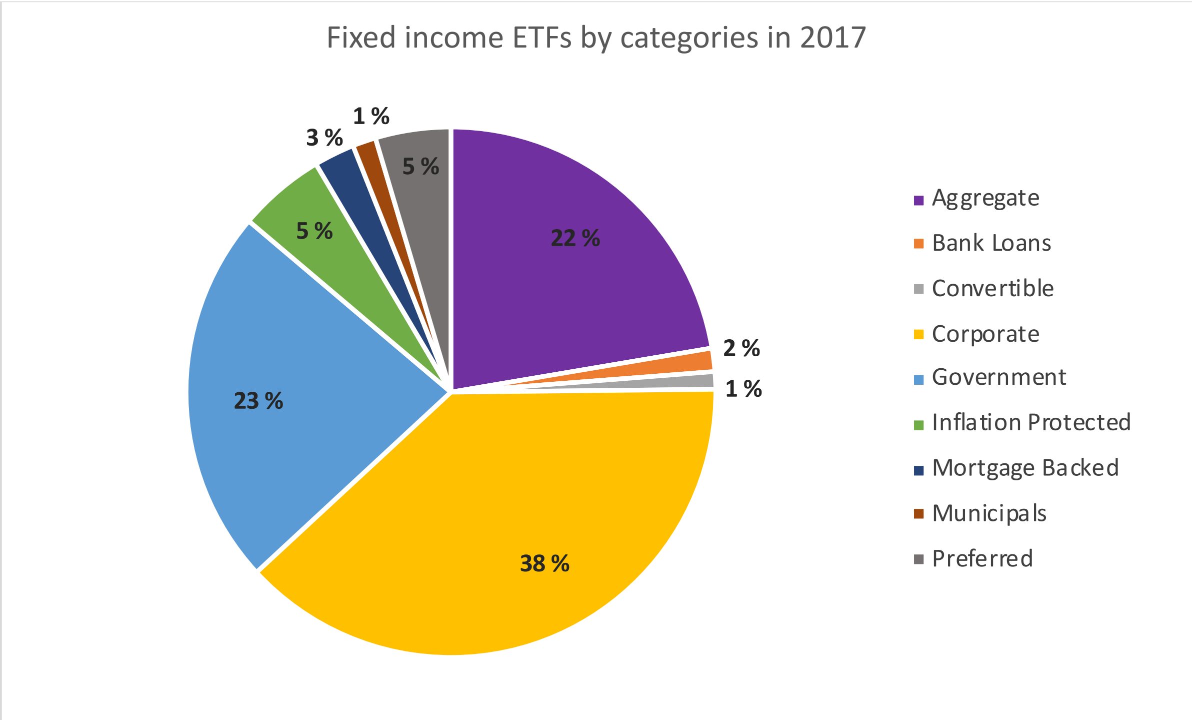


Figure 2.1: Fixed income ETFs by category, source: Bloomberg (2018)

These nine types of fixed income ETFs are divided into categories based on their investment mandate and scope of securities selection. The largest type is by far corporate bond with a share of 38% in numbers of different ETFs. Next, we find government bond and aggregate bond ETF investing in the broad market with shares of 23% and 22%, respectively. Hence, more than 80% of all ETFs invest in either corporate bonds, government bonds or both.

The rest of the ETFs invest in somewhat more specialised securities. Inflation protected securities do what its name says, provide protection against inflation. This is done by having a principal payment that increases with inflation and decreases with deflation (U.S. Department of the Treasury, 2013). The funds invest in TIPS (Treasury Inflation Protected Securities), which provide investors with inflation protection. This category represents five percent of existing fixed income ETFs. Preferred ETFs are funds focused on investing in preferred shares that have a higher priority in dividend claims than common stocks and may therefore be considered safer (ETF.com, 2018). These funds provide exposure to such securities and can serve as a complement to a fixed income ETF portfolio, especially in a low interest rate environment (McCullough, 2017).

ETFs investing in mortgage-backed securities represented three percent of the ETFs in 2017 and offer investors exposure to mortgage payments on both commercial and residential property (Maverick, 2018). ETFs investing in senior bank loans represented two percent and provides exposure to more risky bank loans given to corporations that in turn are bundled into bonds (Li, 2018). The two smallest categories of fixed income ETFs in 2017 were municipal bonds and convertible bonds, each representing one percent of all fixed income ETFs. Municipal bonds are bonds issued by government on the state, municipality or county level to finance capital expenditures on public goods such as infrastructure or schools (ETF.com, 2018). Convertible bond ETFs give investors exposure to debt securities that can be turned into equity at the discretion of the bondholder (ETF.com, 2018).

Figure 2.2 below illustrates the growth in number of ETFs for the different categories since the inception of the first fund in 2002.

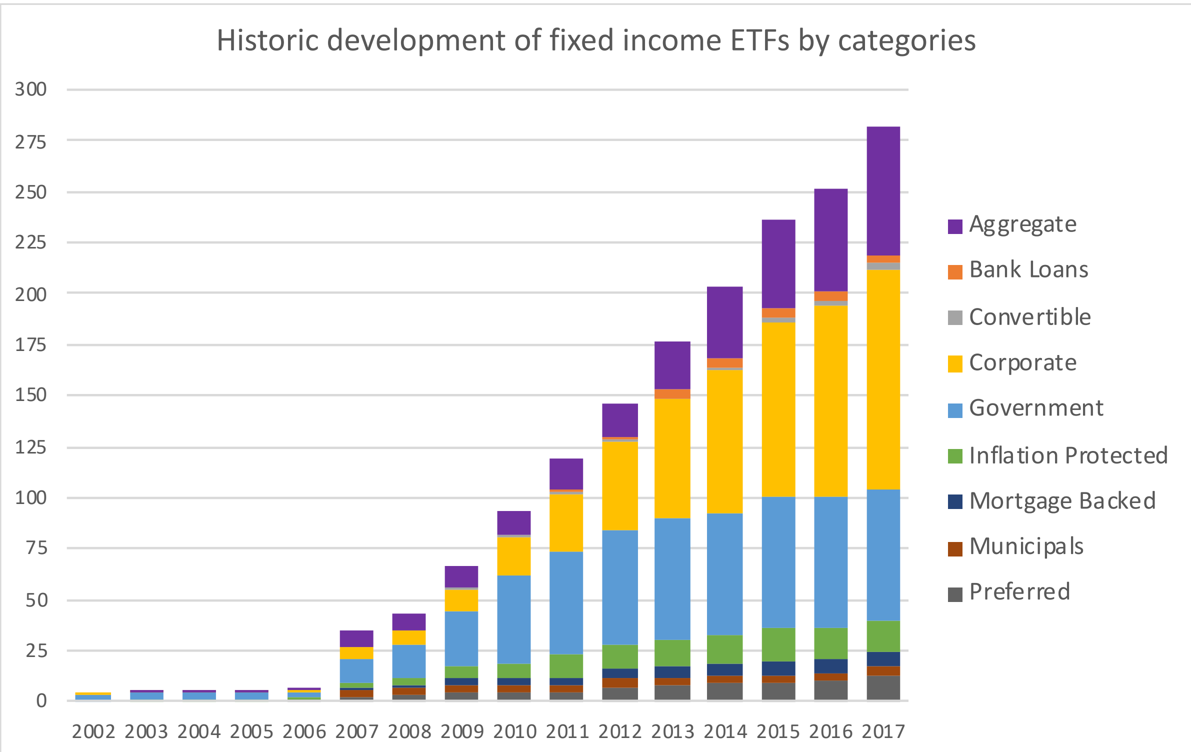


Figure 2.2: Growth in fixed income ETFs by category, source: Bloomberg (2018)

Between 2002 and 2006, the growth was not impressive with the inception of only three ETFs. The market consisted of three government, one corporate and one aggregate bond ETF in 2006, in addition to one preferred and one inflation protected bond ETF. However, the growth in these investment vehicles seems to lift off in 2007. The growth was mostly present in government bond ETFs, but both aggregate and corporate bond ETFs experienced a significant increase in number of funds. In 2008, the growth was rather modest. This is likely linked with the Great financial crisis in 2007-2008. As investors are more cautious during market downturns, attracting capital to new ETFs may be difficult. As the markets improved in 2009, so did the growth in fixed income ETFs with a growth rate of 53%[3], which is the second highest growth rate throughout this period with only 2007 as a stronger year. Post-crisis, the growth of alternative fixed income ETFs soared, especially inflation protected, preferred and mortgage-backed securities. These categories went from only five funds in 2007, to 24 ETFs by 2012. In 2014, corporate bond ETFs took the lead in the fixed income ETF market as the category with the highest number of ETFs, passing government bond ETFs. The growth in government bond ETFs has been rather stagnant since 2013, while the growth in corporate bond ETFs have remained strong until the end of 2017. Aggregate bond ETFs investing in the broad market have also experienced solid development after the global financial crisis growing from eight ETFs in 2007 to 63 ETFs ten years later. End-of-year in 2017, there were in total 282 American fixed income ETF which is a considerable increase from only 35 ETFs in 2007, so there is no doubt that investors have shown a lot of interest in these investment vehicles over the last decade.

Table 2.4 Number of fixed income ETFs

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Aggregate	0	1	1	1	1	8	8	10	12	15	16	24	35	43	51	63
Bank Loans	0	0	0	0	0	0	0	0	0	1	2	4	4	5	4	4
Convertible	0	0	0	0	0	0	0	1	1	1	1	1	1	2	3	3
Corporate	1	1	1	1	1	6	7	11	19	29	43	58	71	86	93	108
Government	3	3	3	3	3	12	17	27	44	50	56	60	60	64	65	65
Inflation Protected	0	1	1	1	1	2	3	6	7	12	12	13	13	16	15	15
Mortgage Backed	0	0	0	0	0	1	1	3	3	3	5	5	6	7	7	7
Municipals	0	0	0	0	0	4	4	4	4	4	4	4	4	4	4	4
Preferred	0	0	0	0	1	2	3	4	4	4	7	8	9	9	10	13
Total	4	6	6	6	7	35	43	66	94	119	146	177	203	236	252	282

[3] See Appendix A for tables of growth rates in number of ETFs

While the number of exchange traded funds investing in debt securities is a relevant measure to look at, it is possibly more relevant to look at the assets under management of these funds to perceive how large this market is.

We observe in figure 2.3 that the development in AUM looks exponential, at least after a more troubling 2013. The distribution of AUM among the different categories appears to be similar to the number of ETFs discussed above. However, there are some exceptions. For instance, aggregate bond ETFs was the leading category in AUM in 2017, with approx. $16 billion more than corporate bond ETFs. Perhaps better shown in table 2.5, we find that aggregate ETFs surpassed corporate ETFs in 2015, while government ETFs was the leading category from the inception of bond ETFs in 2002 until corporate became largest in 2009.

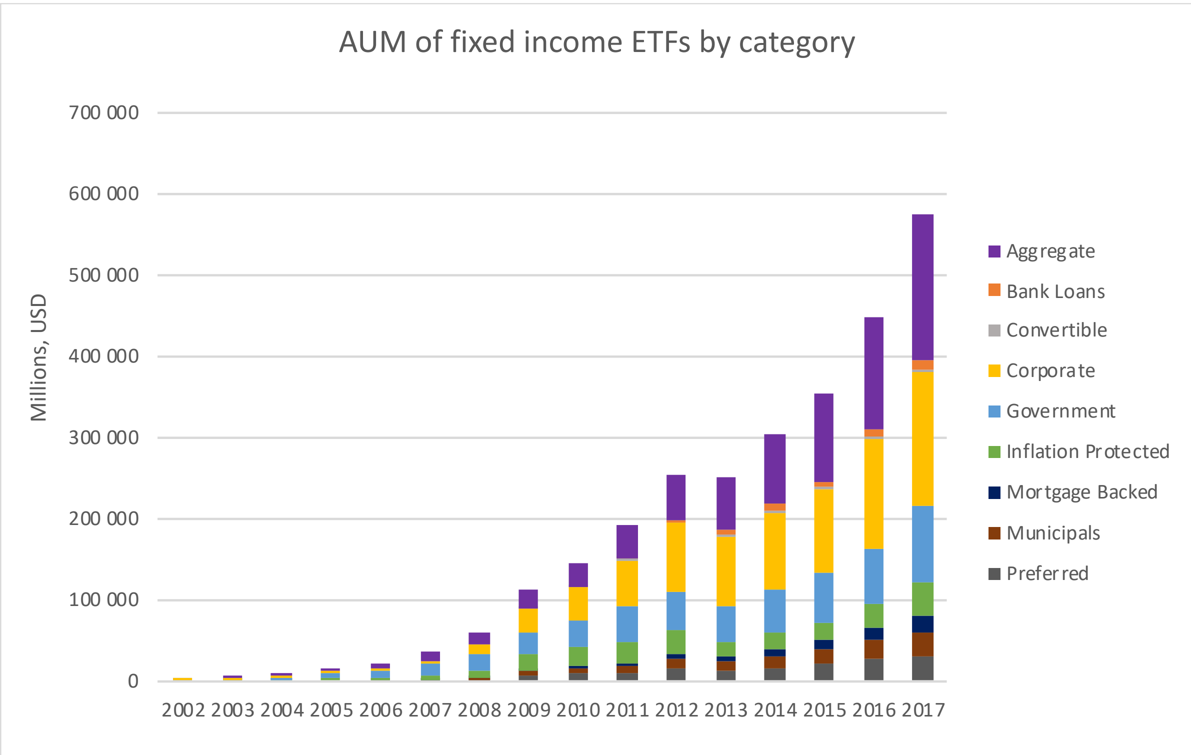


Figure 2.3: Fixed income ETFs' assets under management, source: Bloomberg (2018)

The growth of AUM in 2017 was highest among government (42%), convertible (41%) and mortgage-backed securities (40%). However, the growth was at least 15% for all categories, so the growth appears to be broad. Since the inception of bond ETFs, the growth has mostly proven positive. The total growth rate was only negative in 2013, even during the Global financial crisis of 2007-2008 assets under management increased markedly among fixed income ETFs. In 2009, the total growth in assets under management of fixed income ETFs was 86%, which is the highest growth rate recorded since the inception of these investment vehicles in 2002. Table 2.5 shows the total assets under management for each of the different fixed income ETF categories. As of 2017, the American fixed income ETFs had assets under management of $575 billion (Bloomberg, 2018). This was up 29%[4] from 2016 and represented almost a doubling of the assets under management in 2014.

Table 2.5 Fixed Income ETFs assets under management (Million, USD)

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Aggregate		215	993	2,907	4,925	9,736	15,128	23,360	29,675	42,461	55,881	62,265	86,038	107,406	136,088	178,531
Bank Loans										195	1,539	7,316	6,792	5,498	10,339	12,289
Convertible								234	538	680	923	2,023	2,860	2,533	3,161	4,465
Corporate	1,887	2,311	2,521	2,423	2,731	3,940	11,351	28,432	39,682	56,765	84,444	86,166	95,526	103,917	133,441	162,435
Government	1,996	2,008	3,469	6,388	8,891	14,909	19,346	25,968	34,470	42,900	47,370	42,929	52,301	60,759	67,833	96,578
Inflation Protected		142	1,513	3,311	4,001	5,373	9,236	20,107	21,718	26,240	28,248	20,002	20,276	22,423	31,999	40,852
Mortgage Backed						183	860	1,809	2,295	4,234	6,916	5,784	8,053	9,833	13,268	18,561
Municipals						543	2,196	5,916	7,270	8,767	12,495	10,930	14,667	18,790	24,831	30,726
Preferred					28	201	1,669	5,389	9,310	9,899	15,195	12,169	16,497	21,137	26,094	30,089
Total	3,883	4,676	8,496	15,029	20,576	34,886	59,787	111,215	144,959	192,141	253,011	249,584	303,011	352,295	447,055	574,524

To get an idea of how large the trading in fixed income ETFs is, figure 2.4 illustrates the monthly dollar volume of trade in these investment vehicles that we have in our sample, collected from the Center for Research in Security Prices (CRSP).

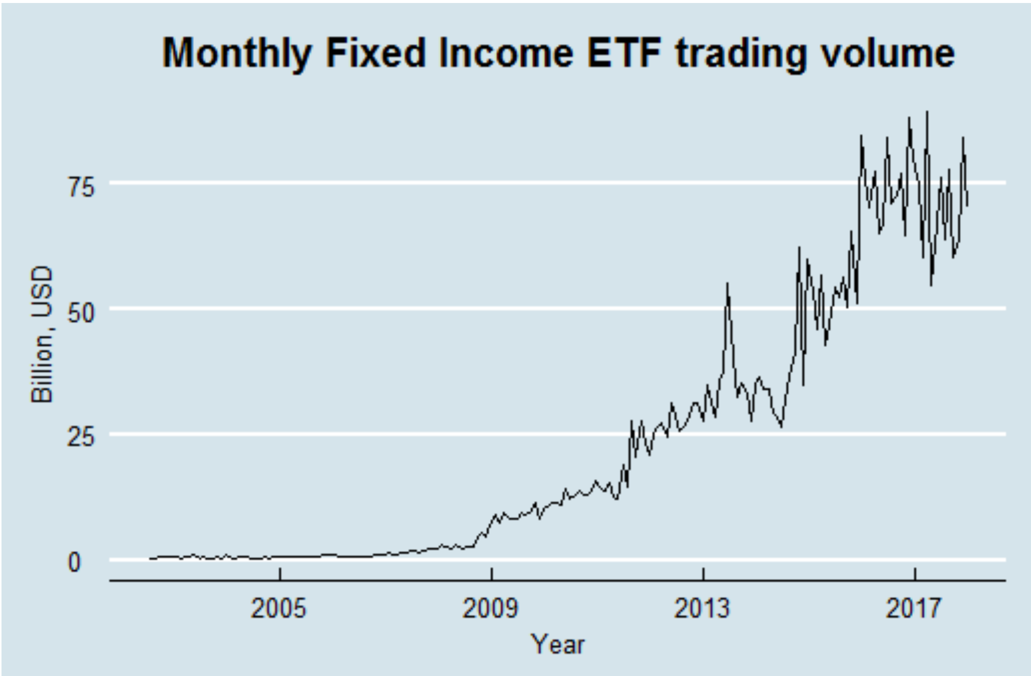


Figure 2.4: Monthly trading volume of bond ETFs in sample, source: CRSP (2018)

[4] See Appendix A for table of growth rates in ETF AUM.

We observe that for a long time the trading volume of fixed income ETFs was low, but since the aftermath of the financial crisis around 2009, the growth accelerated. The growth rate has remained strong with a trading volume that has tripled from 2013 to 2017. Today, the monthly trading volume is approximately $75 billion.

2.2.4 Creation/redemption cycle

A unique feature of ETFs compared to mutual funds is the creation and redemption mechanism. The shares of an ETF can be created or redeemed at the end of each trading day in exchange for a basket of securities from the underlying index. This transaction is performed exclusively between the ETF creator also called the sponsor and market participants called authorised participants (AP). The market where this mechanism occurs is commonly referred to as an ETFs primary market and it is solely through the creation/redemption process that the number of outstanding ETF shares change (Pan & Zeng, 2017). The ETF secondary market is the venue where ETF shares trade intraday by both APs and all other investors.

According to a survey by ICI in 2015, the average number of APs for a U.S. domiciled bond and hybrid ETF is 32 (Antoniewicz & Heinrichs, 2015). APs are typically either large financial institutions or specialised market makers. To perform their objective it is important that the APs have extensive trading experience in the underlying market (Novick et al., 2017). The APs' role in the market is to function as a liquidity provider by having the ability to change the supply of ETF shares in the market. Institutional investors that are interested in buying a large block of ETF shares could for instance contact APs, which in turn are able to facilitate the purchase by creating the requested shares in the primary market. In this case, the institutional investor pays the AP in cash or securities. The AP will then deliver a basket of the underlying securities to the ETF sponsor in exchange for new ETF shares (a creation) which in turn are handed to the institutional investor. Conversely, in a redemption process, the AP exchanges ETF shares for the underlying basket of securities and the ETF creator eliminates (redeems) the shares. Figure 2.5 illustrates the creation/redemption mechanism.

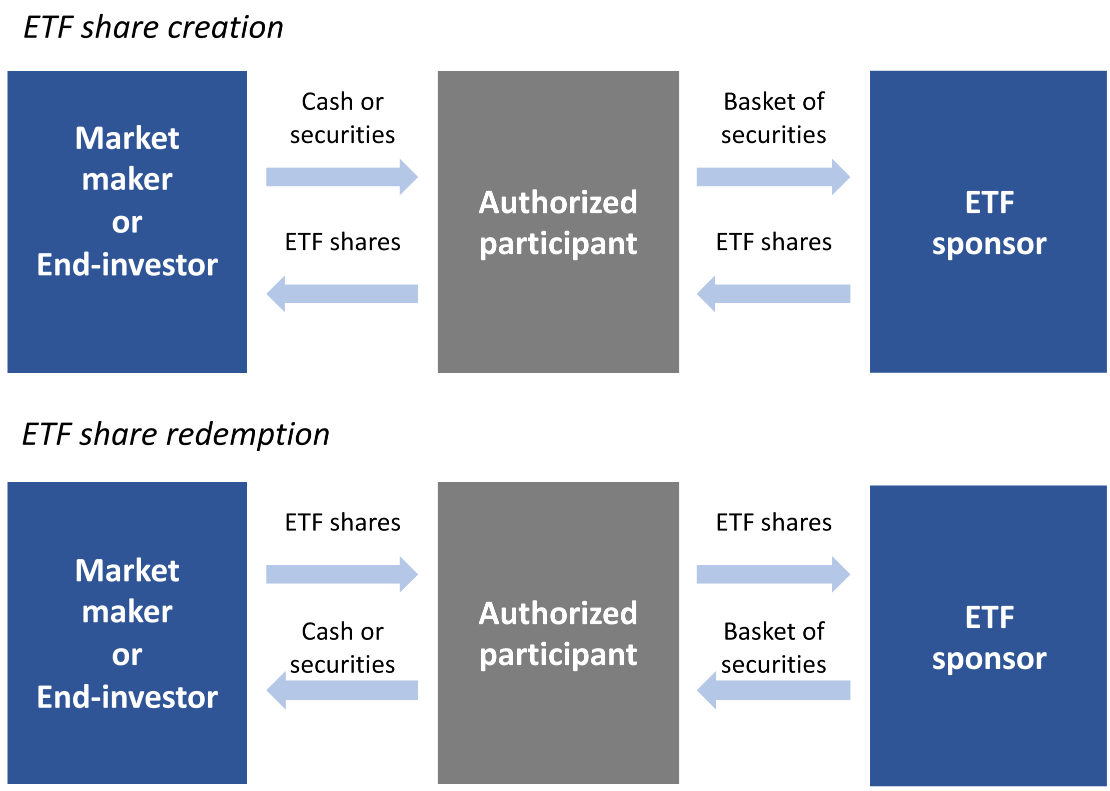


Figure 2.5: Creation/redemption cycle, source: Novick et al. (2017)[5]

APs also have the possibility to perform arbitrage trading. For instance, if the ETF share price is trading above the underlying basket value after transaction costs, the APs can make an arbitrage profit. In this situation, the AP could buy the underlying securities and exchange the basket for ETF shares in the primary market at the end of the trading day. An alternative way to earn an arbitrage profit in the secondary market will be to short the ETF shares and buy the underlying securities. These arbitrage mechanisms make sure the ETF market price keeps close to the value of underlying holdings. We suspect that the arbitrage mechanisms connected to ETFs could lead to increased commonality in the movement of the basket securities. Findings by Da & Shive (2018) and Grant & Turner (2018) suggest that ETF could generate an impact on underlying assets by facilitating conditions for secondary market arbitrage.

[5] Reprinted from *A primer on ETF trading activity and the role of authorized participants,* by Novick et al.(2017), retrieved from: https://www.blackrock.com/corporate/literature/whitepaper/viewpoint-etf-primary-trading-role-of-authorized-participants-march-2017.pdf, published by Blackrock

2.3 Comovement

Barberis, Schleifer & Wurgler (2002) propose three different views on sources of security comovement. The *Fundamentals* view explains that comovement arises from positive correlations in the determinants of a securities value, e.g. cash flows and discount rates. The *Fundamentals* view is based on an economy without friction and with rational investors were the price of an asset equals its fundamental value. This view explains why the security price of companies operating in the same sector move closely since the cash flows and risks of these securities is similar in nature. A number of papers presents evidence that the traditional fundamental view of comovement is incomplete (e.g. Bodurtha Jr., Dong-Son, & Lee (1995) and Pindyck & Rotemberg (1990)). Froot & Dabora (1999) studied Siamese twin stocks that are securities with claims to the same cash-flow stream that are traded in different locations. They find that these securities tend to co-move more with their local market index than their Siamese twin does. Similar findings suggest that investor trading patterns can also be an important source of determining comovement.

The first of the trading induced models that is analysed in Barberis et al. (2002) is the *Category-based* view, which occurs when investors categorise securities into different asset classes (e.g. stocks and bonds) and shift resources in and out of these asset classes in correlated ways. The comparable *Habitat-based* comovement arises when a group of investors move in and out of a specific set of securities in tandem. The specific habitat could be an investor's home country due to the investor having more information about their local market or lower trading costs. The comovement resulting from ETF and index mutual funds flows may be a form of trading induced view since investors track specific categories or habitats of securities.

2.3.1 Fundamental factors

In our analysis, we control potential findings for fundamental effects that could possibly drive comovement in the corporate bond market. We include four different variables that may affect commonality in order to separate trading induced and fundamental effects. The four variables we use are investor sentiment, inflation risk, interest rate risk and credit risk. In this part, we briefly explain why the following factors could drive comovement of corporate bonds and how we calculate the different variables.

Investor sentiment

Investor sentiment can be defined as the optimism or pessimism about financial securities and is transmitted to the market through individual investors transactions (Baker & Wurgler, 2006). An explanation of how sentiment drives comovement in the corporate bond market is investigated by Bethke, Gehde-Trapp & Kempf (2017). Investors with bad sentiment are more sensitive to negative information and avoid risky assets. When sentiment is bad, investors are less prone to invest in bonds with higher credit risk, which means that these bonds become less liquid compared to when sentiment is good. In turn, this implies that liquidity risk premiums increase more with credit risk premiums when sentiment is bad, higher risk factor correlation may translate into higher bond correlation. As a gauge for investor sentiment, we use the CBOE VIX index, a measure of the implied volatility of options on the S&P 500. VIX reflects investors' expectations of future market volatility and is widely used as a proxy for investor sentiment (e.g., Kurov (2010) and Smales (2015)). In our robustness tests in section 5.4.1, we calculate the percentage change in the end of month adjusted closing price of the VIX index.

Inflation risk

Inflation risk is sometimes called purchasing power risk and is the risk that the yield of bonds will not keep pace with inflation (FINRA, 2018). If an investor buys a bond and the rate of inflation rises, the purchasing power of the coupon payments is reduced if it is not inflation protected. Most bonds are exposed to some form of inflation risk. An exception is TIPS, as described in section 2.2.3. Brandt & Wang (2003) present unexpected inflation as a measure of time-varying risk aversion. According to the authors, risk aversion can lead to higher prices of risk from all sources that in turn might lead to comovement of securities. Therefore, we are interested in controlling our results for this effect. We use a method from Bethke et al. (2017) to compute the monthly unexpected return for each month with the following regression:

$$Inflation_t = \alpha + \beta_1 Inflation_{t-1} + \beta_2 Inflation_{t-2} + \varepsilon_t \qquad (2.1)$$

Where we use the residuals from the regression as a measure of the unexpected inflation. We retrieve monthly inflation data from the Consumer Price Index (CPI) published in the FRED database.

Interest rate risk

Interest rate risk is the risk that changes in the interest rate level may reduce or increase the value of a bond (FINRA, 2018). There is an inverse relation between the price of bonds and the market interest rate. If a bondholder owns a bond paying a coupon of 4% and the interest rates in the market rise, new bonds become relatively more attractive to investors. In turn, this leads to a reduction of the value of the old 4% coupon bonds. Changes in interest rates influence the pricing of all fixed coupon corporate bonds independent of a bond’s rating and other features. Therefore, comovement in corporate bonds may be induced by changes in the interest rate levels. To control for this effect, we employ time series of 5-year constant maturity treasuries, since these instruments carry close to no credit risk and closely match the average duration of 5.6 in our bond sample. We adopt a similar method to Bethke et al. (2017) and use the monthly changes in yield of the constant maturity 5-year treasuries as a proxy for interest rate risk.

Credit risk

When investing in bonds, investors are taking a risk on the issuers ability to pay interest on the agreed upon dates and repay the principal (FINRA, 2018). Most bonds face a probability of default, which could mean delayed interest payments or in a worst-case scenario a loss of the bondholder’s principal. Since credit risk is a risk factor for all corporate bonds, changes in the level of credit risk perceived by the market can be a possible explanatory variable for changes in the comovement of individual corporate bonds. We use credit default swap (CDS) indices to control for potential comovement effects driven by the perception of credit risk in the market. A CDS contract is a contract between two parties: A buyer who is paying fixed periodic payments for a credit insurance on a corporation or sovereign entity´s debt, and a seller who collects premiums in exchange for making the buyer whole in the case of default or other credit events (Markit Group ltd, 2008). The indices represent the average protection premium (spread) of the most liquid bonds in the investment grade and high yield market (Markit Group ltd, 2008). Higher (lower) premiums indicate higher (lower) credit risk and this could in turn translate into either lower or higher bond prices. Since credit risk is a bond specific (idiosyncratic) risk factor, changes in the perceived factor risk could e.g. influence the dispersion of corporate bond returns. For our credit risk measure based on CDSs, we use the monthly change of the credit spread levels of the CDX investment grade and high yield index from Markit.

2.4 Index Replication

Funds follow different methodologies when it comes to how they replicate a target index. In deciding which methodology to use the fund provider has to make a trade-off between the tracking error of the fund and transaction costs.

When it comes to replication of an index there are generally three groups of methodologies, *Full replication*, *Sampling* and *Optimisation* (Vanguard Group, 2018). Under *Full replication,* the index is replicated by buying the index constituent securities relative to their weight in the target index. This is a common replication technique when tracking indices with few constituents in liquid markets (e.g. S&P 500). Under some circumstances full replication that yields the lowest tracking error, is not possible due to the target index having many illiquid constituents that are difficult and expensive to trade.

In *Sampling* replication, the fund holds a representative sample of the index constituents. Most bond ETFs replicate their target index with a sampling technique. Managers can replicate the target index by matching bond characteristics such as average duration, sector allocation and rating. According to a study performed by MSCI, bond ETFs tracking error varies widely between different types of bond ETFs (Sparks, 2018). The study shows that high yield ETFs has the highest tracking error at 67 basis points while investment grade funds have a significantly lower tracking error at close to 10 basis points in 2017. The difference may partly be explained by the difficulty to replicate the high yield index due to lower liquidity and higher trading costs. As mentioned, ETF sponsors might look at bond specific factors when replicating the benchmark index. This is something we take into consideration when creating our panel model as explained in 3.2.1.

The last replication technique is *Optimisation* where the ETF sponsor use quantitative multifactor models instead of industry and security characteristics to optimise index tracking.

3. Methodology

In this chapter, we describe the methods and considerations we have taken in the empirical investigation of our research question. In the first part of this chapter, we describe the time series approach, while in section 3.2 we focus on panel data. Under section 3.3 and 3.4 we describe the correlation measurements and ETF activity variables which are the dependent and explanatory variables in our analysis.

We initiate our analysis by calculating the Pearson correlation of the dependent and explanatory variables. We also perform naïve OLS with the same variables to explore the relationships in our data further. Since the Pearson calculations and naïve OLS is not a part of our main analyses and assumed known to the reader, we will not describe these methods. Our main analyses consist of two parts: time series and panel data estimation. First, we use time series estimation to investigate the relationship between the U.S. corporate bond market and the growth of fixed income ETFs using a range of bond commonality measures as dependent variables and different measures of ETF activity as explanatory variables. The bond market variables are calculated on an aggregate level for investment grade and high yield bonds and aim to depict different types of comovement in e.g. returns, trading volume, yields and liquidity between individual bonds. Second, we make use of panel data as an alternative approach to investigate the impact from the growth in ETFs, where the aim is to get a more precise picture of the effects on bonds with a panel of bonds that are owned by two of the largest corporate bond ETFs in the U.S. market.

3.1 Time series

To examine the effects on an aggregate level in the bond market, we use time series regression. Using time-series data enables the researcher to investigate dynamic effects between x and y, i.e. effects between two variables across time. However, having time-series data that are repeated recording of the same variable throughout a given period of time, there are several properties of the data that must be considered when conducting regression analysis (Wooldridge, 2016, p. 7). Such properties comprise the relationship between different points in time of a variable and how this affect the error term in a regression, i.e. autocorrelation. Further, it is not given that a series have a constant mean and variance across time, also called stationarity. The issue of having non-stationary properties in both the dependent and independent variable, may lead to spurious regression and false conclusions about a relationship, especially if there is a common underlying trend that both variables follow (Wooldridge, 2016, p. 346). Hence, such properties must be accounted for in building the estimation model.

3.1.1 Stationarity

Stationarity is a key property in building a time series model with estimates of coefficient ($\widehat{\beta_i}$) that remain constant over time. Wooldridge (2016, p. 345) defines a stationary time series process to be: "one whose probability distributions are stable over time", meaning that at any point in time the probability distribution of possible values must remain unchanged. More formally, this can be stated as a stochastic process $\{x_t: t = 1, 2, ...\}$ is stationary for every collection of time indices $1 \leq t_1 < \ldots < t_m$, the joint distribution of $(x_{t_1}, x_{t_2}, ... x_{t_m})$ is the same as the joint distribution of $(x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_m+h})$ for all integers $h \geq 1$. One implication is that for any choice of *m* and *t*, e.g. *m* = 1 and *t* = 1, x_t has the same distribution as $x_1$for all *t* = 2, 3, …. This implies that the sequence $\{x_t: t = 1, 2, ...\}$ is identically distributed. In addition, stationarity requires that the joint distribution of (x_1, x_2) must be the same as for (x_t, x_{t+1}) for any $t \geq 1$. The correlation between the two adjacent terms may be high, but it must be the same for any adjacent terms across all time periods (Wooldridge, 2016, p. 345).

A weaker form of stationarity is called covariance stationary. This type of stationarity focuses only on the first two moments of a stochastic process and a process is covariance stationary if

i. $\mathrm{E}(x_t)$ is constant
ii. $\mathrm{Var}(x_t)$ is constant
iii. $\mathrm{Cov}(x_t, x_{t+h})$ depends only on *h* and not *t*, for any *t, h* ≥ 1

Further, the concept of weak dependence is also important as this restricts how strong the relationship between two random variables x_t and x_{t+h} can be as *h* increase. To relate this to covariance stationary processes, such a process is said to be weakly dependent if the correlation between x_t and x_{t+h} drops quickly enough towards zero as $h \rightarrow \infty$. The importance of weak dependence is that it replaces the assumption of random sampling in implying that the law of large numbers and the central limit theorem hold (Wooldridge, 2016, p. 346). The central limit theorem for times series data implies that weakly dependent time series that also are stationary, are ideal for conducting multiple regression analysis (Wooldridge, 2016, p. 346).

In order to investigate whether a time series is stationary, one may begin by looking at the autocorrelation function (ACF) plot. In this plot, the ACF will drop relatively quickly to zero for a stationary process, while it decreases rather slowly for non-stationary data. Although this might give an indication of whether one has stationarity in the different series, a more formal approach is to conduct unit root tests on the variable. Our choice of tests for unit root is the Augmented Dickey-Fuller (ADF) test with a null hypothesis of non-stationarity, supplemented by the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) with a null hypothesis of stationarity.

Augmented Dickey-Fuller test

This procedure for testing whether a time series is non-stationary was originally developed by David A. Dickey and Wayne A. Fuller in 1979. We use an extended version of the original model, which allows for the inclusion of a trend component, a drift and multiple lags. However, the original test with no trend, drift nor multiple lags is stated below:

$$\Delta y_t = \theta y_{t-1} + e_t, \tag{3.1}$$

where

$\theta = \rho - 1$ and ρ comes from the autoregressive model of order one, AR (1):

$$y_t = \alpha + \rho y_{t-1} + e_t,$$

and

$$H_0: \theta = 0$$
$$H_1: \theta < 0$$

If one fails to reject the null hypothesis, one cannot infer that the series does not have a unit root ($\rho = 1$) (Dickey & Fuller, 1979). If H_0 is rejected at e.g. the 5% significance level, stationarity is assumed. As stated above, there are several extensions to this original test where additional characteristics can be accounted for (Wooldridge, 2016, p. 576). These are presented below:

Augmented Dickey-Fuller with drift term (α) around a non-zero mean:

$$\Delta y_t = \alpha + \theta y_{t-1} + e_t, \tag{3.2}$$

Augmented Dickey-Fuller with trend component (δ):

$$\Delta y_t = \alpha + \delta t + \theta y_{t-1} + e_t, \tag{3.3}$$

and Augmented Dickey-Fuller with multiple lags:

$$\Delta y_t = \alpha + \theta y_{t-1} + \gamma_1 \Delta y_{t-1} + \gamma_2 \Delta y_{t-2} + e_t \tag{3.4}$$

We conduct the Augmented Dickey-Fuller tests using selected packages in R, such as *urca*, *CADFtest* and *tseries*, where we have the opportunity to use the different specifications of the ADF test. Most variables in our data are not stationary in levels. Hence, it is necessary to transform the variables in order to obtain stationarity. This is done by taking the first-differences of the variables, before we run the ADF tests again to determine whether the level variables are integrated of order one, I(1). This is the case for the majority of the variables, and hence the time series regression contains first-differenced variables for the sake of avoiding spurious regressions. In addition, the lag order for each variable is determined using the embedded Akaike Information Criteria function in the ADF tests. The relevant lags are then included in the regressions.

Kwiatkowski-Phillips-Schmidt-Shin test

We make use of an alternative test in cases where the conclusion from the ADF test is not clear and in order to double-check the results. Wooldridge (2016, p. 575) suggests an approach developed by Kwiatkowski, Phillips, Schmidt and Shin (KPSS) in 1992 as a possible alternative. The main difference between the ADF and KPSS test is the null hypothesis, in which KPSS states stationarity or I(0) as the null hypothesis. The alternative hypothesis is non-stationarity and consequently the variable must be integrated of a higher order.

Suppose we have a series with n observations $y_t, t = 1, 2, \dots, T$ that is the object of an investigation on stationarity. Kwiatkowski et al. (1992) suggests that this series can be decomposed into a deterministic trend, a random walk and a stationary error term:

$$y_t = \xi t + r_t + \varepsilon_t,$$

where r_t is a random walk

$$r_t = r_{t-1} + u_t,$$

where the innovation term, u_t is identically and independently distributed (i.i.d.) with mean = 0 and variance = σ_u^2, and the initial value of r_0 serves the role of an intercept term. This gives the stationarity hypothesis of $\sigma_u^2 = 0$. Since the error term ε_t is assumed to be stationary, y_t is stationary around a trend under the null hypothesis. Setting $\xi = 0$, yields a case where y_t is stationary around a level of r_0, instead of a trend. Kwiatkowski et al. (1992) suggest using a one-sided LM-statistic (Lagrange Multiplier) and further derives their model to be equivalent to an ARIMA model:

$$y_t = \xi + \beta y_{t-1} + w_t, \qquad w_t = v_t + \theta v_{t-1}, \qquad \beta = 1,$$

where they let $\lambda = \sigma_u^2 / \sigma_\varepsilon^2$, giving the connection between θ and λ:

$$\theta = -\frac{\left\{(\lambda + 2) - [\lambda(\lambda + 4)]^{\frac{1}{2}}\right\}}{2}, \qquad \lambda = -\frac{(1 + \theta)^2}{\theta}, \tag{3.5}$$

Where

$$\lambda \geq 0, \qquad |\theta| < 1.$$

Thus, $\lambda = 0$ implies $\theta = -1$ (stationarity), while $\lambda = \infty$ corresponds to $\theta = 0$, which implies that y is a pure random walk. Their approach is to effectively test $\theta = -1$ assuming $\beta = 0$, whereas the Dickey-Fuller approach tests $\beta = 1$ assuming $\theta = 0$ (Kwiatkowski et al., 1992). The main difference is therefore that the KPSS test has a null hypothesis of stationarity in the series, while Dickey-Fuller has a null hypothesis of non-stationarity. KPSS tests for the relevant variables are conducted in R, using the package *urca*. The conclusion of the KPSS tests are in line with the results from the Augmented Dickey-Fuller tests. However, if there are any discrepancies, we decide to assume non-stationarity and use the first-difference of the variable. This way we avoid the risk of including non-stationary series in the regressions.

3.1.2 Other assumptions

According to Wooldridge (2016, pp. 345-365), there are five assumptions that must hold in order to perform large-sample inference for time series regression. These are called the Gauss-

Markov assumptions and state that the stochastic process is stationary, weakly dependent and follows a linear model. Further, the explanatory variables have no perfect collinearity and have zero conditional mean, while the error terms are contemporaneously homoscedastic and uncorrelated over time. We assume that these Gauss-Markov assumptions are known to the reader and will therefore not elaborate them and their accompanied diagnostic tests, apart from the stationarity property of time series that was elaborated in section 3.1.1. It should be mentioned that we make use of Newey-West heteroscedasticity and autocorrelation (HAC) standard errors to try and overcome potential serial correlation and heteroscedasticity in the error terms (Newey & West, 1987; Wooldridge, 2016, p. 389).

3.1.3 Model specification

To examine the impact of ETF activity on the underlying U.S. corporate bond market on an aggregate level, we estimate the relationship between selected risk commonality measures and metrics of ETF activity through several time series models. The models differ in terms of lags included, both of the dependent variable and the explanatory variables.

Model 1

The first model is a static model, investigating the contemporaneous relationship between the variables. This is a simple model, where we focus on an instant impact on the underlying bond market from ETF activity. The *ETF%* variable was found to be a process of neither I(0) nor I(1) and therefore not included in the time series regressions. This is unfortunate, as the variable has a highly significant relationship with several commonality measures in the naïve OLS estimation. The model for the commonality measures (in the following referred to as CM) that are stationary in differences is stated below:

$$\Delta CM_t = \beta_1 \Delta Flow_t + \beta_2 \Delta ETFturnover_t + \beta_3 \Delta SDshares_t + \varepsilon_t \tag{3.6}$$

The included commonality measures are specified in sections 3.3 and 3.4 together with the ETF activity measures.

Model 2

In this model, we take into account the lag structure of both the dependent and independent variables. This yields different autoregressive distributed lag model, ARDL(p, q,...,q) with lags for dependent variables given by p and lags for the independent variables given by q. The lag structure for each independent variable is not necessarily the same but for notational simplicity, we specify the model using q as the maximum lag used for all explanatory variables. This model is specified below:

$$\begin{aligned}\Delta CM_t &= \sum_{i=1}^{p} \gamma_1^i \Delta CM_{t-i} + \sum_{j=0}^{q} \beta_1^j \Delta Flow_{t-j} + \sum_{k=0}^{q} \beta_2^k \Delta ETFturnover_{t-k} \\ &+ \sum_{l=0}^{q} \beta_3^l \Delta SDshares_{t-l} + \varepsilon_t\end{aligned} \tag{3.7}$$

3.2 Panel data

To examine the impact of ETFs on the underlying market further we construct two panel data sets based on ETF bond ownership information from iShares. The construction of the panel data set is elaborated under 4.1.3 in the data chapter, while the variables we create for the panel regressions are elaborated under 3.3.5 and 3.4.5. iShares is the sponsor of the largest investment grade (LQD) and high yield (HYG) bond ETFs (ETF Database, 2018). It may therefore be suitable to investigate how the correlation of the underlying bonds is influenced by the degree of ETF ownership and trading activity. Panel data is used in other papers investigating the effects of bond ETF in the underlying market such as in Sultan (2015) and Dannhauser (2017). To get a more detailed perspective of possible interactions between the ETF and corporate bond market, we exploit the possibilities provided by different fixed effects models. As a monthly correlation measure in the panel models, we use the 12-month rolling return correlation between individual bonds and the ETFs' benchmark indices.

3.2.1 Fixed effects

When performing pooled panel regressions, an important assumption is that time-invariant characteristics (α_i) of each individual (bond) are not correlated with the explanatory variables (Wooldridge, 2016, p. 413). If this is not the case, we introduce heterogeneity bias caused by omitting time-constant variables. However, when using panel data with fixed effects, we allow for unobserved individual effects α_i to be correlated with explanatory variables. For this reason, we add fixed effects in our models. When running panel regressions, it is possible to add individual (α_i) and time (α_t) fixed effects in the model. Individual fixed effect estimators capture characteristics about individuals that are constant over time. While time-specific effects capture unobserved variations in specific time periods. To illustrate how a fixed effects transformation work we consider a model from Wooldridge (2016, p. 435) with a single explanatory variable (β_1) for each individual (i):

$$y_{i,t} = \beta_1 x_{i,t} + \alpha_i + u_{i,} \qquad t = 1,2,\dots,T. \tag{3.8}$$

For each i we average this equation over time and wind up with:

$$\bar{y}_i = \beta_1 \bar{x}_i + \alpha_i + \bar{u}_i, \tag{3.9}$$

where $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$, and so on. Because α_i is fixed over time, it appears in both (3.13) and (3.14) for each t, we end up with:

$$y_{i,t} - \bar{y}_i = \beta_1 \left(x_{i,t} - \bar{x}_i\right) + u_{i,t} - \bar{u}_i \qquad t = 1,2,\dots,T,$$

which can be written as:

$$\ddot{y}_{i,t} = \beta_1 \ddot{x}_{i,t} + \ddot{u}_{i,t} \qquad t = 1,2,\dots,T, \tag{3.10}$$

where $\ddot{y}_{i,t} = (y_{i,t} - \bar{y}_i)$ is the time-demeaned data of y and similar interpretation is made for $\ddot{x}_{i,t}$ and $\ddot{u}_{i,t}$. This transformation is also called the within transformation and after performing it the individual effect a_i has been differenced away and we can perform pooled OLS without heterogeneity bias in the estimators.

There are some important considerations we have to make when it comes to the interpretation of the results from our unobserved effects model by fixed effects (Wooldridge, 2016, p. 437). The first is that we cannot include time-constant variables by themselves in our model. The second consideration is that when we include time specific effects we cannot estimate the effect of any variable whose change across time is constant. This is the case if we for instance include a variable that increases by the same amount between two periods for every individual in the sample. We will not be able to distinguish the increase from the aggregate time effect and therefore a possible element of bias in introduced. Therefore, we remove variables with these qualities in the models where time fixed effects are included. A third consideration and key assumption for inference is no serial correlation in the idiosyncratic errors (Wooldridge, 2016, p. 459). To obtain fully robust standard errors we apply clustering at the individual (bond) level of our time-demeaned data. Clustering makes inference robust to heteroscedasticity (Millo, 2017). Therefore, we use Arellano cluster-robust standard errors at the individual level in our models (Arellano, 1987). The last concern is strict exogeneity that will be commented in the next section. With the following considerations in mind the fixed effect estimator is roughly unbiased (Wooldridge, 2016, p. 435).

3.2.2 Endogeneity concerns

Endogeneity issues can arise in several ways in our data and we want to perform analysis that is robust to these issues (Sultan, 2015). The first way in which these concerns could arise is if ETFs pick bonds based on some characteristics that make the selected bonds correlated to the market index. As mentioned in section 2.4, bond ETFs replicate their index by using sampling techniques. Some potential bond factors ETF sponsors look at when replicating the market index that are time-variant, are duration and credit rating. To tackle this possible endogeneity issue, we include credit rating and duration as control variables in all the model specifications. The included credit rating is the numerical S&P rating, while the duration is stated in years. A second concern is that there could be other unobserved components that drive correlation that varies across bonds, while a third concern might be that bond correlations are affected by time-specific events. To alleviate the two last concerns, we include individual (bond) and time (month) fixed effects in our panel regressions. According to Dannhauser (2017), the inclusion of time fixed effects additionally controls for common trends in the corporate bond markets.

3.2.3 Model specification

To investigate the relationship between ETF activity and bond correlations we specify three different models where we include various explanatory variables and fixed effects. We run each of the models separately on the investment grade and high yield bond subsample. In this part, we explain the model specifications and variables. The regression results are discussed under 5.3.

Model 1

The dependent variable for all the panel regression specification is $Corr_{i,t}$, which is a monthly observation of the 12-month rolling correlation between individual bonds and their respective segment Markit index. In model 1, we include four different ETF variables. *ETFshare* is the share of the par value of an investment grade bond that is owned by the LQD or the share of a high yield bond that is owned by HYG for a given month. The three next measures we include are *ETFturnover*, *SDshares* and *Flows* that we describe under section 3.4**.** We run the ETF measures for LQD on the investment grade bond sample and HYG on the high yield sample. Next, we run four different regressions: Pooled OLS without fixed effects (3.11), panel regression with individual (bond) fixed effect (3.12), panel regression with time fixed effect (3.13) and a two-way fixed effect regression that adjusts for both time and individual fixed effects (3.14). We further include the control variables mentioned in section 3.2.1. Adding the individual bond credit rating and duration should alleviate some of the raised endogeneity concerns and strengthen the interpretation of our results.

We show the formula of the four different specifications of model 1 below. *ETFturnover*, *SDshares* and *Flows* are time series that will change by the same amount between two periods for all individual bonds when added. As mentioned under section 3.2.1, including them in the regressions with a time fixed effect yield biased estimates. Hence, we remove these explanatory variables in regression (3.13) and (3.14) where a time fixed effect is added.

$$Corr_{i,t} = \beta_1 ETFshare_{i,t} + \beta_2 ETFturnover_t + \beta_3 SDshares_t + \beta_4 Flows_t + \beta_5 Rating_{i,t} + \beta_6 Duration_{i,t} + \varepsilon_{i,t} \tag{3.11}$$

$$Corr_{i,t} = \alpha_i + \beta_1 ETFshare_{i,t} + \beta_2 ETFturnover_t + \beta_3 SDshares_t + \beta_4 Flows_t + \beta_5 Rating_{i,t} + \beta_6 Duration_{i,t} + \varepsilon_{i,t} \tag{3.12}$$

$$Corr_{i,t} = \alpha_t + \beta_1 ETFshare_{i,t} + \beta_2 Rating_{i,t} + \beta_3 Duration_{i,t} + \varepsilon_{i,t} \tag{3.13}$$

$$Corr_{i,t} = \alpha_i + \alpha_t + \beta_1 ETFshare_{i,t} + \beta_2 Rating_{i,t} + \beta_3 Duration_{i,t} + \varepsilon_{i,t} \tag{3.14}$$

Model 2

For model 2, we create three new variables to take a closer look at the effect of ETF ownership. We multiply the *ETFshare* variable with *ETFturnover*, *SDshares* and *Flows* in order to create the new variables *Turnover%*, *SDshares%* and *Flows%*. By doing this, we are able to possibly get a more detailed look at how ETF activity influence the correlation of bonds. We also include the control variables from model 1, credit rating and duration. In order to avoid possible collinearity issues due to the new variables being derived from *ETFshare*, we run separate regressions for each. In addition, we specify two different regressions for each variable including individual and two-way fixed effects. The three variations of the last two-way specification regressions are illustrated below.

$$Corr_{i,t} = \alpha_i + \alpha_t + \beta_1 SDshares\%_{i,t} + \beta_2 Rating_{i,t} + \beta_3 Duration_{i,t} + \varepsilon_{i,t} \tag{3.15}$$

$$Corr_{i,t} = \alpha_i + \alpha_t + \beta_1 ETFturnover\%_{i,t} + \beta_2 Rating_{i,t} + \beta_3 Duration_{i,t} + \varepsilon_{i,t} \tag{3.16}$$

$$Corr_{i,t} = \alpha_i + \alpha_t + + \beta_1 Flows\%_{i,t} + \beta_2 Rating_{i,t} + \beta_3 Duration_{i,t} + \varepsilon_{i,t} \tag{3.17}$$

3.3 Risk commonality measures

To explore if fixed income ETFs have influenced the underlying bond market, we look at several correlation, commonality and liquidity measures. The increase in the trading volume of bond ETFs during the last decade is substantial and these funds now account for 8% of the trading volume in the investment grade segment (PPM America, 2018). Hence, investigating measures of trading commonality is of interest.

3.3.1 Commonality measures

To calculate the comovement of fixed income securities, we use two types of measures: Dispersion and correlation. Both measures are commonly used when investigating the relationship of the performance of individual securities or even entire asset classes in relation to others (Janus Henderson Investors, 2017). In this part, the difference between the two types of measures is briefly explained.

Dispersion can be seen as the difference between the best and worst performers in an index or group of securities (Janus Henderson Investors, 2017). For stock pickers, high return dispersion is positive since this indicates that there are opportunities to generate excess returns. Dispersion is usually measured as a standard deviation. If the average deviation between security returns and a market index is high, this could be an indication of high dispersion. In our analysis we will look at the dispersion of monthly returns and also other measures such as volume and yield. Unlike dispersion, correlation does not measure the difference in level of performance between two variables, but rather the directional relationship between them (Janus Henderson Investors, 2017). Figure 3.1 below illustrates the difference between correlation and dispersion. The two assets have experienced high correlation, but their return dispersion is also high.

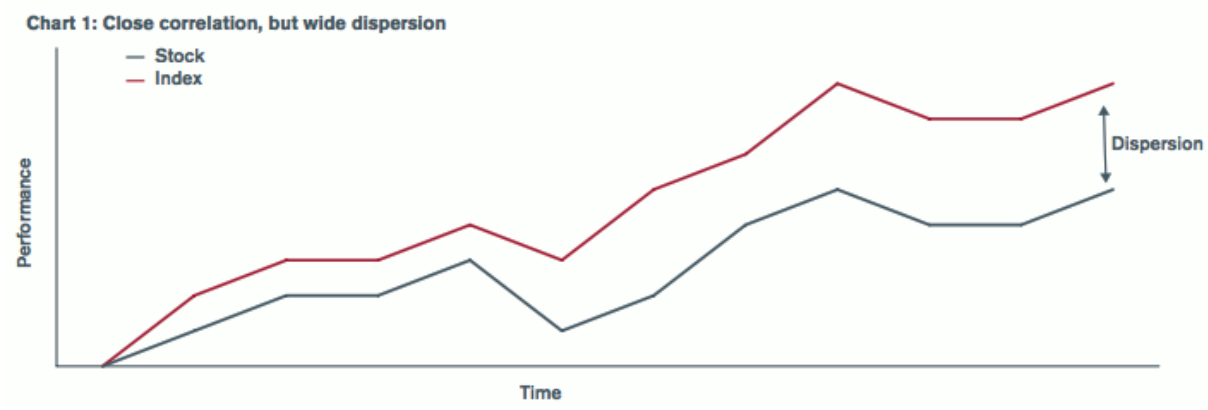


Figure 3.1: Correlation vs dispersion, source: Janus Henderson investors (2017)[6]

3.3.2 Commonalities in trading pattern

Sullivan & Xiong (2012) argues that to understand the market impact of index trading, it is more helpful to look at the dispersion in trading volume-changes rather than dispersion of its

[6] Note: Reprinted from *A simple guide to dispersion and correlation,* by Janus Henderson investors (2017), retrieved from: http://az768132.vo.msecnd.net/documents/103808_2017_05_24_10_51_29_557.gzip.pdf

absolute level. An implication of increased index trading is that a larger share of trades occurs in basket transactions, as ETFs and passive mutual funds buy (sell) securities in groups in response to capital inflow (outflow) (Sullivan & Xiong, 2012). Such basket orders can be spread out over several transactions to minimise price impact. This transaction pattern creates trading volume-changes in the underlying securities that are more similar. Hence, we want to look at measures that capture the effect of changing trading patterns. In constructing the volume-change dispersion, we start by measuring the logarithmic change in the dollar trading volume between two periods as:

$$\Delta V_{i,t} = \ln\left[\frac{V_{i,t}}{V_{i,t-1}}\right] \tag{3.18}$$

where the subscript *i* denotes the bonds in the market and *t* refers to the time period.
Following Bolla et al. (2016), the cross-sectional dispersion of the change in trading volume can be written as:

$$VolDisp_t = \sqrt{\frac{1}{I-1}\sum_{i=1}^{I}(\Delta V_{i,t} - \overline{\Delta V_t})^2} \tag{3.19}$$

which also can be described as the cross-sectional standard deviation of volume changes at time period *t*. *I* refers to the total number of bonds analysed (Bolla et al., 2016). If commonalities in trading volume-change among bonds have increased, then the volume-change dispersion would be lower, and the opposite for less commonality in volume-changes.

Another commonality measure related to the trading patterns of bonds is the average pairwise correlation of volume change in trading of bonds. Bolla et al. (2016) constructs this measure as follows:

$$Vcorr_t = \left[\sum_{i=1}^{I}\sum_{j>i}^{J}\frac{(T-1)^{-1}\sum_{t=1}^{T}(\Delta V_{i,t} - \overline{\Delta V_i})(\Delta V_{j,t} - \overline{\Delta V_j})}{\sigma_{\Delta V_i}\sigma_{\Delta V_j}}\right] / N \tag{3.20}$$

where

$$N = \frac{I(I-1)}{2}, \qquad \overline{\Delta V_i} = \frac{1}{T}\sum_{t=1}^{T}\Delta V_{i,t}$$

T refers to the total number of time steps and $\sigma_{\Delta V_i}$ is the standard deviation of the logarithmic volume change in trading of bond *i*. The rational behind the avereage pairwise correlation measure is that if commonality in bond trading is higher, then the average correlation between pairs of bond volume-changes is higher. We construct *Vcorr* with a rolling window of 12 months.

Both of these commonality measures for trading, focus on the volume side of bond trading and development in the trading pattern. With the surge of passive and index-linked investing during last decades, we expect the trading related to these investment vehicles to have some impact on the trading pattern.

3.3.3 Commonalities in returns

While we previously focused on the trading pattern of passive investment funds, we now put our focus to commonalities in returns and yields of the underlying bonds the passive funds invest in. In commonalities of returns we want to look at whether there is some indication of bond returns being less (or more) dispersed in the period after bond ETFs and other passive investment vehicles have grown to be significant players in the market. Da & Shive (2018) find evidence of a link between measures of ETF activity and return comovement among stocks. Construction of similar measures in returns for bonds will help in investigating such relationships between bonds and passive investing.

First, we define return for bonds. Bond returns encompass coupon payment, change in the price of bonds and accrued interest (Bai, Bali, & Wen, 2016).

$$R_{i,t} = \frac{P_{i,t} + AI_{i,t} + C_{i,t}}{P_{i,t-1} + AI_{i,t-1}} - 1$$

We construct a similar dispersion measure for returns as for volume change in bond trading. The return dispersion may also be thought of as the cross-sectional variation in performance across the respective bonds.

$$RetDisp_t = \sqrt{\frac{1}{I-1}\sum_{i=1}^{I} (R_{i,t} - \overline{R_t})^2} \tag{3.21}$$

where

$$\bar{R}_t = \frac{1}{I}\sum_{i=1}^{I} R_{i,t}$$

The dispersion in bond returns tells something about the disparity in the level of returns in the market. If return dispersion is low, then there is high similarity in returns among the bonds in a chosen sample. Vice versa, if dispersion is high, there is less similarity in the return of bonds. An appropriate sample could for instance be a market, an index or a rating class.

Average pairwise correlation of returns is another commonality measure that Bolla et al. (2016) make use of. This measure estimates average of all pairwise correlation at a particular point in time, where pairwise correlations are equally weighted. It is calculated in a similar way as for volume change (*Vcorr*) through the following formula

$$Rcorr_t = \left[\sum_{i=1}^{I}\sum_{j>i}^{J} \frac{(T-1)^{-1}\sum_{t=1}^{T}(R_{i,t}-\bar{R}_i)(R_{j,t}-\bar{R}_j)}{\sigma_{R_i}\sigma_{R_j}}\right] / \, N \qquad (3.22)$$

where

$$N = \frac{I(I-1)}{2}, \qquad \bar{R}_i = \frac{1}{T}\sum_{t=1}^{T} R_{i,t}$$

Higher average pairwise correlation among bonds would indicate that the bonds in sample behave more similar. A consequence of higher correlation among bonds could be an environment where it is more difficult to diversify a portfolio consisting of bonds (Markowitz, 1952). This correlation measure is constructed using a rolling windows of 12 months on the observations.

A third commonality measure in returns we look at is similarities in the change of yield of bonds. The yield to maturity of a bond is the expected return on a bond if it is held until maturity (Tuckman & Serrat, 2012, p. 100). To construct this measure, we use a similar approach as Bolla et al. (2016) in calculating return dispersion, but replace return with yield. We choose to look at both the logarithmic change in yield and the simple difference between two consecutive yields for the bonds in sample. The logarithmic change in yield of bond *i* at time *t* can be defined as

$$\Delta_1 yield_{i,t} = \ln\left[\frac{yield_{i,t}}{yield_{i,t-1}}\right]$$

The simple difference in yield of bond *i* at time *t* can be defined as

$$\Delta_2 yield_{i,t} = yield_{i,t} - yield_{i,t-1}$$

The measure of change in yields can then in turn be used to define yield-change dispersion

$$YDisp_t = \sqrt{\frac{1}{I-1}\sum_{i=1}^{I} (\Delta yield_{i,t} - \overline{\Delta yield_t})^2} \tag{3.23}$$

where

$$\overline{\Delta yield_t} = \frac{1}{I}\sum_{i=1}^{I} \Delta yield_{i,t}$$

If the yield-change dispersion measure is declining over time, it indicates that yields among the bonds in sample have become more similar throughout the time window. Therefore, yield-change dispersion can be a measure of more comovement among the debt securities. This could in turn make it more difficult for a bond investor to diversify risk. If the yield-change dispersion over time is rising, then it is a signal of more variation in yields between the bonds. These are market conditions that possibly could make it easier for an investor to diversify risk.

Similar to *Rcorr,* the average pairwise correlation of yield-change can be constructed in the following:

$$Ycorr_t = \left[\sum_{i=1}^{I}\sum_{j>i}^{J} \frac{(T-1)^{-1}\sum_{t=1}^{T}(\Delta yield_{i,t} - \overline{\Delta yield_i})(\Delta yield_{j,t} - \overline{\Delta yield_j})}{\sigma_{\Delta yield_i}\sigma_{\Delta yield_j}}\right] / N \tag{3.24}$$

where

$$N = \frac{I(I-1)}{2}, \qquad \overline{\Delta yield_i} = \frac{1}{T}\sum_{t=1}^{T} \Delta yield_{i,t}$$

As mentioned, the yield correlation measure is similar to return correlation. If the average correlation of yield-change in bonds have risen over time, this may be an indicator of more comovement and similarities in the bond market. The correlation measure is constructed over a rolling window of 12 months.

3.3.4 Commonalites in liquidity

Compared to the equity markets, the liquidity of bond markets is usually lower (Kiernan, 2015). Additionally, investors are often large and institutional, and often a significant part of the bonds are held to maturity and therefore not trading. Some bonds might not be traded on a single day, week or even within a month. It seems thus that liquidity in bond market has some special features. To investigate effects on liquidity as a result of growth in passive investing, we need to establish some appropriate measures of liquidity.

As liquidity is not a specific size that can be observed directly, it needs to be estimated (Kamara, Lou, & Sadka, 2008). Amihud (2002) propose a liquidity measure for stocks focused on the relationship between the return of a security on a single day and the dollar volume of trading in that security on that particular day ($Vol_{i,t}$). The Amihud illiquidity measures the price impact of a trade per unit traded (Dick-Nielsen, Feldhütter, & Lando, 2012) and can be defined as

$$ILLIQ_{i,t} = \frac{\left|\frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}}\right|}{Vol_{i,t}}$$

If the absolute return of security on a particular day is large relative to the dollar trading volume, it implies that the volume traded has a high influence on the price movement in that security. This may be an indication of low liquidity. If the return of a security is low relative to the dollar trading volume, then a low volume transaction would not affect the price much and hence the liquidity is relatively high. The Amihud illiquidity measure can thus be thought of as the price impact of trades in a security (Lin, Wang, & Wu, 2011). Due to non-stationary properties of the time series of the Amihud measure (Kamara et al., 2008), we define the change in illiquidity as

$$\Delta ILLIQ_{i,t} = \ln\left[\frac{ILLIQ_{i,t}}{ILLIQ_{i,t-1}}\right]$$

If the liquidity in security *i* improves from time *t-1* to time *t* variable $\Delta ILLIQ_{i,t}$ takes a negative value. $\Delta ILLIQ_{i,t}$ takes a positive value if liquidity measured by Amihud declines between the time periods.

Commonalities in liquidity among bonds can be mesaured by the average pairwise correlation of liquidity (Bolla et al., 2016). This correlation measure is constructed in the following way

$$Lcorr_t = \left[\sum_{i=1}^{I}\sum_{j>i}^{J}\frac{(T-1)^{-1}\sum_{t=1}^{T}(\Delta ILLIQ_{i,t}-\overline{\Delta ILLIQ_i})(\Delta ILLIQ_{j,t}-\overline{\Delta ILLIQ_j})}{\sigma_{\Delta ILLIQ_i}\sigma_{\Delta ILLIQ_j}}\right] / N \quad (3.25)$$

where

$$N = \frac{I(I-1)}{2},\ \ \overline{\Delta ILLIQ_i} = \frac{1}{T}\sum_{t=1}^{T}\Delta ILLIQ_{i,t}$$

and $\sigma_{\Delta ILLIQ_i}$ refers to the standard deviation of change in the Amihud illiquidity of security *i*. If the liquidity among bonds moves in the same direction and with similar magnitude, then *Lcorr* will be high, and low in the opposite case. We are thus interested in the development in liquidity correlation over time to investigate whether passive-investing has increased risk commonalities among bonds. We use a window of 12 months to construct a rolling correlation measure of *Lcorr*.

3.3.5 Bond-index correlation

For the panel data analysis, we create a correlation measure that is unique for each individual bond. We calculate the 12-month rolling window correlation between individual bond returns and the returns of a corporate bond market index. We use the Markit iBoxx USD Liquid high yield index and investment grade index. These are the indices that the iShares high yield and investment grade ETFs track. For high yield bonds, the correlation with the high yield index is calculated, while investment grade bonds correlations are calculated with the investment grade index. To create the measure, we need at least 12 bond-month observations. The formula specification of the rolling window measure is similar to the specification from Duda & Augustynek (2005).

$$Corr_{i,t} = \sum_{i=t-n+1}^{t}\frac{(r_{Bond,i}-\bar{r}_{Bond,i})(r_{index,i}-\bar{r}_{index,i})}{\sigma_{r_{Bond,i}}\sigma_{r_{Index,i}}} \quad (3.26)$$

where

$$\bar{r}_{Bond_i} = \frac{1}{n}\sum_{i=t-n+1}^{t} r_{Bond_i}, \qquad \bar{r}_{Index} = \frac{1}{n}\sum_{i=t-n+1}^{t} r_{Index}$$

and n is the length of the rolling window.

3.4 ETF activity variables

In the following section, we explain how our explanatory ETF activity variables are created. We are interested in investigating the relationship between the following measures and the commonality measures that we identified in section 3.3. If relationships are identified it is possible that ETFs could propagate comovement in the underlying corporate bond market.

3.4.1 ETF assets under mangement

In order to create measures of bond ETFs assets under management (AUM) as a percentage of the total amount of corporate debt outstanding we need to find a measure that captures the development of the size of the market. As a proxy for the market size we use a data series from the Financial Accounts of the United States published by the Federal Reserve. To our knowledge, there is no common proxy used by practitioners to measure the size of the corporate bond markets. We decide to use "Nonfinancial corporate business corporate bonds liability level" as this series was readily accessible and seemed to capture the size of the corporate bond market well when we compare with other data sources such as SIFMA (2018). The series tracks the amount outstanding of US nonfinancial business corporate bonds at a quarterly frequency.

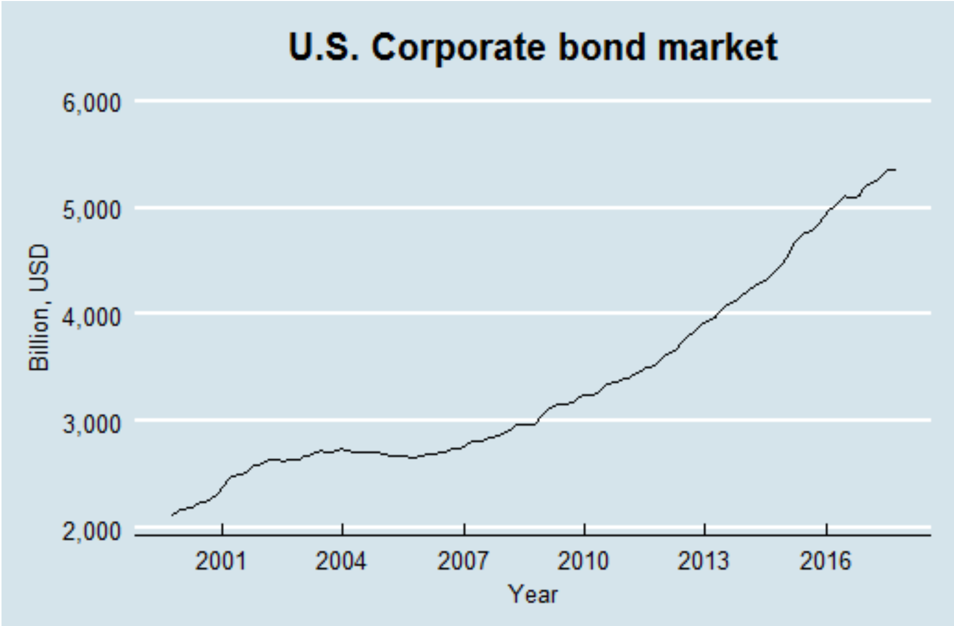


Figure 3.2: U.S. corporate bond market, source: (Board of Governors of the Federal Reserve System (US), 2018)

To compare monthly ETF AUM with the size of the market we need an estimation of monthly amount outstanding. In order to convert the series from quarterly to monthly data we employ cubic spline interpolation, which is a technique used to estimate higher frequency data points from low frequency data (Torres-Reyna, 2014). The method we apply is explained in detail in Torres-Reyna (2014). Figure 3.2 shows the interpolated time series of monthly corporate bond liability levels. When performing regressions on interpolated data, it is important to consider that a systematic source of serial correlation in the regressor is introduced (Dezhbakhsh, 1994). The interpolated data points are related to each other in a systematic way by a cubic polynomial, which will lead to a violation of the OLS assumption of no autocorrelation. To control for this issue, we use Newey-West standard errors when testing for significance of regression coefficients as they are robust to autocorrelations (Wooldridge, 2016, p. 389). Newey-West standard errors additionally correct for heteroscedasticity in the error terms.

Similar to Bolla et al. (2016) we define the variable for ETF holdings as a share of the total bond market as:

$$ETF\%_t = \frac{ETF\ AUM_t}{Market\ size_t} \tag{3.27}$$

where *ETF AUM* is the monthly aggregate assets under management for the exchange traded funds in our sample. We collect *ETF AUM* from Bloomberg. Market size is as discussed in this section, the interpolated quarterly data series of corporate bond liability level published by FRED. We calculate this measure both for aggregate bond ETFs and for corporate bond ETFs. Separating these two types of bond ETFs enables us to explore both the impact from more general, aggregate fixed income ETFs, while using corporate bond ETFs may be more precise when looking at specific impacts on the underlying market of corporate bonds.

3.4.2 ETF turnover

A relevant measure of ETF activity is the turnover of shares in a ETF. According to Da & Shive (2018), this is defined as:

$$ETFturnover_t = Monthly\ mean\left(\frac{1}{I}\sum\nolimits_{i=1}^{I}\frac{Daily\ number\ of\ shares\ traded_{i,d}}{Shares\ outstanding\ that\ day_{i,d}}\right) \tag{3.28}$$

where d is the daily time index for ETF i, while I is the total number of ETFs in sample. We calculate the daily mean of turnover for all the ETFs in our sample. We use this to calculate a monthly mean of turnover, which is the regressor in our estimation models. The rationale of using ETF turnover as a measure of ETF activity, is that it is likely to be positively correlated with arbitrage activity (Da & Shive, 2018). As the ETF seeks to track a given benchmark index, market participants and traders will correspondingly try to make arbitrage trades on deviations between he ETF's net asset value and the market value of the underlying securities. If there is a high turnover rate in the ETF shares, it may be an indication of high arbitrage activity among market participants and can in turn impact the underlying bond market. The APs are also able to perform arbitrage activity by selling (or buying) ETF shares and taking the opposite position in the underlying security if there is a mismatch between the price of ETF shares and the components' prices. However, it is not expected that all of the trading in ETF shares is due to arbitrage activity. A number of investors are likely to merely look for exposure to the underlying market as a part of their investment strategies (Da & Shive, 2018).

3.4.3 ETF creation/redemption activity

As discussed in section 2.2.3, creation/redemption activity is a unique property of exchange traded funds, as APs have the right and opportunity to write (create) new shares and withdraw (redeem) existing shares. Similar to Da & Shive (2018), we create a proxy measure of creation/redemption activity as

$$SDshares_t = \frac{1}{I}\sum_{i=1}^{I} \frac{Stdev(daily\ number\ of\ shares\ outstanding_i)}{Mean(shares\ outstanding\ during\ the\ month_i)} \quad (3.29)$$

where *i* refers to the individual ETFs and *I* is the total number of ETFs in sample. The variable *SDshares* has a monthly time index *t.* This variable is supposed to measure the intensity of changes in the shares outstanding in the ETFs, and as the APs have the opportunity to create and redeem shares, a change in number of shares outstanding indicates such activity. In turn, this activity will be associated with volatility of demand for the underlying bonds of the ETF. A high value of *SDshares* can indicate that the primary activity between APs and the ETF provider is large, which in turn may affect the underlying security markets. There is a possibility that creation and redemption could drive correlation in the bond market, if the APs buy and sell securities in blocks that are relatively large (Da & Shive, 2018). However, it may be the case that APs trade in smaller blocks and over several days in order to minimise market impact and thus the influence from creation and redemption activity is reduced. In addition, as mentioned in section 2.2.3 the APs have an arbitrage motive that may generate more creation and redemption activity motivated by deviations in the pricing of shares versus the underlying securities. Such arbitrage mechanisms can in turn generate more primary market activity which will result in a higher standard deviation of shares outstanding.

3.4.4 ETF flows

A fourth measure of ETF activity, is the net inflow of capital into ETFs. As ETFs attract more capital from both individual and institutional investors, these flows may induce trading in the bond market unless the APs already have these underlying securities on book. If they have the shares on book, they can just trade the shares for the cash inflow from the investors without making trades in the underlying securities. This measure may capture such behaviour better than for instance our proxy for creation/redemption activity that depends on the issuance or withdrawal of ETF shares. We define the ETF flows as the following:

$$Flows_t = \frac{1}{I}\sum_{i=1}^{I} \frac{Net\ inflow_{i,t}}{AUM_{i,t-1}} \quad (3.30)$$

where *i* refers to the individual ETF and *I* as the number of ETFs in sample. This variable is a monthly time series, with a time index *t*. In order to make this measure a ratio, we divide the net inflow of capital by the AUM of the ETF the previous month. This makes it easier to consider the relative size of the inflows or outflows the fund experiences for a given month. We average this ratio over ETFs in our sample at time *t*.

3.4.5 ETF share

For the panel regressions, we create a new ETF measure, *ETFshare*. *ETFshare* represents how much of the amount outstanding of a single bond that is owned by an ETF in a given month. Da & Shive (2018) employ *ETFshare* as an explanatory variable for equity commonalities. The reason for including this measure is that when a larger share of a bond is held by ETFs the bond could move more in line with the market. In order to create this measure, we use two samples: One for high yield bonds and one for investment grade bonds. We only include bonds that are owned by either the iShares iBoxx $ Investment Grade Corporate Bond ETF (LQD) or iShares iBoxx $ High Yield Corporate Bond ETF (HYG) in the period from July 2002 to June 2016. Each monthly observation of a bonds amount outstanding is then divided by the monthly dollar amount owned by the ETFs.

$$ETFshare_{i,t} = \frac{ETF\ invested\ dollar\ amount_{i,t}}{Bond\ amount\ outstanding_{i,t}} \tag{3.31}$$

In model 2 of the panel regressions in section 3.2.3, we create the new variables *Turnover%, SDshares%* and *Flows%* by multiplying *ETFshare* with *ETFturnover, SDshares* and *Flows* from the LQD and HYG funds.

4. Data and descriptive statistics

In the following chapter, we provide an overview of the data used in the analysis. Section 4.1 presents the data collection process of bond, fund, panel and fundamental data and the different sources we use to obtain these data. In section 4.2 we give a short guide to the data cleaning procedure. The last section provides descriptive statistics for our aggregate bond sample, commonality measures and the ETF activity variables. Plots of the ETF activity and bond commonality measures are found in Appendix A. In addition, an overview of our empirical methods and their corresponding bond subsamples is presented in figure 4.2 under section 4.3.1.

4.1 Data collection

To perform empirical tests and investigate the relationship between ETFs and U.S. corporate bonds, we collect data from a multitude of sources. We divide our data sources into four separate groups: Bond data, fund data, panel data and fundamental data. For all the data sets, we collect information from July 2002 to June 2016 if possible since this is the period with complete observations from the Wharton Research Data Services' (WRDS) bond returns database.

4.1.1 Bond data

We collect bond data from the WRDS bond return database, a dataset that combines bond transaction data from FINRA Trade Reporting and Compliance Engine (TRACE) and data on bond issue characteristics (e.g. rating) from Mergent FISD. TRACE is a dataset created and compiled by the Financial Industry Regulatory authority (FINRA). Since July 2002 all broker dealers that are members of FINRA have an obligation to report fixed income transactions. In the WRDS bond database the transactional data are cleaned and aggregated to a monthly level. The reason for this is that unlike the equity market, several bonds trade only once or a few times per month. Hence, looking at returns at higher frequencies can be problematic. WRDS performs a cleaning process that we describe in Appendix B. We collect a sample of 1,289,222 bond-month observations between July 2002 and June 2016, with 40 different variables such as the end of month return of the issue, dollar trading volume, yield and rating. Summary statistics of the variables we use in our analysis is provided in table 4.1.

4.1.2 Fund data

We gather fund data from three different sources: Bloomberg, Center for Research in Security Prices (CRSP) and iShares (Blackrock's ETF provider branch). From Bloomberg, we collect data on net fund flows, in addition to data the ETFs' assets under management (AUM) at the end of each month.

For the exploratory analysis of the data with Pearson correlation and naïve OLS, we decide to make use of all bond ETFs with a mandate to invest in American corporate bonds. These ETFs include both aggregate and corporate bond ETFs. As we want to include only funds that exclusively trade in the U.S. market, we manually remove 20 ETFs with international mandates based on their fund name. This selection gives 62 aggregate and 107 corporate bond ETFs, thus a total of 169 funds. To construct the measures of ETF activity explained in chapter 3.4, we collect daily data of the bond ETFs from CRSP. In the stock/security file from CRSP, we are merely interested in daily number of shares outstanding and the trading volume. To extract information for each ETF in the sample, we filter the dataset by the ETFs' CUSIP code, stock ticker and setting security share code to 73, which is a security identifier for exchange traded funds. CUSIP is a unique number that is used to identify American financial instruments (SEC, 2018a). Out of the 169 ETFs in the data from Bloomberg, we find information about 151 in the CRSP stock/security files. In our dataset we end up with some duplicated daily observations because two different dividend amounts were recorded for the same trading day. Since we do not use the dividend amount in our analysis we only keep one of the trading day observations.

In the time series regression, we decide to only make use of the corporate bond ETFs that have clear mandate to invest exclusively in either investment grade (IG) or high yield (HY) bonds. The reason for this is that we believe that the level of precision is higher when separating the two rating classes, thus we can connect bond ETF activity in the two categories to their respective underlying investment segments. We find the funds' mandate by using the database of ETF.com, where we search for each ETF's ticker. In the Bloomberg data, we find 64 ETFs that are exclusively investing in investment grade U.S. corporate bonds and 36 ETFs in high yield bonds. After extracting the CRSP data, we are left with 40 IG and 20 HY bond ETFs. The respective losses in number of ETFs, occur due to the fact that some ETFs are incepted after the data period in our TRACE data (June 2016).

To calculate a measure of ETF AUM as a percentage of the total market value, we need a proxy for the size of the corporate bond market. As described in section 3.4.1, we use the data series “Nonfinancial corporate business corporate bond liability level” as our proxy. We retrieve this data from the FRED database of the Federal Reserve Bank of St. Louis.

4.1.3 Panel data

In our panel data estimation, we make use of monthly holding information for two of the largest corporate bond ETFs, namely iShares iBoxx $ Investment Grade (LQD) and iShares iBoxx $ High Yield (HYG). To create the investment grade and high yield panel data sets, we use data from TRACE, iShares, CRSP and Bloomberg. We source the panel data bond subset from TRACE by only including corporate bonds that are owned by either iShares LQD or HYG in the period from July 2002 to June 2016. We obtain ETF ownership from files providing monthly holdings disclosure from the iShares web page. Data for February 2014 is missing, but we include ownership data for all other months. We further use the constituents' unique CUSIP code to identify the bonds in the TRACE data set and create two subsets, one for the investment grade bonds and one for the high yield bonds. Next, we create monthly measures of bond-index correlation, ETF ownership and ETF activity. To create the ETF activity measures, we extract information for LQD and HYG from the CRSP database. Summary statistics for the measures that are included in the panel data set are described in Appendix A. In order to create a measure of monthly bond-index correlation, we download monthly return series for the Markit iBoxx USD Liquid Investment Grade Index and the Markit iBoxx USD Liquid High Yield Index, as these are the indices that the LQD and HYG track. For each bond in the sample, we calculate the 12-month rolling correlation between the bond and the market index that we use as the correlation measure. Since we need 12 months of observations to calculate the rolling window correlation, we drop all bonds with too few observations from the sample. In total, we end up with 2170 investment grade and 1293 high yield bonds in the subsamples.

4.1.4 Fundamental data

To find time series for the fundamental factors described under 2.3.1. for the robustness tests of our results, we use a variety of sources. For investor sentiment we download a monthly time series of adjusted closing prices for the CBOE VIX index from Yahoo finance. As our measure of credit risk in the corporate bond markets, we use two credit default swap indices, the CDX

HY index and the CDX IG Index from Markit. We download monthly changes in the price of the two indices from the Bloomberg terminal. For inflation and interest rate risk, we collect monthly data of the “Consumer price index for all urban consumers: all items” and the “5-year constant maturity rate” from the FRED database, respectively.

4.2 Data cleaning process

4.2.1 Cleaning and sampling of bond data

As explained Appendix B, the WRDS bond dataset we use in our analysis is a result of several cleaning procedures of TRACE raw data. We perform some further sampling procedures that we explain in the following.

As Badoer & Demiroglu (2017) suggest, we drop all convertible bonds in our sample. This is due to convertible bonds having an embedded option component that may be included in the price and this may in turn complicate the comparison of returns and comovement. The issue sizes of the bonds are not always correctly reported, and some bonds have a reported issue size of \$0. With that in mind we remove all bonds with an issue size of less than \$100 million, as smaller bonds are likely to be less liquid and rarely traded (Fitch Ratings, 2015). In addition, bonds with an issue size under a \$100 million often fall outside of the index methodologies of the indices that ETFs track. For instance, Bloomberg Barclays fixed income indices raised the minimum amount outstanding of included bonds to \$300 million in 2017 (Stone, 2017). We also decide to exclude non-rated bonds in our final sample, since we conduct the main analysis on both the investment grade and high yield segment. This gives us a sample of a total of 978 059 bond-month observations. Figure 4.1 shows the monthly number of bond return observations in our TRACE dataset. When creating the subsample for the panel data regressions (4.1.3), we apply the same cleaning and sampling procedures prior to the panel creation.

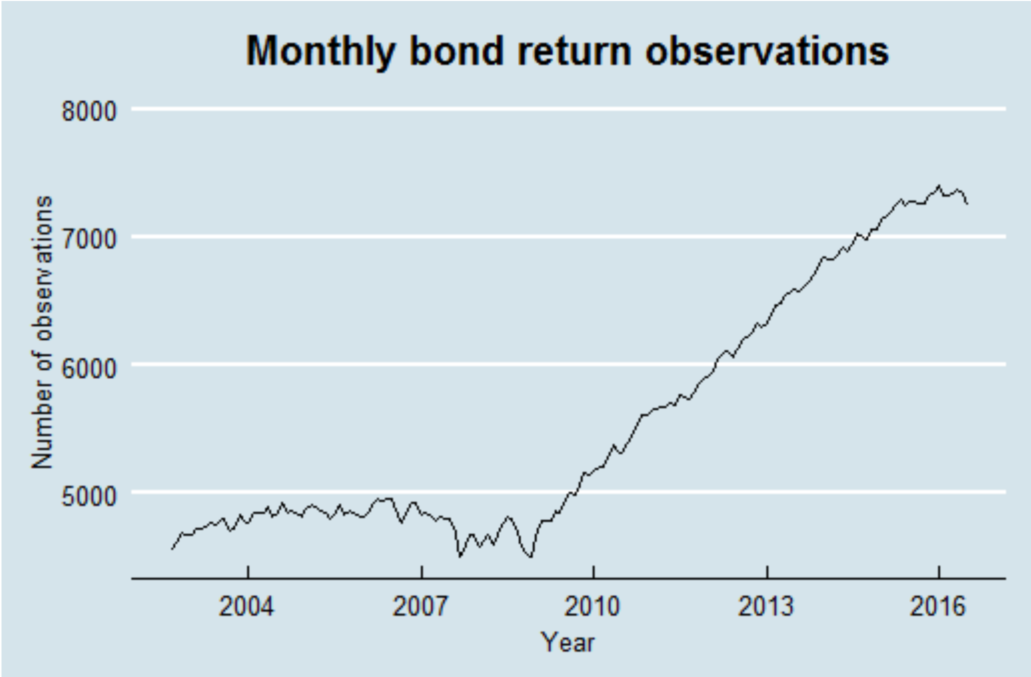


Figure 4.1: Number of bond return observations in TRACE dataset after cleaning

4.3 Descriptive statistics

4.3.1 Bond sample

In figure 4.2, we provide an overview of the empirical methods and corresponding samples for the analysis. The samples we use in the time series regressions, is the full sample split into a high yield and investment grade sample. Due to migration between rating classes, the total number of bonds in the IG and HY samples add up to more than 18 099 bonds. The panel regression subsamples only include bonds that are owned by either the LQD or the HYG ETF during the period between July 2002 and June 2016.

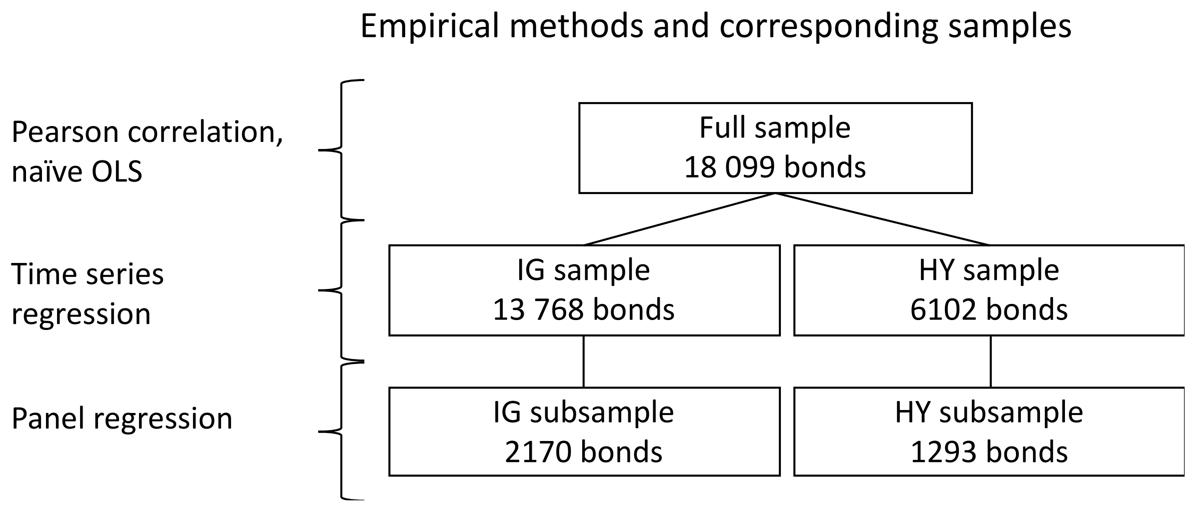


Figure 4.2: Empirical methods and corresponding data samples

The table below describes summary statistics for the most important variables from our full sample of bond-month observations from July 2002 to June 2016. Summary statistics for the time series samples and panel data subsamples is provided under Appendix A. The average bond issue in our sample has an offering amount of $565 million and yield of 5.6%. The largest bond issue of $15 billion is a Verizon bond with 30-year maturity issued in 2013, paying a coupon of 6.55 and with a rating of BBB+. The rating number variable indicates that the numeric average rating of the bonds in the sample is 8.9, which translates to BBB in the S&P rating system[7] . The average trading spread in the sample was 0.7% (70 basis points) while the average monthly return for bonds in the sample was 0.6%.

Table 4.1 Descriptive statistics of bonds

Variable	N	Mean	St. Dev.	Min	Max
Offering amount	978,059	565,727.4	551,369.8	101,000.0	15,000,000
Offering price	744,790	99.4	3.2	1.0	109.7
Principal amount	978,059	1,116.0	3,752.7	0.0	250,000
Coupon	978,059	6.2	2.0	0.0	15.5
Rating number	977,075	8.9	3.7	1	22
Trading Volume (MUSD)	978,059	52,526	201,939	2	43,763,849
Trading Spread	850,664	0.007	0.013	0.0	2.0
Yield	918,577	0.056	0.059	-1.0	1.0

[7] In the numerical code conversion of the S&P bond rating system, the highest quality bonds (AAA) have a numerical code of 1 while lowest rated bonds (D) receive a rating of 22. Hence, a lower number indicates higher credit quality (Jewell & Livingston, 1999).

Price End of month	978,059	104.3	14.2	0.0	3,348.0
Return End of month	929,621	0.006	0.045	-1.0	1.0
Duration	916,855	5.6	4.0	0.0	30.0

The average rating indicates that most of the bonds in the sample are investment grade. Table 4.2 below, illustrates the frequency of observations from different rating classes and credit ratings.

Table 4.2: S&P Ratings

Investment grade		**High yield**	
Rating	Frequency	Rating	Frequency
AAA	411	BB+	1,834
AA+	309	BB	1,574
AA	631	BB-	1,802
AA-	1,447	B+	1,937
A+	2,304	B	1,874
A	3,867	B-	1,774
A-	3,724	CCC+	1,224
BBB+	4,233	CCC	709
BBB	4,635	CCC-	473
BBB-	3,561	CC	410
		C	174
		D	550
		Missing	1,109
		Total	**40,566**

Some of the observations in our sample are missing their Standard & Poor rating and are therefore listed as missing. In addition, some bonds have migrated between rating classes during the sample period and will therefore generate more than one bond/rating combination. In total, we end up with 40,566 observations where 64% of the bond/rating combinations are registered as investment grade and the remaining 36% are high yield.

4.3.2 Risk commonality measures

In the table below, we list summary statistics for our risk commonality measures. In total, we have 167 observations of the four dispersion measures. For three of the 12-month rolling average pairwise correlation measures, we have 156 observations. In order to generate a 12-month rolling correlation, we need to make use of eleven previous data points of the underlying variable, and hence we have no recorded correlation for the first eleven months of these time series. For the Amihud illiquidity change correlation, we have 155 observations as this is constructed by using the price changes from the first month to the following to get the Amihud illiquidity measure. When constructing the change in Amihud illiquidity between two periods, we lose another data point. The maximum observation of volume change and yield change dispersion was recorded in October 2008. Return dispersion reached its maximum value in December 2008, while the maximum value for logarithmic yield change dispersion was reached in September the same year. Apparently, the financial crisis influenced these variables to a large extent, as periods with market distress may have diverse impacts on securities within the same asset class. For instance, investors are prone to seek safer assets during periods of market turmoil.

Table 4.3 Descriptive statistics of risk commonality measures

Risk commonality measure	N	Mean	St. Dev.	Min	Median	Max
Volume change dispersion	167	1.648	0.172	1.282	1.681	1.973
Return dispersion	167	0.037	0.024	0.013	0.029	0.135
Yield change dispersion	167	0.022	0.012	0.005	0.019	0.069
Ln (yield change) dispersion	167	0.191	0.055	0.068	0.202	0.309
Volume change correlation	156	0.021	0.007	0.004	0.021	0.037
Return correlation	156	0.210	0.066	0.097	0.198	0.380
Log yield correlation	156	0.021	0.007	0.004	0.021	0.037
Amihud illiquidity correlation	155	0.047	0.017	0.018	0.044	0.085

4.3.3 ETF activity summary statistics

Table 4.4 lists summary statistics for each ETF activity variable. For the time series variables, we have 167 monthly observations from July 2002 to June 2016. In the data period, the ETFs' market share grew from 0.1 % to 4.9% for aggregate and corporate bond ETFs combined and reached 2.5% at the end of the period for corporate bond ETFs. From the flow measures, it can be observed that the maximum observed flow as a percentage of ETF AUM was almost 40%. This observation was from august 2002 the month after iShares Iboxx $ investment grade which is the oldest ETF in our sample launched. In the sample period the mean monthly flows have been close to 3% of AUM for both ETF subsamples. Daily turnover has a mean of 2% in the sample period with a maximum value recorded at 10.2% in June 2013. June 2013 was when Governor Ben Bernanke unveiled plans to phase out the Federal Reserve's quantitative easing program, which caused mass outflows from bond funds (Nyaradi, 2013). Standard deviation of ETF shares has kept at 2.5% on average over the period and peaked at 33.4% in October 2003 when the second fund in our sample was launched.

Table 4.4: Descriptive statistics ETF activity measures

ETF activity variables	N	Mean	St. Dev.	Min	Median	Max
ETF% Aggregate bond ETFs	167	0.018	0.016	0.001	0.014	0.049
ETF% Corporate bond ETFs	167	0.010	0.009	0.001	0.008	0.025
ETF flows Aggregate bond ETFs	167	0.033	0.039	-0.052	0.027	0.396
ETF flows Corporate bond ETFs	167	0.030	0.054	-0.074	0.022	0.396
ETF turnover	167	0.021	0.016	0.004	0.016	0.102
SD shares	167	0.025	0.033	0.000	0.018	0.334

5. Results

In the following chapter, we present, interpret and discuss the results. The first section consists of results from our initial exploratory analysis of the commonality and ETF activity variables. In section 5.2, the results from our time series regressions are presented, while we show the results from our panel regressions in section 5.3. The results from robustness tests on time series and panel models are displayed in section 5.4. In section 5.5 we interpret the results in relation to our initial research question and discuss limitations of our approach. We use p-values for significance throughout our analysis.

5.1 Exploratory analysis

Note that in the following where we do not separate the sample into investment grade and high yield sample, ETF variables named "all" such as *ETF% all* and *Flows_all*, refers to both corporate bond ETFs and aggregate bond ETFs. For simplicity, we call them "all" when the variables are constructed using numbers from both corporate and aggregate bond ETFs, where the latter can invest in both government and corporate bonds.

5.1.1 Correlation matrix

We start our analysis by looking at a correlation matrix of our ETF activity measures and the commonality measures in the U.S. corporate bond market. First, we have a look at the ETF activity measures. At first glance, we observe that *ETF%_all* and *ETF%_corp* are almost perfectly correlated. This is expected since *ETF%_all* represents the ratio ETF assets under management as a share of the bond market for both aggregate bond ETFs and corporate bond ETFs. Next, we find *Flows* to be negatively correlated with *ETF%*. *ETFturnover* is positively correlated with *ETF%*, indicating that trading activity in the ETF increases with the ETFs' relative market size. Creation/redemption activity measured by *SDshares* has no correlation with *ETF%* but has a weak positive correlation with *Flows* and *ETFturnover*.

Table 5.1 Correlation matrix of variables

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
ETF%_all	(1)	1													
ETF%_corp	(2)	0.994	1												
Flows_all	(3)	-0.366	-0.387	1											
Flows_corp	(4)	-0.116	-0.139	0.776	1										
ETFturnover	(5)	0.363	0.403	-0.047	0.02	1									
SDshares	(6)	-0.007	-0.009	0.272	0.12	0.383	1								
VolDisp	(7)	-0.902	-0.897	0.453	0.299	-0.219	0.023	1							
RetDisp	(8)	-0.144	-0.184	0.466	0.678	-0.174	0.026	0.28	1						
YDisp	(9)	-0.228	-0.261	0.43	0.626	-0.182	0.022	0.378	0.872	1					
LogYDisp	(10)	0.406	0.434	-0.05	0.278	0.322	0.027	-0.278	0.352	0.381	1				
Vcorr	(11)	-0.115	-0.111	0.091	0.171	0.158	0.061	0.251	-0.037	0.073	-0.005	1			
Rcorr	(12)	-0.007	0.006	0.037	0.114	-0.022	0.015	0.063	0.324	0.25	0.059	-0.194	1		
Ycorr	(13)	0.13	0.142	-0.024	0.098	0.028	0.018	-0.066	0.335	0.249	0.139	-0.215	0.972	1	
Lcorr	(14)	-0.019	-0.012	-0.105	-0.164	-0.152	-0.029	-0.081	0.064	0.009	-0.165	-0.264	0.726	0.733	1

We find a negative correlation between three out of four dispersion measures and the ETF activity measures *ETF%* and *ETFturnover*, but they seem to have a positive relationship with ETF flows. The 12-month rolling volume change correlation seems to be weakly negatively correlated with *ETF%* but has a weak positive relationship with the rest of the ETF activity variables. Return correlation (*Rcorr*) has no indication of relationship with *ETF%*, but it has a weak positive correlation with flows. However, between *Rcorr* and both turnover and creation/redemption activity the correlation coefficients are small. Yield-change correlation (*Ycorr*) has a weak positive correlation with *ETF%*, negative with *Flows_all* and weak positive with *Flows_corp*. However, the relationship between yield change correlation and return correlation is strongly positive. Liquidity correlation (*Lcorr*) seems to be negatively correlated with all of the ETF activity measures. We find the highest correlation between any ETF activity measure and the bond commonality measures to be between *Flows_corp,* and *RetDisp* and *YDisp*, with correlation coefficients of 0.678 and 0.626, respectively. The lowest correlation estimated, is between *VolDisp* and *ETF%* with correlation coefficients around -0.90.

5.1.2 Naïve OLS

In the following, we present univariate ordinary least squares regressions where we look at the relationship between the previously mentioned commonality measures in the bond market and the ETF activity measures. This first stage of analysis contains naïve OLS regressions, not taking into account the necessary properties of a time series, such as stationarity in the variables. We present naïve OLS regression output of the dispersion measures in the tables 5.2 and 5.3 below.

Table 5.2 Naïve OLS on return and volume-change dispersion

	Dependent variable:											
	RetDisp						VolDisp					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ETF% all	-0.3428						-9.6115***					
	(0.3246)						(0.0000)					
ETF% corp		-0.6839						-16.8212***				
		(0.2775)						(0.0000)				
ETFflows all			0.2552**						1.6768***			
			(0.0197)						(0.0000)			
ETFflows corp				0.2812***						1.0266***		
				(0.0002)						(0.0002)		
ETFturnover					-0.3035*						-2.5569*	
					(0.0647)						(0.0963)	
SDshares						0.0065						0.0601
						(0.8675)						(0.8702)
Constant	0.0430***	0.0436***	0.0286***	0.0286***	0.0434***	0.0368***	1.8179***	1.8130***	1.5930***	1.6176***	1.7029***	1.6465***
	(0.0004)	(0.0005)	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	167	167	167	167	167	167	167	167	167	167	167	167
R^2	0.0524	0.0675	0.1687	0.3893	0.0418	0.0001	0.8210	0.8126	0.1450	0.1033	0.0591	0.0001
Adjusted R^2	0.0467	0.0618	0.1636	0.3856	0.0360	-0.0060	0.8199	0.8114	0.1398	0.0979	0.0534	-0.0059
Residual Std. Error	0.0237	0.0235	0.0222	0.0190	0.0238	0.0243	0.0729	0.0746	0.1594	0.1632	0.1672	0.1723
F Statistic	9.1318***	11.9374***	33.4776***	105.1780***	7.2029***	0.0128	756.6904***	715.2948***	27.9841***	19.0136***	10.3657***	0.0218

Note: We use Newey-West HAC standard errors. P-values are reported in the brackets.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5.3 Naive OLS on simple yield change and logarithmic yield change dispersion

	Dependent variable:											
	YDisp						LogYDisp					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ETF% all	1.0252						-0.2146*					
	(0.4264)						(0.0510)					
ETF% corp		1.9933						-0.4086**				
		(0.2692)						(0.0484)				
ETFflows al			0.0796						0.1211***			
			(0.4243)						(0.0014)			
ETFflows corp				0.2940***						0.1286***		
				(0.0002)						(0.0000)		
ETFturnover					0.9154						-0.1465**	
					(0.1241)						(0.0139)	
SDshares						0.0209						0.0045
						(0.9068)						(0.8096)
Constant	0.1726***	0.1712***	0.1881***	0.1821***	0.1711***	0.1902***	0.0257***	0.0259***	0.0180***	0.0181***	0.0251***	0.0218***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	167	167	167	167	167	167	167	167	167	167	167	167
R^2	0.0897	0.1096	0.0031	0.0814	0.0728	0.0002	0.0881	0.1032	0.1628	0.3488	0.0418	0.0002
Adjusted R^2	0.0842	0.1042	-0.0029	0.0758	0.0671	-0.0059	0.0826	0.0978	0.1578	0.3448	0.0359	-0.0059
Residual Std. Error	0.0531	0.0525	0.0555	0.0533	0.0536	0.0556	0.0112	0.0111	0.0107	0.0095	0.0115	0.0117
F Statistic	16.2570***	20.3037***	0.5193	14.6148***	12.9467***	0.0252	15.9389***	18.9893***	32.0941***	88.3702***	7.1893***	0.0259

Note: We use Newey-West HAC standard errors. P-values are reported in the brackets.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We start by looking at return dispersion in table 5.2. We do not find ETFs' AUM-to-market cap ratio to be significant for neither corporate bond ETFs nor all bond ETFs. However, we find that net inflow into these funds tends to move in the same direction as return dispersion. The coefficients are significant at 5% level for both corporate and all bond ETFs. For corporate bond ETF flows, the model explains 38.9% of the variance in return dispersion, which is the highest recorded in the return dispersion OLS. We find that *ETFturnover* has a negative relationship with return dispersion significant at the 10% level. The proxy for creation/redemption activity, *SDshares,* seems to have no explanatory power at all when looking at return dispersion.

Turning to volume-change dispersion, which is the commonality measure for trading pattern in the underlying bond market, we find that ETFs' market share of the underlying market is negatively related and significant at the 1% level, both for corporate and all bond ETFs. This implies that ETFs holding a larger share of the U.S. corporate bond market tends to correspond with a lower volume change dispersion. With coefficients of -16.8 for corporate ETFs, it follows that a one percentage point increase in ownership share of underlying, implies a 16.8 percentage point lower volume change dispersion among corporate bonds. These two univariate regressions yield an R^2 of 81% and 82% for corporate and all bond ETFs, respectively. For fund net inflow, we find a positive relationship with volume change dispersion, also significant at 1% for both corporate and all bond ETFs. With a coefficient of 1.03 for corporate ETF flows, it follows that a one percentage point increase in net inflow to the funds would imply a 1.03 percentage point increase in volume change dispersion. This is unexpected given that fund inflow is likely to induce more block transactions of underlying securities. Block transaction might contribute to lower volume change dispersion as discussed in section 3.2.2. *ETFturnover* has a coefficient of -2.557 and is significant at the 10% level, while *SDshares* shows no explanatory power for volume change dispersion.

Next, we turn our attention to the OLS regression for yield change dispersion, both in simple differences and logarithmic differences. These results are reported in table 5.3. We find *ETF%_corp* to have a coefficient of -0.409 with significance at the 5% level, implying that a one percentage point increase in ownership of underlying securities to market capitalisation would imply a lower yield change dispersion of about 0.4 percentage points. For *ETF%_all* we get a coefficient of -0.215 that is significant at the 10% level, while *Flows_corp* and *Flows_all* are significant at 1% level with coefficients of 0.129 and 0.121, respectively. The *ETFturnover* coefficient is estimated to be -0.146, significant at 5%, implying that a percentage point increase in *ETFturnover*, is accompanied by a 0.146 percentage point decrease in yield change dispersion. For *SDshares* we find no evidence of impact as the coefficient is close to zero, a very low t-statistic and R^2 at zero. Further, we look at the logarithmic yield change dispersion. We only find *Flows_corp* to be significant with a coefficient 0.294 and a p-value below 1%. The general result for this commonality measure is low R^2 and little significance among the included explanatory variables.

The naïve OLS regression outputs of the 12-month rolling correlation measures, are shown in appendix C. In these models, we find no significant relationships between any of the ETF activity variables and the correlation measures.

5.2 Time series regressions

In the following section, we present the results from the time series regression of commonality measures in bonds regressed on different ETF activity measures. We conduct the regressions as described in section 3.1. This approach is inspired by similar studies in the stock market performed by Bolla et al. (2016) and Sullivan and Xiong (2012). We estimate two models for each of the commonality measures, where the first model is a static time series model looking at the contemporaneous relationship. The second model incorporates the lag structure of both dependent and independent variables, and the result is a so-called autoregressive distributed lag model (ARDL).

5.2.1 Investment grade sample

In this section, we provide results from the time series regression of investment grade bonds and ETFs with a specific mandate to invest in such bonds. We present results from estimating both time series models on the eight different commonality measures presented in section 3.2. We start by looking at the first two dispersion measures, return and volume-change dispersion.

Table 5.4 Time series regression: Investment grade dispersion measures (a)

	Dependent variable:			
	ΔVolDisp		ΔRetDisp	
	Model 1	Model 2	Model 1	Model 2
	(1)	(2)	(3)	(4)
ΔETFturnover	0.2562***	0.2426*	0.0419**	0.0429**
	(0.0005)	(0.0644)	(0.0205)	(0.0192)

ΔETFturnover, t-1		-0.2384		-0.0121
		(0.2217)		(0.4437)
ΔETFturnover, t-2		0.1191		-0.0035
		(0.4821)		(0.7314)
ΔSDshares,	0.0725	0.1089	-0.0088*	-0.0112**
	(0.6258)	(0.1837)	(0.0578)	(0.0318)
ΔSDshares, t-1		-0.0777**		0.0098
		(0.0206)		(0.1770)
ΔFlows	0.0392	-0.1260	0.0189	-0.0004
	(0.6132)	(0.2949)	(0.1691)	(0.9821)
ΔFlows, t-1		-0.2115		0.0011
		(0.1977)		(0.8736)
ΔFlows, t-2		0.0950		-0.0137**
		(0.2868)		(0.0285)
ΔVolDisp, t-1		0.5129***		
		(0.0000)		
ΔRetDisp, t-1				0.5097***
				(0.0000)
Observations	166	164	166	164
R^2	0.0163	0.4455	0.0253	0.5478
Adjusted R^2	-0.0018	0.4133	0.0073	0.5216
Residual Std. Error	0.0514	0.0393	0.0066	0.0046
F Statistic	0.9012	13.8388***	1.4091	20.8653***

Note: We use Newey-West HAC standard errors. All variables are on differenced form in order to obtain stationarity. Due to differencing, no constant is estimated.
P-values are reported in the brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The regression output table shows that *ETFturnover* have a positive significant relationship with *VolDisp* in the static model, while neither *SDshares* nor *Flows* is estimated to have a significant relationship with volume-change dispersion, having high p-values. In column 1, the coefficient is significant at the 1% level. As we have differentiated variables, the interpretation of the coefficient is how e.g. an increase in the change of the independent variable affects the change in the dependent variable. The exact interpretation of the coefficient's magnitude is then less straightforward. Thus, we emphasise the interpretation of significance levels and signs in our analysis. Turning to column 2 and the ARDL model, we find a negative relationship between the first lag of *SDshares* and volume-change dispersion that is significant at the 5% level, implying that an increase in the change of creation/redemption activity corresponds with a decrease in the change of volume dispersion.

We still find the change in *ETFturnover* to have a positive relationship with *VolDisp*. However, in this model it is significant at the 10% level. Adding a lag of the dependent variable is necessary due the lag structure and consequently, we observe that R^2 increases from 1.6% in model 1 to 44.6% in model 2.

We observe from table 5.4 that the static model in column 1 estimates a positive significant relationship between *ETFturnover* and *RetDisp*, with 5% significance. *SDshares* is significant at the 10% level with a coefficient implying a negative relationship return dispersion. We add the relevant lags of the dependent and independent variables in the ARDL model (model 2), and this model confirms that *ETFturnover* and *SDshares* have the same relationship and approximately the same magnitude as in model 1. Additionally, the second lag of *Flows* is estimated to have a negative coefficient at the 5% level, implying that an increase in net inflow of funds to ETFs corresponds to a decrease in change of return dispersion with a delay. We also observe that adding the lagged dependent variable has a large impact on the R^2 with an increase from 2.5% to 54.8%.

We report results from regressions of the next two dispersion measures, dispersion of yield change and dispersion of logarithmic yield change in the regression output table below.

Table 5.5 Time series regression: Investment grade dispersion measures (b)

	Dependent variable:			
	ΔYDisp		ΔLogYDisp	
	Model 1	Model 2	Model 1	Model 2
	(1)	(2)	(3)	(4)
ΔETFturnover	-0.0742**	-0.0870**	0.2711**	0.2845**
	(0.0209)	(0.0363)	(0.0196)	(0.0306)
ΔETFturnover, t-1		0.0779*		-0.1977**
		(0.0740)		(0.0270)
ΔETFturnover, t-2		-0.0475		-0.0526
		(0.2289)		(0.6126)
ΔSDshares	0.0133	0.0221**	-0.0098	0.0015
	(0.1595)	(0.0110)	(0.6924)	(0.9617)
ΔSDshares, t-1		-0.0108		0.0457*
		(0.2403)		(0.0697)
ΔFlows	0.0038	0.0069	0.0294	0.0155
	(0.8257)	(0.8103)	(0.5878)	(0.8593)
ΔFlows, t-1		-0.0095		-0.0112

		(0.6285)		(0.8693)
ΔFlows, t-2		0.0095		-0.0178
		(0.4375)		(0.5914)
ΔYDisp, t-1		0.5005***		
		(0.0000)		
ΔLogYDisp, t-1				0.5007***
				(0.0000)
Observations	166	164	166	164
R^2	0.0193	0.3220	0.0286	0.4246
Adjusted R^2	0.0012	0.2826	0.0107	0.3912
Residual Std. Error	0.0092	0.0079	0.0279	0.0219
F Statistic	1.0684	8.1775***	1.5992	12.7072***

Note: We use Newey-West HAC standard errors. All variables are on differenced form in order to obtain stationarity. Due to differencing, no constant is estimated. P-values are reported in the brackets.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We continue, by looking at the simple yield-change dispersion. From column 1, we observe an estimated coefficient of -0.074 for *ETFturnover* that is significant at the 5% level. This indicates a negative relationship between the turnover of ETF shares and the dispersion in yield-changes among the underlying investment grade bonds. *Flows* have a p-value of 0.8 and this indicates no significant statistical relationship with yield-change dispersion. *SDshares* has a positive estimated coefficient, but with a p-value of 0.16. However, this changes in model 2 where *SDshares* has an estimated coefficient of 0.02 and significant at the 5% level. This indicates that an increase in creation/redemption activity of ETF shares corresponds to an increase in yield-change dispersion. From the same column, we observe that the contemporaneous effect of *ETFturnover* on *YDisp* remains positive when expanding the model. Interestingly, the model estimates that the first lag of *ETFturnover* has the opposite sign but slightly lower magnitude. This coefficient is significant at the 10% with a p-value of 0.074. The effect of adding the first lag of the dependent variable is the same with the first two dispersion measures, where the coefficient is highly significant and the model's R^2 increase from 1.9% to 32.2%.

Turning to the dispersion measure of logarithmic yield-change in columns 3 and 4, the regression results in general differ from those in columns 1 and 2 where the estimated coefficients of *ETFturnover* have the opposite signs. This is an unexpected finding, since the direction of the yield-change should be the same for simple-differenced and logarithmic differenced. Hence, we cannot rule out measurement errors in this dependent variable. The contemporaneous variable is estimated to have a positive relationship with *LogYDisp*, while the first lag has a negative estimated relationship where both are significant at the 5% level. Further, we find that the first lag of creation/redemption activity proxied by *SDshares* have a positive relationship with *LogYDisp* with a significance of 10%. We also observe that inclusion of the lagged dependent variable increases R^2 from 2.9% to 42.5%.

Next, we present the results from the regressions of the first two 12-month rolling correlation measures in the output table below.

Table 5.6 Time series regression: Investment grade correlation measures (a)

	Dependent variable:			
	ΔVcorr		ΔRcorr	
	Model 1	Model 2	Model 1	Model 2
	(1)	(2)	(3)	(4)
ΔETFturnover	-0.0094	-0.0053	-0.2881**	-0.3594***
	(0.2608)	(0.5482)	(0.0293)	(0.0037)
ΔETFturnover, t-1		-0.0047		-0.3691*
		(0.7400)		(0.0536)
ΔETFturnover, t-2		-0.0025		0.7558**
		(0.8372)		(0.0265)
ΔSDshares	-0.0005	0.0002	0.0515	0.0470
	(0.8806)	(0.9486)	(0.2587)	(0.2686)
ΔSDshares, t-1		0.0012		-0.0407
		(0.5429)		(0.3099)
ΔFlows	0.0045	0.0077	0.0298	-0.0369
	(0.5140)	(0.4805)	(0.7996)	(0.7632)
ΔFlows, t-1		-0.0009		-0.0934
		(0.9345)		(0.2981)
ΔFlows, t-2		-0.0035		0.2113*
		(0.6494)		(0.0537)
ΔVcorr, t-11		0.0813***		
		(0.0006)		
ΔRcorr, t-11				0.1039***
				(0.0000)
Observations	154	143	154	143
R^2	0.0063	0.0830	0.0138	0.2069
Adjusted R^2	-0.0134	0.0214	-0.0057	0.1536
Residual Std. Error	0.0029	0.0028	0.0438	0.0409
F Statistic	0.3188	1.3471	0.7066	3.8840***

Note: We use Newey-West HAC standard errors. All variables are on differenced form in order to obtain stationarity. Due to differencing, no constant is estimated. P-values are reported in the brackets.
* p < 0.1, ** p < 0.05, *** p < 0.01

In the first two columns, we find that none of the ETF activity measures have any significant relationships with volume-change correlation of U.S. investment grade corporate bonds and these regression models have very low explanatory power.

The results from the regressions of return correlation display more significant coefficients, where *ETFturnover* is estimated to have a significant negative relationship at the 5% level with return correlation in model 1. The two other ETF activity measures have no clear estimated relationship with return correlation. The ARDL model in column 4, shows that the contemporaneous variable and both lags of *ETFturnover* have significant estimated coefficients. The second lag is significant at 5% with a coefficient of 0.756, while the contemporaneous and first lag have coefficients of -0.359 and -0.369 with significance at the 1% and 10% level, respectively. We also find that the second lag of *Flows* to have a positive coefficient that is significant at the 10% level.

In the table below, we present the last regression output of investment grade corporate bonds from the estimations of 12-month rolling correlations of yield change and Amihud illiquidity change.

Table 5.7 Time series regression: Investment grade correlation measures (b)

	Dependent variable:			
	ΔLcorr		ΔYcorr	
	Model 1	Model 2	Model 1	Model 2
	(1)	(2)	(3)	(4)
ΔETFturnover	0.0044	-0.0170	-0.1486	-0.2007
	(0.8875)	(0.6394)	(0.3043)	(0.2183)
ΔETFturnover, t-1		0.0558*		-0.3112*
		(0.0727)		(0.0574)
ΔETFturnover, t-2		-0.0153		0.6417**
		(0.5775)		(0.0336)
ΔSDshares	0.0162*	0.0198	0.0268	0.0067
	(0.0692)	(0.1824)	(0.4977)	(0.8518)
ΔSDshares, t-1		-0.0105		-0.0096
		(0.4087)		(0.8236)
ΔFlows	-0.0239	-0.0325	0.0597	-0.0002

	(0.2063)	(0.2797)	(0.5795)	(0.9984)
ΔFlows, t-1		-0.0342		-0.0501
		(0.2248)		(0.4481)
ΔFlows, t-2		0.0439***		0.1531*
		(0.0037)		(0.0889)
ΔLcorr, t-11		0.0633***		
		(0.0022)		
ΔYcorr, t-11				0.0950***
				(0.0001)
Observations	154	143	154	143
R^2	0.0125	0.0889	0.0059	0.1693
Adjusted R^2	-0.0071	0.0277	-0.0138	0.1135
Residual Std. Error	0.0102	0.0094	0.0421	0.0400
F Statistic	0.6387	1.4523	0.3007	3.0342***

Note: We use Newey-West HAC standard errors. All variables are on differenced form in order to obtain stationarity. Due to differencing, no constant is estimated. P-values are reported in the brackets.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Model 1, where we regress *Lcorr* on the ETF activity measures, shows that *SDshares* has a positive relationship with the commonality measure that is significant at the 10% level. This indicates that an increase in creation/redemption activity corresponds to an increase in the Amihud illiquidity change correlation among investment grade corporate debt securities. However, this relationship does not persist when estimating the full model in column 2 where the same variable has a p-value of 0.18. In this model, we find that the first lag of *ETFturnover* has a positive relationship with *Lcorr* that is significant at the 10% level. In addition, we find the second lag of *Flows* to have a positive relationship with the dependent variable with significance at the 1% level, indicating that increased growth rate in net inflow of funds to investment grade ETFs corresponds to an increase in the correlation of liquidity changes, with a delay.

The static regression model of *Ycorr* does not yield any significant estimates of the ETF activity measures. Thus, we turn to the ARDL model in column 4. In this model we find that the first and second lag of *ETFturnover* are both significant where they have 5% and 1% significance, respectively. The first lag has an estimated coefficient of -0.31, while the second lag has a coefficient of 0.64. This indicates that the first and second lag has opposite relationships with the correlation measure. In addition, we find that second lag of *Flows* have a positive relationship with yield-change correlation. This coefficient is significant at the 10%

level and is in line with the estimates in the *Lcorr* and *Rcorr* ARDL model with the same delay as observed for *Lcorr.*

5.2.2 High yield sample

In this section, we present the results from running the same regression models on the high yield sample. In addition, we compare the results with the output from the investment grade sample. As with investment grade, we start by looking at the dispersion measures. The results from volume-change and return dispersion are presented in the table below.

Table 5.8 Time series regression: High yield dispersion measures (a)

	Dependent variable:			
	ΔVolDisp		ΔRetDisp	
	Model 1	Model 2	Model 1	Model 2
	(1)	(2)	(3)	(4)
ΔETFturnover	0.7695^{**}	1.1335^{**}	0.0808^{*}	0.1312
	(0.0168)	(0.0314)	(0.0831)	(0.1098)
ΔETFturnover, t-1		-0.6675		-0.2398^{***}
		(0.2030)		(0.0045)
ΔETFturnover, t-2		0.0473		0.0801^{*}
		(0.7589)		(0.0705)
ΔSDshares	-0.0767	-0.1417	0.0213^{**}	0.0401^{**}
	(0.2160)	(0.1121)	(0.0224)	(0.0149)
ΔSDshares, t-1		0.1992^{*}		0.0176
		(0.0883)		(0.5678)
ΔFlows	-0.0752	-0.1189	0.0022	0.0059
	(0.2412)	(0.1492)	(0.8144)	(0.4155)
ΔFlows, t-1		0.0342		-0.0112
		(0.5880)		(0.3052)
ΔFlows, t-2		-0.0078		-0.0073
		(0.8468)		(0.4427)
ΔVolDisp, t-1		0.4956^{***}		
		(0.0000)		
ΔRetDisp, t-1				0.4934^{***}
				(0.0000)
Observations	109	107	109	107
R^2	0.0469	0.4427	0.0313	0.6301

Adjusted R^2	0.0200	0.3915	0.0039	0.5961
Residual Std. Error	0.0726	0.0577	0.0151	0.0097
F Statistic	1.7404	8.6495^{***}	1.1428	18.5499^{***}

Note: We use Newey-West HAC standard errors. All variables are on differenced form in order to obtain stationarity. Due to differencing, no constant is estimated. P-values are reported in the brackets.
* p < 0.1, ** p < 0.05, *** p < 0.01

We start by looking at the static model in column 1. This regression model displays a positive relationship between *ETFturnover* and volume-change dispersion that is significant at the 5% level with a p-value of 0.0168 and a coefficient of 0.7695. Neither *SDshares* nor *Flows* are significant in this model with p-values of 0.216 and 0.2412, respectively. Their p-values decrease somewhat in model 2, but not enough to conclude with a significant statistical relationship. However, the first lag of *SDshares* is significant at the 10% level with an estimated coefficient of 0.1992, indicating a positive relationship with *VolDisp*. This coefficient is of the opposite sign than the same variable for the investment grade sample. *ETFturnover* remains significant at 5% level in model 2 but has a higher estimated coefficient in this model of 1.1335, which is larger than the same coefficient for the IG sample (0.2426). We also find that adding lags of the dependent and independent variables to the model increase the R^2 from 2.0% to 39.2%, which is consistent with the findings in the IG sample.

The static model of return dispersion, estimates a positive relationship between *ETFturnover* and the dependent variable that is significant at the 10% level. *SDshares* has a coefficient of 0.0213 that is significant at 5%, while *Flows* show no indication of any relationship with return dispersion. In model 2, the contemporaneous variable of *ETFturnover* has a p-value of 0.1098 and hence lose significance. However, for the first lag we estimate a negative relationship with *RetDisp* that is significant at the 1% level, indicating that increased trading activity in the ETF shares corresponds to a lower dispersion in returns of high yield bonds. *SDshares* remain positively related to *RetDisp* at the 5% level, while *Flows* also in this model show no indication of any relationship with the dependent variable. We observe that model 2 has an R^2 of 63.0% compared to 3.1% for model 1.

Next, we turn to simple yield-change dispersion and log yield-change dispersion. We report these results in the table 5.9.

Table 5.9 Time series regression: High yield dispersion measures (b)

	Dependent variable:			
	ΔYDisp		ΔLogYDisp	
	Model 1	Model 2	Model 1	Model 2
	(1)	(2)	(3)	(4)
ΔETFturnover	0.0448	-0.0059	0.2835***	0.3402*
	(0.2761)	(0.9506)	(0.0015)	(0.0583)
ΔETFturnover, t-1		0.0181		-0.5750**
		(0.8786)		(0.0207)
ΔETFturnover, t-2		0.0184		0.2645***
		(0.6447)		(0.0027)
ΔSDshares	-0.0394***	-0.0346*	-0.0724***	-0.0030
	(0.0000)	(0.0956)	(0.0009)	(0.9145)
ΔSDshares, t-1		0.0014		0.0515
		(0.9507)		(0.4665)
ΔFlows	-0.0039	-0.0009	0.0082	0.0244
	(0.4853)	(0.8860)	(0.6778)	(0.2871)
ΔFlows, t-1		-0.0003		-0.0042
		(0.9759)		(0.8634)
ΔFlows, t-2		0.0072		-0.0048
		(0.2441)		(0.8298)
ΔYDisp, t-1		0.5156***		
		(0.0000)		
ΔLogYDisp, t-1				0.4863***
				(0.0000)
Observations	109	107	109	107
R^2	0.0550	0.3075	0.0346	0.4607
Adjusted R^2	0.0282	0.2439	0.0073	0.4112
Residual Std. Error	0.0099	0.0088	0.0297	0.0231
F Statistic	2.0549	4.8360***	1.2679	9.3028***

Note: We use Newey-West HAC standard errors. All variables are on differenced form in order to obtain stationarity. Due to differencing, no constant is estimated. P-values are reported in the brackets.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The static regression model of yield-change dispersion in column 1 estimates *SDshares* to have a negative relationship with the dependent variable that is significant at the 1% level, while none of the other two ETF activity measures' coefficients are significant. In the ARDL model in column 2, *SDshares* still has a negative relationship with *YDisp* but is significant at 10% level in this model. Compared to the same model in the IG sample, we find a coefficient of *SDshares* that is of the opposite sign. In the IG sample, we found *ETFturnover* to have a significant relationship with *YDisp*, but this is not the case for the HY sample where none of the *ETFturnover* variables have significant relationships with the dependent variable.

Next, we address the time series regression of *LogYDisp*. For this dependent variable, we identify a positive relationship of *ETFturnover* and a negative relationship of *SDshares* in the static model. Both coefficients are significant at the 1% level. The size of the *ETFturnover* coefficient is very similar to the estimation in the IG sample, whereas *SDshares* insignificant in that sample. Model 2 changes our perception of the relationship between *SDshares* and *LogYDisp*, as the coefficient show no indication of significance in this model with a p-value above 0.9. However, for *ETFturnover*, we still find a significant relationship with a positive coefficient for the contemporaneous variable and the second lag, while the first lag is negative. The rest of the included ETF activity variables are insignificant.

Table 5.10 presents the results from the 12-month rolling correlations of volume-change and returns of high yield bonds.

Table 5.10 Time series regression: High yield correlation measures (a)

	Dependent variable:			
	ΔVcorr		ΔRcorr	
	Model 1	Model 2	Model 1	Model 2
	(1)	(2)	(3)	(4)
ΔETFturnover	0.5541	2.2399***	-0.0672***	-0.0543
	(0.1847)	(0.0000)	(0.0000)	(0.4586)
ΔETFturnover, t-1		-0.0767***		0.0195
		(0.0000)		(0.8157)
ΔETFturnover, t-2		0.3763***		-0.0256
		(0.0000)		(0.3682)
ΔSDshares	-0.1104	-0.2835***	0.0023	0.0070
	(0.1176)	(0.0000)	(0.5178)	(0.8075)
ΔSDshares, t-1		-0.1328***		-0.0112
		(0.0000)		(0.5754)
ΔFlows	-0.0323	0.0326***	-0.0041*	-0.0029
	(0.2909)	(0.0029)	(0.0787)	(0.7152)
ΔFlows, t-1		-0.0877***		0.0111
		(0.0000)		(0.3581)
ΔFlows, t-2		0.0493***		-0.0113
		(0.0000)		(0.2013)

ΔVcorr, t-11		0.1058^{***}		
		(0.0000)		
ΔRcorr, t-11				0.0593^{***}
				(0.0024)
Observations	109	98	109	98
R^2	0.0295	0.2069	0.0692	0.0835
Adjusted R^2	0.0020	0.1267	0.0428	-0.0092
Residual Std. Error	0.0586	0.0555	0.0050	0.0052
F Statistic	1.0728	2.5801^{**}	2.6252^{*}	0.9007

Note: We use Newey-West HAC standard errors. All variables are on differenced form in order to obtain stationarity. Due to differencing, no constant is estimated. P-values are reported in the brackets.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The static model of the 12-month rolling volume-change correlation, estimates no significant coefficients for the ETF activity variables. The ARDL model returns a contrasting output. The contemporaneous *ETFturnover* variable is significant at the 1% level with a positive coefficient of 2.399, which is quite different from the IG regression where no significant relationship is estimated. In addition, both lags are significant at 1% with negative and positive coefficients, respectively. *SDshares* is also estimated to have a negative relationship with *Vcorr* at the highest significance level, where this applies for both the contemporaneous and lagged variable. For the ETF activity measure *Flows* we also estimate all variables to be significant at 1% level, where the contemporaneous and second lag shows a negative relationship with *Vcorr*, while the first lag is of the opposite sign. We note that all explanatory variables are significant at the 1% level in this model. This is something we want to investigate closer in the robustness section.

The static model of *Rcorr* estimates a negative relationship with *ETFturnover* that is significant at the 1% level, while *Flows* is estimated to have a negative relationship at the 10% level. However, these variables turn insignificant in the ARDL model in column 4. In the full model, we find none of the ETF activity measures to have significant relationships with the 12-month rolling return correlation for high yield bonds. The model explains little of the variation in return correlation compared to the investment grade sample. The IG regression displays significance on all *ETFturnover* variables in addition to the second lag of *Flows.*

The last time series regression output is reported in the table below, looking at the 12-month rolling correlations of yield-change and Amihud illiquidity change.

Table 5.11 Time series regression: High yield correlation measures (b)

	Dependent variable:			
	ΔLcorr		ΔYcorr	
	Model 1	Model 2	Model 1	Model 2
	(1)	(2)	(3)	(4)
ΔETFturnover	-0.0039	0.1596	0.5866	2.3249***
	(0.9476)	(0.1874)	(0.1707)	(0.0003)
ΔETFturnover, t-1		0.1909*		-0.2219
		(0.0786)		(0.8383)
ΔETFturnover, t-2		-0.0763		0.4823
		(0.2816)		(0.2491)
ΔSDshares	-0.0054	0.0324	-0.1192	-0.3411*
	(0.5774)	(0.4085)	(0.1201)	(0.0848)
ΔSDshares, t-1		-0.0370		-0.0713
		(0.3333)		(0.8187)
ΔFlows	-0.0021	-0.0066	-0.0219	0.0357
	(0.7155)	(0.5412)	(0.4400)	(0.4587)
ΔFlows, t-1		-0.0014		-0.0750
		(0.9012)		(0.1588)
ΔFlows, t-2		0.0051		0.0627
		(0.6114)		(0.1908)
ΔLcorr, t-11		0.0853***		
		(0.0001)		
ΔYcorr, t-11				0.0808**
				(0.0120)
Observations	109	98	109	98
R^2	0.0018	0.1410	0.0308	0.1962
Adjusted R^2	-0.0264	0.0542	0.0033	0.1149
Residual Std. Error	0.0106	0.0103	0.0591	0.0562
F Statistic	0.0640	1.6236	1.1211	2.4132**

Note: We use Newey-West HAC standard errors. All variables are on differenced form in order to obtain stationarity. Due to differencing, no constant is estimated. P-values are reported in the brackets.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The static model of *Lcorr* returns no significant coefficients of the ETF activity measures. The ARDL model estimates a positive relationship between the first lag of *ETFturnover*, however, it is only significant at 10% level. In general, the regression models of *Lcorr* for the high yield sample suggest no statistical relationship between the variables, which is slightly different from the models for the IG sample. Those models produce significant coefficients for *SDshares* in the static model and for the first lag of *ETFturnover* and second lag of *Flows* in the ARDL model.

The static model for yield-change correlation estimates no significant relationships between the ETF activity measures and the dependent variable. The ARDL model produces a positive coefficient of 2.3249 for the contemporaneous *ETFturnover* variable that is significant at the 1% level. Additionally, *SDshares* is estimated to have a coefficient of -0.3411 that is significant at the 10% level. We find no other statistically significant relationships in the HY sample, whereas for the IG sample both the first and second lag of *ETFturnover* were significant along with the second lag of *Flows*.

5.3 Panel Regressions

In this part, we present the results from the panel regressions of the bond-index correlation measure and ETF activity measures. As described under 3.2 we use ownership and trading activity information from two of the largest corporate bond ETFs in their respective markets and investigate the significance of these measures on the correlation of constituent bonds and the market index. The results from the two panel regression models include different fixed effects and control variables. We look at the two models separately for the investment grade and high yield sub-sample.

5.3.1 Investment grade subsample

We begin by looking at the panel regressions for Model 1 on the investment grade sample. For model 1, we run four different regressions, pooled OLS without fixed effects (1), panel regression with bond fixed effects (2), panel regression with time fixed effects (3) and a regression with both time and individual fixed effects (4). As explained in section 3.2.3 we drop some of the explanatory variables when time fixed effects are introduced to not introduce a potential element of bias. The regression output is printed in table 5.12.

Table 5.12 Panel regressions: Model 1 Investment grade

	Dependent variable:			
	$Y = Corr_{i,t}$			
	Pooled OLS	Bond FE	Time FE	2-way FE
	(1)	(2)	(3)	(4)
ETFshare	1.9249***	0.1036	2.7565***	0.7839***
	(0.0000)	(0.6021)	(0.0000)	(0.0002)
ETFturnover	0.6181	-0.6313		
	(0.2269)	(0.2385)		
SDshares	2.1200***	2.7249***		
	(0.0000)	(0.0000)		
Flows	-0.3281***	-0.3211***		
	(0.0000)	(0.0000)		
Rating	-0.0159***	-0.0132***	-0.0158***	-0.0113***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Duration	0.0145***	0.0387***	0.0137***	0.0384***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Constant	0.6556***			
	(0.0000)			
Observations	105,729	105,729	105,729	105,729
R^2	0.1088	0.0729	0.1187	0.0350
Adjusted R^2	0.1087	0.0534	0.1173	0.0133
F Statistic	2,150.6400***	1,356.2170***	4,737.7390***	1,251.2940***

Note: P-values in brackets. Standard errors are clustered at the individual level.
$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

From the regression output in table 5.12, we observe that *ETFshare* which indicates the percentage of a corporate bonds principal amount that is owned by LQD, is significant at the 1% level for three of the model specifications. The coefficient of *ETFshare* is positive in all model specifications indicating that bonds that have a higher ETF ownership tend to be more correlated with the market index. A coefficient of 0.78 for *ETFshare* as estimated by the two-way FE model indicates that bonds with a 10% higher ETF ownership, ceteris paribus, have a correlation with the market index that is 0.078 higher. We further observe that the coefficient of the monthly standard deviation of ETF shares has a positive sign. This indicates that creation and redemption activity in exchange traded funds may increase the correlation of underlying basket securities. *SDshares* is significant for all model specifications where it is included. For turnover, we observe that the variable is not significant at any of the significance levels. *Flows* appears to have a negative effect on correlation and is significant at the 1% level in regression (1) and (2), indicating that months with high net ETF flows tend to coincide with a reduction of bond correlations. We also find the two control variables to be highly significant in all the regression models. We observe the ETF ownership of bonds seems to influence bond-index correlation, while turnover, which we expected to have an effect, lacked significance. This indicates that ETF ownership possibly is the most relevant of the included ETF variables in explaining variations in the correlation measure.

For model 2 we create new variables based on the monthly ETF bond ownership, the resulting output is illustrated in table 5.13.

Table 5.13 Panel regressions: Model 2 Investment grade

	Dependent variable:					
	Y = $Corr_{i,t}$					
	Bond FE (1)	Bond FE (2)	Bond FE (3)	2-way (4)	2-way (5)	2-way (6)
Turnover%	44.6914^{***} (0.0015)			95.5360^{***} (0.0000)		
SDshares%		53.9597^{***} (0.0000)			40.3438^{***} (0.0000)	
Flows%			-3.7876^{*} (0.0585)			-0.3872 (0.8325)
Rating	-0.0135^{***} (0.0000)	-0.0134^{***} (0.0000)	-0.0138^{***} (0.0000)	-0.0113^{***} (0.0000)	-0.0116^{***} (0.0000)	-0.0119^{***} (0.0000)
Duration	0.0388^{***} (0.0000)	0.0387^{***} (0.0000)	0.0393^{***} (0.0000)	0.0378^{***} (0.0000)	0.0390^{***} (0.0000)	0.0393^{***} (0.0000)
Observations	105,729	105,729	105,729	105,729	105,729	105,729
R^2	0.0677	0.0683	0.0672	0.0364	0.0345	0.0338
Adjusted R^2	0.0481	0.0488	0.0477	0.0147	0.0127	0.0120
F Statistic	$2{,}504.7730^{***}$	$2{,}530.2910^{***}$	$2{,}488.5810^{***}$	$1{,}302.0690^{***}$	$1{,}229.9970^{***}$	$1{,}204.6310^{***}$

Note: P-values in brackets. Standard errors are clustered at the individual level.
$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

In model 2 we adjust *SDshares*, *ETFturnover* and *Flows* for bond ownership by multiplying each variable with the monthly *ETFshare* of each bond. This way we can investigate the effect of the ETF activity variables weighted by ownership, which vary between 0% and 18.6% in the underlying IG sample. For the investment grade sample *SDshares%* and *Turnover%* are significant at the 1% level in the panel regressions. Hence, when adjusting *ETFturnover* for ownership, we identify the relationship that we initially expected. The results seem reasonable since turnover and creation/redemption activity have been found to be relevant in previous research of other markets (Da & Shive, 2018). *Flows%* is significant at the 10% level for the bond fixed effect regression and shows no significance in the two-way regression. The findings indicate that increased trading activity of ETFs in either the primary (*SDshares%*) or secondary market (*Turnover%*) could increase underlying bonds correlation with the market index.

5.3.2 High yield subsample

For the high yield subsample, we use the same models and regression specifications as for investment grade. In the interpretation of the regression output, we make comparisons with the results from the investment grade subsample that are presented under 5.3.1. The output from model 1 is presented below.

Table 5.14 Panel regressions: Model 1 High yield

	Dependent variable:			
	$Y = Corr_{i,t}$			
	Pooled OLS	Bond FE	Time FE	2-way FE
	(1)	(2)	(3)	(4)
ETFshare	0.4792	-1.3822***	1.0775**	1.7406***
	(0.2121)	(0.0007)	(0.0299)	(0.0011)
ETFturnover	1.9062***	4.6632***		
	(0.0000)	(0.0000)		
SDshares	-0.0227	0.0658		
	(0.8881)	(0.6997)		
Flows	-0.0308	-0.2870***		
	(0.2098)	(0.0000)		
Rating	0.0015	-0.0115***	0.0025	-0.0022
	(0.4720)	(0.0011)	(0.2323)	(0.6579)
Duration	0.0394***	0.0524***	0.0395***	0.0614***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Constant	0.2614***			
	(0.0000)			
Observations	39,106	39,106	39,106	39,106
R^2	0.0589	0.1211	0.0500	0.1089

Adjusted R^2	0.0587	0.0909	0.0472	0.0756
F Statistic	407.6211***	868.3368***	683.7388***	1,535.7410***

Note: P-values in brackets. Standard errors are clustered at the individual level.
*p < 0.1, **p < 0.05, ***p < 0.01

For the HY subsample we observe that the coefficient of *ETFshare* is more difficult to interpret. For instance, the sign of the coefficient is negative in specification (2) and the variable is not significant in specification (1). However, *ETFshare* is significant in the 2-way specification that accounts for both time and bond fixed effects. It is also interesting to note that the coefficient of *ETFshare* is 1.74 compared to 0.78 for the IG bond. This indicates that ETF ownership have a larger impact on the correlation between HY bonds and the corresponding index. In addition, compared to the model 1 regressions of the IG bonds where turnover was not significant, we find high significance and a positive coefficient for the high yield bonds. The coefficient of the *Flows* variable is negative and significant at the 1% level for the bond fixed effect regression, which is also the case for the same regression specification in the IG subsample. In addition, the credit rating control variable seems to carry less relevance for the correlation of individual HY bonds compared to IG bonds and is not significant in any of the specifications. Further, the duration control variable is significant in all model specifications. In the following table, we present the regression output for model 2 for the high yield subsample.

Table 5.15 Panel regressions: Model 2 High yield

	Dependent variable:					
	Y = $Corr_{i,t}$					
	Bond FE	Bond FE	Bond FE	2-way	2-way	2-way
	(1)	(2)	(3)	(4)	(5)	(6)
Turnover%	14.5306**			51.5259***		
	(0.0144)			(0.0000)		
SDshares%		39.1172***			33.0495***	
		(0.0000)			(0.0000)	
Flows%			6.8121***			3.8696***
			(0.0000)			(0.0005)
Rating	0.0168***	0.0169***	0.0171***	0.0153***	0.0156***	0.0156***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Duration	0.0434***	0.0421***	0.0420***	0.0430***	0.0470***	0.0473***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Observations	46,455	46,455	46,455	46,455	46,455	46,455
R^2	0.0476	0.0497	0.0492	0.0232	0.0188	0.0176
Adjusted R^2	0.0203	0.0224	0.0219	-0.0083	-0.0129	-0.0140
F Statistic	752.8690***	786.6414***	778.6268***	356.1498***	287.1974***	269.1796***

Note: P-values in brackets. Standard errors are clustered at the individual level.
$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

For HY bonds, all ETF activity variables are significant at either the 1% or 5% level in model 2. The coefficients are more difficult to interpret in model 2 since they constitute of both the ETFs ownership of an individual bond and the aggregated ETF activity. However, we observe that the size of the coefficients for *Turnover%* and *SDshares%* in HY panel regressions are noticeably smaller than the same coefficients for the IG sample. This can indicate that turnover and creation/redemption activity drive less correlation in the high yield segment. This might for instance be due to higher liquidity in the IG bond markets making basket trades easier to implement. We also observe that *Flows%* is significant at the 1% for both the bond fixed effect and two-way regression. Hence, it appears as if the flows into the HYG fund are more influential compared to the LQD flows in the investment grade segment, where we found no significance at the 10% level in the two-way regression in 5.3.1.

5.4 Robustness tests

To investigate the robustness of our results we include fundamental variables in our time series regressions. This way we are able to disentangle the effects of ETF activity and investigate if changes in the comovement measures can be explained by fundamental factors. In order to perform these tests, we include the factors explained in section 2.3.1. As mentioned, these factors can possibly promote comovement in the cross section of securities. Additionally, we include a new control variable in the panel regressions and adjust the time window of the rolling correlation measure. This way we are able to investigate if the ETF measures are still significant under other conditions.

5.4.1 Time series regression

Our time series models do not control for fundamental factors that could potentially drive comovement of bonds. Therefore, we include the four fundamental variables discussed in section 2.3.1 to make the estimates robust to factors that are considered important in the pricing of corporate bonds. These four variables are investor sentiment proxied by the VIX index, inflation risk measured by unexpected inflation, interest rate risk that is measured by changes in 5-year constant-maturity treasuries and credit risk measured by changes in the monthly level of CDS indices for both investment grade and high yield bonds. The sample used in the robustness test of the investment grade sample has a shorter time span than the original model, as the credit spread is available from November 2005, while the original model use bond data from July 2002. Shortening the sample by three years might lead to some differences in the estimated coefficients, before adjusting for the fundamental factors. For the high yield sample, the sample period is the same as in the original model. The results from these robustness tests are shown in Appendix D.

Investment grade

We begin by looking at the investment grade sample. For *VolDisp*, we observe that *ETFturnover* increases both in magnitude and in significance. *SDshares* loses significance and thus it is more difficult to infer that creation/redemption activity drives volume-change dispersion. In return dispersion, *ETFturnover* loses some significance and magnitude, but is still significant at the 10% level. For the two dispersion measures of yield-changes, the ETF activity variables generally experience a small increase in coefficient magnitude. Additionally, the significance level is strengthened for the explanatory variables. For *YDisp*, the lagged variable of *SDshares* is significant at the 10% level as opposed to no significance in the original model. For correlation of bond returns, we observe that the contemporaneous *ETFturnover* is no longer significant. This is in stark contrast to the original model, where the variable is significant at the 1% level. However, the first lag keeps its significance level and the second lag is significant at 5%, opposed to 10% in the original model. The model of yield-change correlation shows that the first lag of *ETFturnover* increases in significance from 10% to 5%, while the second lag of *Flow* is no longer significant. For *Lcorr,* we find that *ETFturnover* is no longer a significant explanatory variable, while the second lag of *Flows* drop to 5% significance level with smaller coefficient magnitude. Of the fundamental variables, interest rate risk appears to be most influential with significant coefficients in five of the regression models. VIX turns out to be significant for one dispersion measure (*LogYDisp)*, while credit risk is significant for yield correlation.

High yield

Addressing the high yield sample, we begin with volume-change dispersion. We find that the contemporaneous *SDshares* variable is significant at the 10% level, while the other coefficients are not considerably affected by controlling for fundamental factors. For return dispersion, we observe that the contemporaneous variable of *ETFturnover* is significant at 10% level as opposed to no significance initially, while the second lag of the same variable loses significance. The two other dispersion measures are relatively consistent with the original findings, but *ETFturnover* has higher significance and a lower magnitude in the contemporaneous variable for *LogYDisp*. The robustness test of yield-change correlation indicates both a higher significance and higher magnitude of *SDshares*. The correlation of Amihud illiquidity change no longer has any significant explanatory variables due to the loss of significance in the first lag of *ETFturnover*. Interestingly, the robustness test has a major influence on the explanatory variables in the regression of volume-change correlation. In the original model, all ETF activity measures and their lags were significant at 1% but the robustness test returns most of these variables as insignificant. Two variables are left significant: The contemporaneous variable of *ETFturnover* that is significant at the 1% level with no considerable change in coefficient magnitude, and the contemporaneous *SDshares* variable that is significant at the 10% level and have a small increase in coefficient magnitude. The robustness test rejects the original model of *Vcorr* to a large extent, so that the model is more similar to that in the IG sample, which returned no significant ETF activity measures in the original model. This illustrates that explaining the volume-change correlation among corporate bonds is a difficult task. While the fundamental factors are significant for a range of commonality measures in the IG sample, this is only partly the case in the HY sample. Interest rate risk is only significant for *LogYDisp*, VIX is significant for *Lcorr*, while inflation risk exhibits some explanatory power for *Vcorr* and *Ycorr*.

5.4.2 Panel data regression

Inherently, the time fixed effect modification of the panel regression models adjusts for the effects from omitted variables that vary over time but not across individuals. Therefore, our two-way and time fixed effect models should already account for the effect of time-variant fundamental variables such as investor sentiment, inflation expectation, credit risk and interest rate risk. Including these variables in the model might be problematic since we introduce variables whose change over time is constant across individuals as commented in section 3.2.1. In our robustness tests, we use the two-way fixed effect specifications of model 2 where the ETF activity variables are weighted by ETF ownership.

To test the robustness of our panel data results we adjust the rolling window for the calculation of the dependent variable. This way we are able to investigate if the explanatory variables are robust to different lengths of the rolling window. We therefore calculate the 18 and 24 month rolling window correlation between the HY and IG market index and the individual bonds in the two subsamples. In addition, we add a new control variable, the monthly average trading spread for each individual bond, since the variable may add insight by reflecting the liquidity in underlying bonds (Bao, Pan, & Wang, 2008).

We present the results from the robustness tests in Appendix D. As illustrated, the significance and size of the coefficients for *SDshares%* and *Turnover%* does not change considerably when changing the rolling window of the dependent variable and including trading spread. The coefficient of the last ETF activity measure, *Flows%* increases considerably in size for the IG sample when increasing the length of the rolling window correlation between individual bonds and the market index. However, this variable is still not significant in the IG regressions when changing the calculation method of the dependent variable. When using alternative rolling windows of the correlation measure and adding a new control variable, our ETF measures remain significant. The results from the robustness test indicate that the ETFs probably have an effect on the correlation of underlying securities.

5.5 Discussion

In the following section, we summarise and discuss our key findings from the various methods, and link them to the problem statement and hypotheses in section 1.2. In addition, we include the findings from the different robustness tests in our interpretation. Finally, we discuss potential limitations to our approach. In the conclusion under chapter 6, we provide suggestions for future research on the topic of this thesis based on the identified limitations.

ETFs and variability in corporate bond commonality

As stated under section 1.2 our first hypothesis is: "*Measures of bond ETF activity can explain parts of the variation in the commonality of securities in the underlying market*". In order to investigate the hypothesis, we create measures of bond commonality based on both correlation and dispersion as explained in section 3.3.1. We further conduct various empirical tests on different bond samples. A mapping of the methods and their corresponding bond samples is provided under section 4.3.1.

Our initial exploration of the dependent and explanatory variables is in section 5.1.1 with the Pearson correlation matrix. Here, the highest correlations are identified between some of the bond dispersion measures and the activity measures *Flows* and *ETF%*. The *Flows* variable shows a substantial negative correlation with volume dispersion indicating that increasing flows into ETFs coincides with lower dispersion of monthly volume-changes in bonds. A pattern of similar volume-changes (i.e. lower dispersion) could be an indication of more automated or basket trading in the market. In addition, we find that flows into corporate bond funds exhibit a relatively high correlation (0.678 and 0.626) with return and yield-change dispersion. Hence, high dispersion tends to coincide with high ETF net inflows in the data period. In the naïve OLS regressions, our findings identify that the same corporate bond flows measure is strongly significant at the 1% level for return, volume-change, yield-change and log yield-change dispersion. No other ETF measures are significant for all commonality measures, but *ETFturnover* and *ETF%* show significance in some of the models. Without correcting for the time series properties of our variables, it seems like there are indications of a relationship between ETFs and bond commonalities.

For the time series regressions, we investigate the dependent and explanatory variables and use models that account for their time series properties, such as stationarity and lag structure. We also include fundamental factors[8] in robustness tests to investigate if the most relevant variables remained significant after including other potential sources of commonality in the models. The output from the robustness tests in Appendix D provides evidence that several ETF measures have a significant relationship with the commonality measures. Particularly *ETFturnover* shows strong significance, i.e. at 1% level, for several of the dispersion and correlation measures in both the high yield and investment grade sample. The intensity of creation/redemption activity as expressed in the *SDshares* measure, also shows significance for both dispersion and correlation measures. However, the significance level is generally higher for the turnover variable. An interesting finding is that the *Flows* measure displays low significance after adjusting for time series properties and the influence of fundamental factors in the bond market. The only specification of the fund flow measure that shows significance is the second lag (t-2) for return dispersion and liquidity correlation in the IG sample. Hence, in the time series models we find significant relationships between ETF related variables and

[8] The included fundamental factors and the way in which they might influence movements of corporate bonds is discussed in section 2.3.1

bond commonality measures. Our findings, moreover, suggest that the relationship is more pronounced for trading related measures, such as turnover of ETF shares. This resonates with our expectations and findings in other asset classes, such as the relationship Da & Shive (2018) identify in the equity market.

In the last part of the analysis, we investigate the ETF measures on subsamples of IG and HY bonds that are held by the iShares IG and HY ETF during the sample period. This way we are able to adjust the ETF measures for how much of each bond that is owned by the ETF. In the robustness test we change the rolling window length of the dependent variable, include one additional control variable to account for possible endogeneity issues. Accounting for these additions and the fixed effects, our key findings still indicate that particularly *ETFturnover* and *SDshares* are significant in explaining the variation in the commonality measure. Interestingly, *Flows%* is significant at the 1% level in model 2 for the HY sample but shows no significance for the IG sample. A possible explanation of this property could be that ETF flows lead to more transactions in the underlying high yield market, while this does not seem to be the case in the investment grade segment. This could be a result of differences in the structure of the HY and IG markets. For instance, higher liquidity in the IG segment may lower the limits to arbitrage and hence the direct impact of flows on underlying securities might be weaker. However, this is only speculative, and the true cause of this phenomenon remains unknown.

Our findings imply that the trading of ETF shares in the primary (*SDshares*) and secondary markets (*ETFturnover*) explain parts of the variation in commonality measures for individual bonds in the underlying markets. Clear direction for several relationships are hard to establish given our results, as the sign of certain coefficients for the explanatory variables vary between the different models. However, establishing the direction of the relationships is not a part of our initial research question. Controlling for fixed effects and fundamental factors also indicate that the comovement could be trading induced and not driven by fundamentals effects (Barberis, 2002). Our empirical findings suggest that there exists a relationship between the variables that we investigated and that hypothesis 1 is supported by our results. Limitations to our approach should be taken into consideration. There could for instance be other fundamental factors that we have not accounted for that are relevant. In addition, we have not established the direction of causality, which we provide as an idea for further research on the topic.

ETF turnover and corporate bond commonality

Our second hypothesis is: *"Turnover of ETF shares have more influence on commonality measures of securities in the underlying market compared to other ETF activity measures"*. We find *ETFturnover* to be a significant explanatory variable for commonality measures of bonds across different estimation techniques. However, the signs of the coefficients are ambiguous. While the naïve OLS regression suggests a negative relationship with the dispersion measures, the time series regressions do to a large extent contradict this with estimated positive relationships for three out of four dispersion measures in both samples. As the time series regressions adjust for potential problematic statistical properties in the data, we emphasise these results. The notion that *ETFturnover* as a proxy for arbitrage activity in the secondary market leads to lower dispersion in the underlying bond market, is not supported in the results and hence somewhat different from our expectations ex ante. However, the results indicate that there is a significant relationship with dispersion measures, but that higher turnover instead coincide with higher dispersion. If there is high dispersion among bonds, it may be more likely that strategies of bond picking will yield profitable returns compared to environments with low dispersion (Janus Henderson Investors, 2017). Thus, increased investments in ETFs that give the investor a broad exposure to the underlying securities, does not fit with this argument if the investor is profit maximising and have the ability to trade individual bonds. Hence, it is harder to picture the causality going from dispersion to turnover in ETFs. However, in periods of high market turmoil, dispersion among assets may increase as investors seek safer securities. It is also likely that ETF turnover is high in such periods, so that there might be a positive relationship between the two variables due to special market conditions. However, including the VIX in the robustness test does not change the result of a positive relationship.

For the high yield sample, we find that turnover in ETFs is strongly positively related to the correlations within the bond market, both in the time series and the panel regression. While the volume-change and yield-change correlation are the commonality measures with significant positive relationships in the time series, the panel regressions exclusively look at the correlation between the return of individual bonds and the market index. These results are to a large extent in line with the expectations ex ante. In the investment grade sample, the conclusions from time series estimation and panel regressions are more ambiguous. The time series regressions produce ambiguous estimates of *ETFturnover* and is significant for both lags for two of the correlation measures. We do, however, emphasise the results from the panel regressions as they provide more accurate inference of model parameters. In addition, panel regression simplifies computation and inference, especially in cases of nonstationary time series data (Hsiao, 2007). The results from the panel estimations clearly indicate that turnover is a significant variable in explaining the correlation of the individual bond with a strong positive relationship. Additionally, we find that the impact on index correlation is higher for investment grade than for high yield when weighing turnover with ownership in individual bonds. This point to ETF measures having a larger impact on correlation in the investment grade segment than for high yield. A possible explanation for this could be that IG bonds are more liquid (Liebschutz & Smith, 2016), and hence it is easier for ETFs to penetrate this segment as trading costs are lower, measured by e.g. bid/ask-spreads. This is also supported by the discussion in section 2.4, where higher trading costs are presented as potential explanatory factors of a higher tracking error among HY bond ETFs compared to IG bond ETFs.

Limitations

In general, comovement in financial securities is a complex phenomenon, as several known and unknown factors can have an influence (Chen, Chen, & Li, 2008). To our knowledge, there has been no research conducted previously on the relationship between bond ETFs and commonality in corporate bond returns, volume- and yield-changes. For this reason, we decide to employ a wide array of empirical methods inspired by research in other asset classes to investigate the relationship closer. We cannot rule out measurement errors in the calculation of dependent and independent variables, as most variables used in the estimation are calculated through algorithms instead of collected directly from a source. A general problem of assessing the effect of innovative products in financial markets is the timespan. As corporate bond ETFs are a relatively novel investment vehicle, the time period we look at in our research is limited. When conducting empirical research over a shorter time horizon there is a higher probability of finding spurious relationship, having noisy data and estimation errors (Gospodinov, 2017). A way to potentially alleviate these concerns is to use higher frequency data and also try to control for other factors that might influence the dependent variables. Causality issues is also a limitation of our approach. A common issue in the financial literature is to establish the direction of causal relationships. In our case, ETFs can for instance select bonds that are expected to have higher correlation with other bonds. This would mean that bond correlation may drive ETF activity and that the causal relationship has the opposite direction. We suggest that quasi-natural experiments can be conducted to establish the direction of causality. For instance Agarwal, Hanouna, Moussawi, & Stahel (2016) established the direction of causality between ETF ownership and comovement in stock liquidity by looking at index reconstitutions and how trading halts in ETFs affect stock comovement. The result from both experiments support the initial view of the researchers, that ETF trading activity drives stock comovement.

6. Conclusion

The introduction of ETFs in the financial markets has raised several questions on how the instruments influence the underlying securities. One of the questions is if ETFs introduce excessive basket trading in the markets, which in turn could lead to increased commonality between basket securities. In our research, we aim to investigate the relationship between fixed income ETFs and U.S. corporate bonds. Our intention is to examine if variation in bond commonality can be partly explained by ETF activity and if trading related ETF measures carry more significance as postulated in our research question. In our empirical research, we employ a wide array of methods and model specifications to investigate the relationship. We initially apply naïve OLS, but our main analysis consists of different time series and panel regressions.

Our key findings from the empirical analysis support our initial hypothesis and we find a significant relationship between ETF measures and corporate bond commonality. When controlling for fundamental factors that could drive commonality of corporate bonds in the time series models (e.g. credit and inflation risk), the turnover of ETF shares and creation/redemption activity estimates remain significant. In our panel models, we also find the same measures to be significant when including time and individual fixed effects as well as several control variables. Flows into corporate bond ETFs are also found to have significance on the variation of the commonality measure in the panel regressions on the high yield sample. Hence, in our main analysis we are able to establish a relationship between fixed income ETFs and the underlying corporate bonds in the market. Previous research suggests that trading related variables such as ETF turnover have a significant influence on underlying stocks (Da & Shive, 2018). For this reason, in the second part of our problem statement we set out to investigate the relevance of turnover on commonality in the U.S. corporate bond market. We find that the turnover of ETF shares has the strongest effect among the ETF activity variables. This impact seems to be larger in the investment grade segment.

For further research on the topic of the relationship between corporate bond ETFs and corporate bonds, we suggest taking into account the limitations we have addressed in our research. Other data frequencies such as weekly or for some more liquid bond samples even daily data could be included to reduce some of the limitations resulting from having a short sample period. In addition, the inclusion of other sources of ownership and trading activity such as both closed-end and open-end mutual funds and institutional investors could provide valuable insight. Extending our panel data analysis by including holding information from more ETFs could also be beneficial. A last suggestion is to address the direction of causality between ETF activity and bond commonality by performing quasi-natural experiments.

References

Agarwal, V., Hanouna, P., Moussawi, R., & Stahel, C. (2016). Do ETFs increase the commonality in liquidity of underlying stocks? Retrieved from http://finance.darden.virginia.edu/wp-content/uploads/2018/01/ETF_Paper_October_2017.pdf

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. Journal of Financial Markets, 5(1), 31-56.

Antoniewicz, R., & Heinrichs, J. (2015). The Role and Activities of Authorized Participants of Exchange-Traded Funds. Retrieved from https://www.ici.org/pdf/ppr_15_aps_etfs.pdf

Arellano, M. (1987). Computing Robust Standard Errors for Within-Groups Estimators. Oxford Bulletin of Economics and Statistics, 49(4), 431-434.

Asquith, P., Covert, T. R., & Pathak, P. A. (2013). The Effects of Mandatory Transparency in Financial Market Design: Evidence from the Corporate Bond Market. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1106852

Badoer, D. C. B., & Demiroglu, C. D. (2017). The Relevance of Credit Ratings in Transparent Bond Markets. Retrieved from https://www.eurofidai.org/sites/default/files/pdf/parismeeting/2017/badoer_2017.pdf

Bai, J., Bali, T. G., & Wen, Q. (2016). Do the Distributional Characteristics of Corporate Bonds Predict Their Future Returns? Retrieved from https://warwick.ac.uk/fac/soc/wbs/subjects/finance/events/seminars/bai-bali-wen.pdf

Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. The Journal of Finance, 61(4), 1645-1680.

Banerji, G. (2017). Institutional Investors Flock to Bond ETFs. Retrieved from https://www.wsj.com/articles/institutional-investors-flock-to-bond-etfs-1493085602

Bao, J., Pan, J., & Wang, J. (2008). Liquidity of Corporate Bonds. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1106852

Barberis, N. S., Andrei; Wurgler, Jeffrey. (2002). Comovement. Retrieved from http://www.econ.yale.edu/~shiller/behfin/2002-04-11/barberis-etal.pdf

Bethke, S., Gehde-Trapp, M., & Kempf, A. (2017). Investor sentiment, flight-to-quality, and corporate bond comovement. Journal of Banking & Finance, 82, 112-132.

BlackRock. (2017). Global ETP Landscape Monthly Snapshot December 2017. Retrieved from https://www.blackrock.com/investing/literature/market-commentary/global-etp-landscape-monthly-snapshot-en-us.pdf

Bloomberg. (2018). Retrieved March 10 2018

Board of Governors of the Federal Reserve System (US). (2018). Nonfinancial Corporate Business; Corporate Bonds; Liability, Level [CBLBSNNCB]. Retrieved March 10 2018, from Federal Reserve Bank of St. Louis https://fred.stlouisfed.org/series/CBLBSNNCB

Bodurtha Jr., J. N., Dong-Son, K., & Lee, C. M. C. (1995). Closed-end Country Funds and U.S. Market Sentiment. The Review of Financial Studies, 8(3), 879-918.

Bojinov, S. (2015). A Brief History of ETFs. Retrieved from http://etfdb.com/etf-industry/a-brief-history-of-etfs/

Bolla, L., Kohler, A., & Wittig, H. (2016). Index-Linked Investing - A Curse for the Stability of Financial Markets around the Globe? Journal of Portfolio Management, 42(3), 26-43.

Brandon, K., Sung, S., & Podziemska, J. (2018). SIFMA US Quarterly Highlights 1Q'18. Accessible from SIFMA: https://www.sifma.org/wp-content/uploads/2018/04/US-Quarterly-Highlights-2018Q1-2018-04-06-SIFMA.pdf

Brandt, M. W., & Wang, K. Q. (2003). Time-Varying Risk Aversion and Unexpected Inflation. Journal of Monetary Economics, 50(7), 1457-1498.

Chambers, R., Kokic, P., & Cruddas, M. (2000). Winsorization for Identifying and Treating Outliers in Business Surveys. Retrieved from https://www.researchgate.net/publication/307632859_Winsorization_for_Identifying_and_Treating_Outliers_in_Business_Surveys

Chen, H. J., Chen, H. J., & Li, F. (2008). Firm-Level Return Comovement. Retrieved from http://people.ucalgary.ca/~nfa99/2008/papers/67.pdf

Clark, J., & Mauck, N. (2014). An Analysis of the Growth in the Fixed Income ETF Market. Journal of Stock & Forex Trading, 3(3).

Da, Z., & Shive, S. (2018). Exchange traded funds and asset return correlations. European Financial Management, 24(1), 136-168.

Dannhauser, C. (2017). The impact of innovation: Evidence from corporate bond exchange-traded funds (ETFs). Journal of Financial Economics, 125(3), 537-560.

Dezhbakhsh, H. L., Daniel. (1994). Periodic properties of interpolated time series. Economics letters, 44(1), 221-228.

Dick-Nielsen, J., Feldhütter, P., & Lando, D. (2012). Corporate bond liquidity before and after the onset of the subprime crisis. Journal of Financial Economics, 103(3), 471-492.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. Journal of American Statistical Association, 74(366), 427-431.

Duda, J. T., & Augustynek, A. (2005). A study of cross-correlation nonstationarity of world economy indices and energy prices. Retrieved from http://www.zarz.agh.edu.pl/augustyn/pliki/publikacje/Duda_Aug_WZ.pdf

ETF Database. (2018). Retrieved March 12 2018. From: http://etfdb.com/etfs/

ETF.com. (2018). What Kinds Of Bond ETFs Are There? Retrieved from http://www.etf.com/etf-education-center/what-kinds-bond-etfs-are-there?nopaging=1

FINRA. (2018). Understanding Bond Risk. Retrieved May 1.2018 from: http://www.finra.org/investors/understanding-bond-risk

Fitch Ratings. (2015). Bond Trade Frequency Strongly Linked to Issue Size. Retrieved from https://www.fitchratings.com/gws/en/fitchwire/fitchwirearticle/Bond-Trade-Frequency?pr_id=978847

Froot, K. A. D., Emil M. (1999). How are stock prices a!ected by the location of trade? Journal of Financial Economics, 53(1), 189-216.

FTSE Russel. (2017). The growth of passive investing: Has there been an impact on the US equity market? Retrieved from http://investmentnews.co.nz/wp-content/uploads/ftsepass.pdf

Gorman, L. R., Sapra, S. G., & Welgand, R. A. (2010). The role of cross-sectional dispersion in active portfolio management. Investment Management and Financial Innovations, 7(3), 58-68.

Gospodinov, N. (2017). Asset Co-movements: Features and Challenges. Retrieved from https://www.frbatlanta.org/-/media/documents/research/publications/wp/2017/11-asset-co-movements-features-and-challenges-2017-11-16.pdf

Hendershott, T., & Madhavan, A. (2015). Click or Call? Auction versus Search in the Over-the-Counter Market. The Journal of Finance, 70(1), 419-447.

Hsiao, C. (2007). Panel data analysis - advantages and challenges. TEST, 22(1), 1-22.

Janus Henderson Investors. (2017). A Simple Guide to Dispersion and Correlation. Retrieved March 15.2018 from http://az768132.vo.msecnd.net/documents/103808_2017_05_24_10_51_29_557.gzip.pdf

Jewell, J., & Livingston, M. (1999). A Comparison of Bond Ratings from Moody's S&P and Fitch IBCA. Financial Markets, Institutions & Instruments, 8(4), 1-45.

Kamara, A., Lou, X., & Sadka, R. (2008). The divergence of liquidity commonality in cross-section of stocks. Journal of Financial Economics, 89, 444-466.

Kiernan, K. (2015). Bond Liquidity — Factors to Consider and Questions to Ask. Retrieved from http://www.finra.org/investors/alerts/bond-liquidity-factors-questions

Kurov, A. (2010). Investor sentiment and the stock market's reaction to monetary policy. Journal of Banking & Finance, 34(1), 139-149.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics, 54, 159-178.

Lettau, M., & Madhavan, A. (2018). Exchange-Traded Funds 101 for Economists. Journal of Economic Perspectives, 32(1), 135-154.

Li, Y. (2018). Asset managers eye loan ETFs as interest rates rise. Retrieved from https://www.reuters.com/article/loans-etf-rates/asset-managers-eye-loan-etfs-as-interest-rates-rise-idUSL2N1RH1AL

Liebschutz, J., & Smith, B. (2016). Examining Corporate Bond Liquidity and Market Structure. Retrieved from https://www.treasury.gov/connect/blog/Pages/Examining-Corporate-Bond-Liquidity-and-Market-Structure.aspx

Lin, H., Wang, J., & Wu, C. (2011). Liquidity risk and expected corporate bond returns. Journal of Financial Economics, 99, 628-650.
Markit Group ltd. (2008). Markit Credit Indices, A Primer. Retrieved from https://www.markit.com/news/Credit%20Indices%20Primer.pdf

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

Martin, A. (2017). The Rise of Passive and Indexed Investing, and its Effect on Market and Liquidity Risk. Retrieved March 1.2018 from: https://www.advisorperspectives.com/articles/2017/02/27/the-rise-of-passive-and-indexed-investing-and-its-effect-on-market-and-liquidity-risk

Maverick, J. B. (2018). Top 3 Mortgage Backed Securities (MBS) ETFs (MBB, MBG). Retrieved from https://www.investopedia.com/articles/investing/081215/top-3-mortgage-backed-securities-mbs-etfs.asp

McCullough, M. (2017). A tantalizing fixed-income alternative. Retrieved from http://www.moneysense.ca/save/investing/etfs/fixed-income-alternative/

Millo, G. (2017). Robust Standard Error Estimators for Panel Models: A Unifying Approach. Journal of Statistical Software, 82(3), 1-27.

Nam, J. (2017). Market Accessibility, Corporate Bond ETFs, and Liquidity. Indiana University Kelley School of Business Resarch Papaper 18-1, 1-70.

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica, 55(3), 703-708.

Novick, B., Madhavan, A., Cohen, S., Samandar, S., Van Nugteren, S., & Rosenblum, A. (2017). A Primer on ETF Primary Trading and
the Role of Authorized Participants. Retrieved from https://www.blackrock.com/corporate/literature/whitepaper/viewpoint-etf-primary-trading-role-of-authorized-participants-march-2017.pdf

Nyaradi, J. (2013). The ‘Great Bond Massacre’ of 2013. Retrieved May 10.2018 from: https://www.marketwatch.com/story/the-great-bond-massacre-of-2013-2013-07-03

Pan, K., & Zeng, Y. (2017). ETF arbitrage under liquidity mismatch. Retrieved from http://www.esrb.europa.eu/pub/pdf/wp/esrb.wp59.en.pdf

Pindyck, R. S., & Rotemberg, J. J. (1990). The Excess Co-Movement of Commodity Prices. The Economic Journal, 100(403), 1173-1189.

PPM America. (2018). The Growth, and Impact, of Fixed Income ETFs. Retrieved from https://www.ppmamerica.com/assets/media/pdfs/PPM_Market-Spotlight_ETFs.pdf

Riedl, D. (2018). Size matters when it comes to ETFs. Retrieved from https://www.justetf.com/uk/news/etf/size-matters-when-it-comes-to-etfs.html

Schacht, K. N. (2016). Should Bonds Trade More Like Equities? CFA Insititute Magazine, 27, 51.

SEC. (2018a). CUSIP Number. Retrieved May 2.2018 from: https://www.sec.gov/answers/cusip.htm

SEC. (2018b). Securities Act Rule 144. Retrieved May 2.2018 from: https://www.sec.gov/fast-answers/answersrule144htm.html

SIFMA. (2018). US Bond Market Issuance and Outstanding. Retrieved from: https://www.sifma.org/resources/research/us-bond-market-issuance-and-outstanding/

Simpson, S. D. (2018). A Brief History Of Exchange-Traded Funds. Retrieved from Investopedia website: https://www.investopedia.com/articles/exchangetradedfunds/12/brief-history-exchange-traded-funds.asp

Smales, L. A. (2015). The role of political uncertainty in Australian financial markets. Accounting & Finance, 56(2), 545-575.

Sparks, A. (2018). MSCI: How easy is it to track a bond market index? Retrieved from https://www.etfstrategy.co.uk/msci-how-easy-is-it-to-track-a-bond-market-index-65487/

Stone, A. (2017). Index Changes Could Cause More Crowding into Biggest Bonds. Retrieved from https://www.barrons.com/articles/change-to-indices-could-cause-funds-to-crowd-into-biggest-bonds-1485289697

Sullivan, R. N., & Xiong, J. X. (2012). How Index Trading Increases Market Vulnerability. Financial Analysts Journal, 68(2), 70-84.

Sultan, S. G. (2015). A Study on Bond Exchange-Traded Funds (ETFs) and Corporate Bond Liquidity. (Doctor of philosophy), University of Washington. Retrieved from https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/33732/Sultan_washington_0250E_14919.pdf?sequence=1&isAllowed=y

Sushko, V., & Turner, G. (2018). The implications of passive investing for securities markets, 19. Retrieved from https://www.bis.org/publ/qtrpdf/r_qt1803j.htm

Tang, K., & Xiong, W. (2012). Index Investment and the Financialization of Commodities. Financial Analysts Journal, 68(6), 54-74.

Torres-Reyna, O. (2014). Cubic interpolation using R. Retrieved from http://www.princeton.edu/~otorres/InterpolationR101.pdf

Tuchman, M. (2013). What is an ETF? Three simple answers. Retrieved from https://www.forbes.com/sites/mitchelltuchman/2013/06/28/what-is-an-etf-three-simple-answers/2/#4f2caeac612b

Tucker, T. (2016). How ETFs Work. Retrieved from https://www.bloomberg.com/features/2016-etf-files/toy/

Tuckman, B., & Serrat, A. (2012). Fixed Income Securities: Tools for Today's Markets (Vol. 3). Hoboken, New Jersey: John Wiley & Sons, Inc.

U.S. Department of the Treasury. (2013). Treasury Inflation-Protected Securities (TIPS). Retrieved May 9.2018 from https://www.treasurydirect.gov/indiv/products/prod_tips_glance.htm

Vanguard Group. (2018). How are ETFs indexed? Retrieved May 8.2018 from: https://advisors.vanguard.com/VGApp/iip/site/advisor/etfcenter/article/ETF_HowETFIndexed

Wigglesworth, R. (2018). ETF growth is 'in danger of devouring capitalism'. Retrieved from https://www.ft.com/content/09cb4a5e-e4dc-11e7-a685-5634466a6915

Wooldridge, J. M. (2016). Introductory Econometrics, A Modern Approach. Boston, MA, USA: Cengage Learning.

WRDS. (2017). WRDS Corporate Bond Database: Data Overview and Construction Manual.

WRDS Corporate Bond Database Retrieved March 10, 2018, from Wharton Research Data Services

Wurgler, J. (2011). On the Economic Consequences of Index-Linked Investing Challenges to business in the Twenty first century (pp. 20-34). Cambridge, Massachusetts: American Academy of Arts and Sciences. Retrieved from https://www.amacad.org/pdfs/challenges.pdf.

Appendix

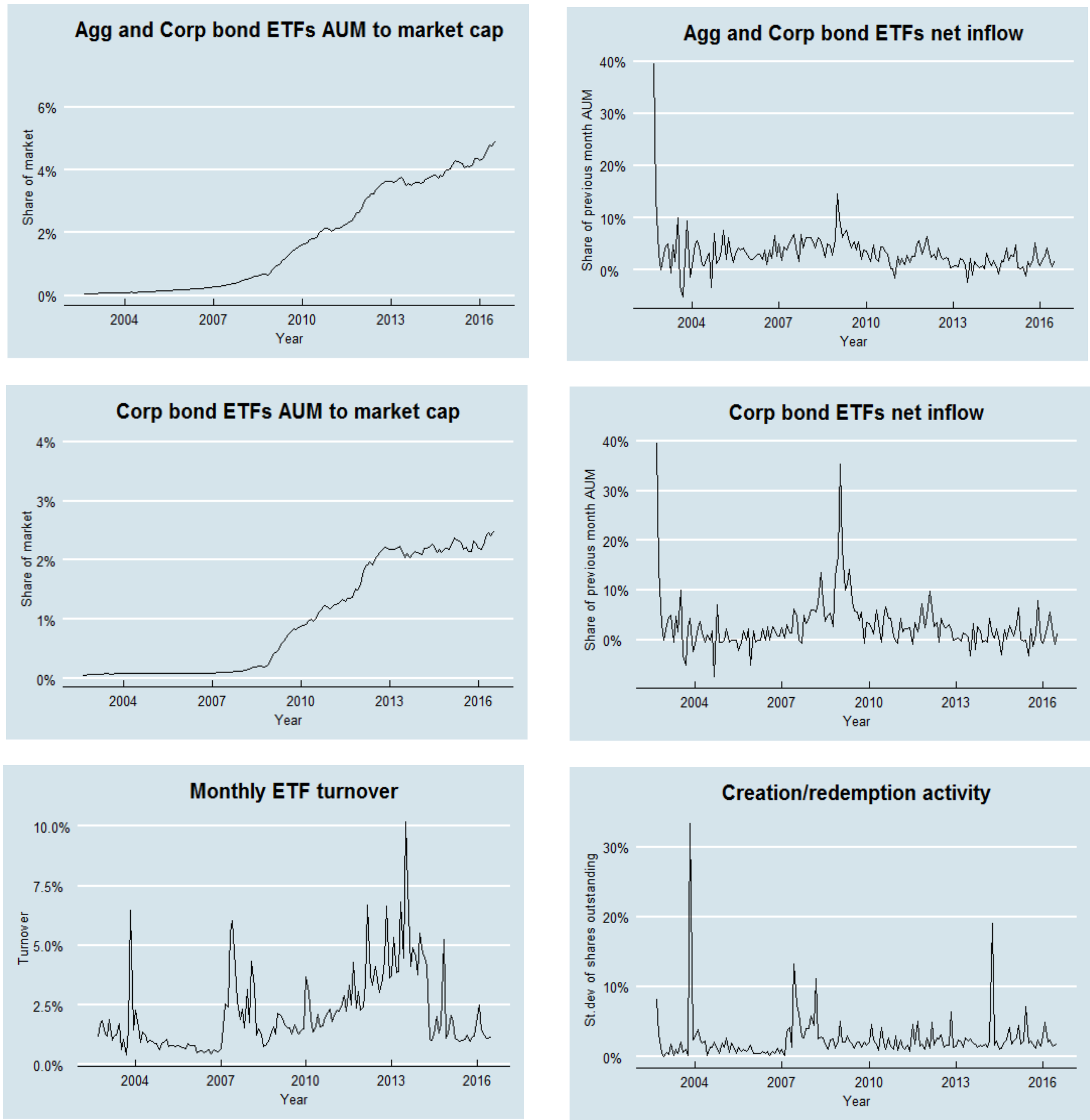
A. Descriptive statistics and variable plots

Development of fixed income ETFs

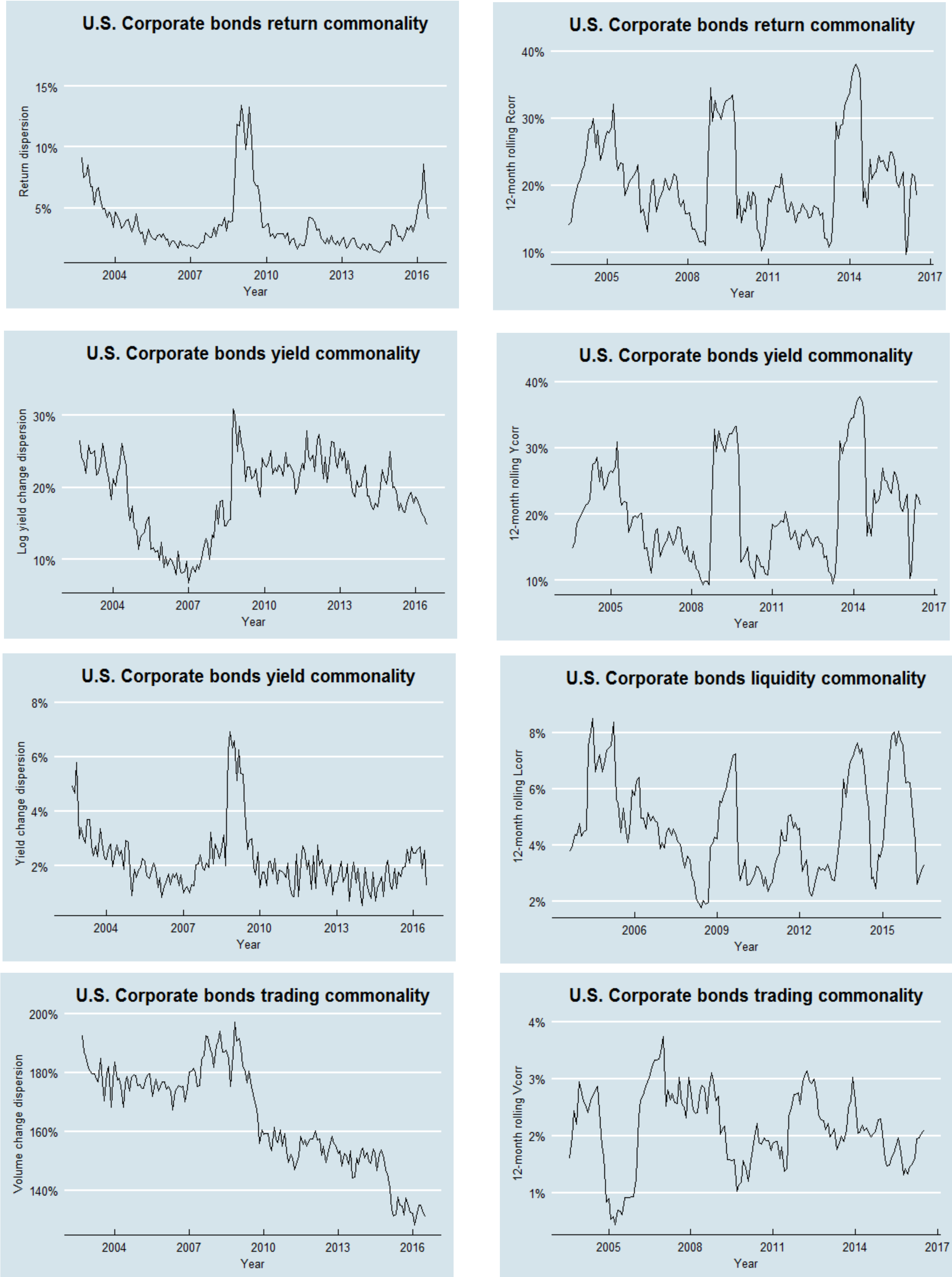
# FI ETF growth Y/Y	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
Aggregate	0 %	100 %	0 %	0 %	0 %	700 %	0 %	25 %	20 %	25 %	7 %	50 %	46 %	23 %	19 %	24 %
Bank Loans	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	100 %	100 %	100 %	0 %	25 %	-20 %	0 %
Convertible	0 %	0 %	0 %	0 %	0 %	0 %	0 %	100 %	0 %	0 %	0 %	0 %	0 %	100 %	50 %	0 %
Corporate	100 %	0 %	0 %	0 %	0 %	500 %	17 %	57 %	73 %	53 %	48 %	35 %	22 %	21 %	8 %	16 %
Government	100 %	0 %	0 %	0 %	0 %	300 %	42 %	59 %	63 %	14 %	12 %	7 %	0 %	7 %	2 %	0 %
Inflation Protected	0 %	100 %	0 %	0 %	0 %	100 %	50 %	100 %	17 %	71 %	0 %	8 %	0 %	23 %	-6 %	0 %
Mortgage Backed	0 %	0 %	0 %	0 %	0 %	100 %	0 %	200 %	0 %	0 %	67 %	0 %	20 %	17 %	0 %	0 %
Municipals	0 %	0 %	0 %	0 %	0 %	100 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
Preferred	0 %	0 %	0 %	0 %	0 %	100 %	50 %	33 %	0 %	0 %	75 %	14 %	13 %	0 %	11 %	30 %
Total		50 %	0 %	0 %	17 %	400 %	23 %	53 %	42 %	27 %	23 %	21 %	15 %	16 %	7 %	12 %

FI ETF AuM growth Y/Y	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
Aggregate	0 %	0 %	363 %	193 %	69 %	98 %	55 %	54 %	27 %	43 %	32 %	11 %	38 %	25 %	27 %	31 %
Bank Loans	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	688 %	375 %	-7 %	-19 %	88 %	19 %
Convertible	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	130 %	26 %	36 %	119 %	41 %	-11 %	25 %	41 %
Corporate	0 %	22 %	9 %	-4 %	13 %	44 %	188 %	150 %	40 %	43 %	49 %	2 %	11 %	9 %	28 %	22 %
Government	0 %	1 %	73 %	84 %	39 %	68 %	30 %	34 %	33 %	24 %	10 %	-9 %	22 %	16 %	12 %	42 %
Inflation Protected	0 %	0 %	963 %	119 %	21 %	34 %	72 %	118 %	8 %	21 %	8 %	-29 %	1 %	11 %	43 %	28 %
Mortgage Backed	0 %	0 %	0 %	0 %	0 %	0 %	369 %	110 %	27 %	85 %	63 %	-16 %	39 %	22 %	35 %	40 %
Municipals	0 %	0 %	0 %	0 %	0 %	0 %	305 %	169 %	23 %	21 %	43 %	-13 %	34 %	28 %	32 %	24 %
Preferred	0 %	0 %	0 %	0 %	0 %	631 %	730 %	223 %	73 %	6 %	53 %	-20 %	36 %	28 %	23 %	15 %
Total	0 %	20 %	82 %	77 %	37 %	70 %	71 %	86 %	30 %	33 %	32 %	-1 %	21 %	16 %	27 %	29 %

ETF activity and bond commonality measures



ETF activity measures calculated, source: CRSP, FRED Economic Data, Bloomberg (2018)



U.S. Corporate bonds calculated commonality measures, source: TRACE (2018)

Descriptive statistics: Investment grade sample

Statistic	N	Mean	St. Dev.	Min	Median	Max
Flows	167	0.027	0.049	-0.074	0.020	0.396
ETFturnover	167	0.027	0.030	0.004	0.014	0.181
SDshares	167	0.020	0.034	0.000	0.013	0.369
RetDisp	167	0.025	0.016	0.010	0.019	0.098
VolDisp	167	1.697	0.192	1.334	1.727	2.060
LogYDisp	167	0.197	0.062	0.061	0.212	0.341
YDisp	167	0.017	0.009	0.003	0.016	0.062
Vcorr	156	0.019	0.006	0.005	0.020	0.031
Rcorr	156	0.337	0.093	0.176	0.344	0.542
Ycorr	156	0.326	0.095	0.170	0.332	0.518
Lcorr	155	0.062	0.025	0.017	0.067	0.127

Descriptive statistics: High yield sample

Statistic	N	Mean	St. Dev.	Min	Median	Max
Flows	110	0.058	0.109	-0.217	0.029	0.565
ETFturnover	110	0.024	0.019	0.009	0.019	0.155
SDshares	110	0.035	0.058	0.005	0.019	0.505
RetDisp	110	0.054	0.041	0.018	0.038	0.208
VolDisp	110	1.453	0.237	0.882	1.518	1.871
LogYDisp	110	0.213	0.049	0.087	0.224	0.341
YDisp	110	0.016	0.009	0.003	0.015	0.062
Vcorr	110	0.334	0.087	0.137	0.330	0.595
Rcorr	110	0.036	0.011	0.017	0.033	0.072
Ycorr	110	0.329	0.098	0.117	0.326	0.576
Lcorr	110	0.063	0.025	0.016	0.065	0.128

Descriptive statistics: Fundamental factors

Statistic	N	Mean	St. Dev.	Min	Median	Max
VIX	167	0.016	0.226	-0.385	-0.017	1.346
UnexpInfl	167	-0.00001	0.003	-0.014	0.00003	0.011
IntRate	167	0.002	0.135	-0.311	-0.007	0.544
IGCDX	129	0.014	0.156	-0.335	-0.010	0.909
HYCDX	128	0.005	0.149	-0.593	-0.012	0.557

Descriptive statistics: Investment grade panel

Statistic	N	Mean	St. Dev.	Min	Median	Max
Bond-index correlation	107,049	0.242	0.679	-0.961	0.745	0.999
ETFshare	105,729	0.007	0.011	-0.0003	0.002	0.186
SDshares	167	0.007	0.010	0.000	0.005	0.082
ETFturnover	167	0.010	0.005	0.003	0.009	0.025
LQDflows	167	0.020	0.054	-0.107	0.015	0.396
SDshares%	105,729	0.00005	0.0001	0	0.00000	0.007
Turnover%	105,729	0.0001	0.0001	0	0.00003	0.003
Flows%	105,729	0.0001	0.001	-0.008	0.000	0.036
RATING_NUM	107,049	6.709	2.343	0	7	22
DURATION	107,049	6.701	4.467	0.000	5.680	19.410
T_Spread	107,049	0.006	0.006	0.000	0.004	0.190

Descriptive statistics: High yield panel

Statistic	N	Mean	St. Dev.	Min	Median	Max
Bond-index correlation	47,092	0.261	0.633	-0.957	0.701	0.999
ETFshare	46,455	0.012	0.013	0.000	0.014	0.125
SDshares	111	0.021	0.035	0	0.01	0.268
Turnover	111	0.030	0.021	0.008	0.022	0.142
HYGflows	111	0.071	0.182	-0.186	0.032	1.320
SDshares%	46,455	0.0001	0.0003	0.000	0.00004	0.012
Turnover%	46,455	0.0004	0.001	0.000	0.0003	0.007
Flows%	46,455	0.0003	0.002	-0.008	0.000	0.056
RATING_NUM	47,092	13.002	2.893	0	13	22
DURATION	47,092	4.292	1.635	0.000	4.430	10.400
T_Spread	47,092	0.007	0.016	0.000	0.005	1.651

B. WRDS data cleaning procedures

The original FINRA TRACE dataset contains raw data on all reported trades in the US OTC corporate bond markets. In order to perform academic research on the TRACE data it is necessary to clean the raw data. The WRDS bond database that we used in our research follows the data filtering and cleaning steps from (Asquith, Covert , & Pathak, 2013) and Dick-Nielsen (2009, 2014), which have been used in several other research papers. The cleaning process is explained in the WRDS Corporate Bond Database Manual (WRDS, 2017). The steps correct for the three most common trade report errors which are cancellations, corrections and reversals**.** Cancellations are instances when a previous same-day trading report has been registered without replacing an updated report. Corrections is an error where a previous same-day trading report replaces an updated report. Reversals occur when a previous trading report from a different date replaces the updated report. Agency reporting conventions can also result to some trades being counted twice, which is also corrected during the steps. The entire WRDS cleaning process is attached in appendix x.

The WRDS dataset is further cleaned by only including fixed and zero-coupon bonds (variable coupons are filtered out) that are not under rule 144a. Securities under rule 144a have to comply with a set of conditions such as holding period and cannot be sold to the public (only to institutional investors) (SEC, 2018b). All bond types that are not US Corporate convertibles (CCOV), US Corporate debentures (CDEB), US corporate medium-term notes (CMTN), US corporate term notes (CMTZ) or US corporate paper, are filtered out. Returns are then constructed by calculating the price change of each individual bonds “dirty price” which is the traded “clean price” plus the coupon interest that has accrued between coupon payment dates. The returns are further winsorized[9] at the 1% level to remove extreme observations. The monthly variables that we included for each bond before performing further cleaning are listed in appendix x.

[9] Winsorization: Statistical technique to deal with outliers by altering or removing the value of extreme outliers or altering their weight (Chambers, Kokic, & Cruddas, 2000)

C. Regression outputs: Naïve OLS

	Dependent variable:											
	Vcorr						Rcorr					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ETF%_all	-0.0489						-0.0268					
	(0.7650)						(0.9788)					
ETF%_corp		-0.0826						0.0460				
		(0.7676)						(0.9817)				
Flows_all			0.0251						0.0981			
			(0.4202)						(0.8436)			
Flows_corp				0.0257						0.1638		
				(0.1196)						(0.6099)		
ETFturnover					0.0647						-0.0855	
					(0.5445)						(0.9030)	
SDshares						0.0125						0.0296
						(0.6404)						(0.8599)
Constant	0.0222^{***}	0.0221^{***}	0.0205^{***}	0.0206^{***}	0.0198^{***}	0.0209^{***}	0.2110^{***}	0.2100^{***}	0.2076^{***}	0.2062^{***}	0.2124^{***}	0.2098^{***}
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	155	155	155	155	155	155	155	155	155	155	155	155
R^2	0.0132	0.0123	0.0084	0.0294	0.0251	0.0038	0.00004	0.00004	0.0014	0.0130	0.0005	0.0002
Adjusted R^2	0.0068	0.0058	0.0019	0.0230	0.0187	-0.0028	-0.0065	-0.0065	-0.0051	0.0065	-0.0061	-0.0063
Residual Std. Error	0.0068	0.0068	0.0068	0.0068	0.0068	0.0069	0.0658	0.0658	0.0658	0.0654	0.0658	0.0658
F Statistic	2.0474	1.9040	1.2904	4.6329^{**}	3.9352^{**}	0.5776	0.0066	0.0064	0.2126	2.0097	0.0731	0.0349

Note: We use Newey-West HAC standard errors. P-values are reported in the brackets.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

	Dependent variable:											
	Ycorr						Lcorr					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ETF%_all	0.5733						-0.0206					
	(0.6257)						(0.9749)					
ETF%_corp		1.0978						-0.0226				
		(0.6348)						(0.9839)				
Flows_all			-0.0698						-0.0733			
			(0.9043)						(0.4662)			
Flows_corp				0.1526						-0.0624		
				(0.6740)						(0.2448)		
ETFturnover					0.1200						-0.1574	
					(0.8836)						(0.4325)	
SDshares						0.0384						-0.0152
						(0.8304)						(0.6734)
Constant	0.1912***	0.1906***	0.2042***	0.1981***	0.1995***	0.2011***	0.0469***	0.0467***	0.0487***	0.0482***	0.0500***	0.0469***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	155	155	155	155	155	155	155	155	155	155	155	155
R^2	0.0168	0.0201	0.0006	0.0096	0.0008	0.0003	0.0004	0.0001	0.0111	0.0270	0.0232	0.0009
Adjusted R^2	0.0104	0.0137	-0.0059	0.0031	-0.0057	-0.0062	-0.0062	-0.0064	0.0046	0.0207	0.0168	-0.0057
Residual Std. Error	0.0708	0.0706	0.0713	0.0710	0.0713	0.0714	0.0174	0.0174	0.0173	0.0171	0.0172	0.0174
F Statistic	2.6149	3.1419*	0.0915	1.4788	0.1225	0.0499	0.0559	0.0220	1.7181	4.2495**	3.6343*	0.1317

Note: We use Newey-West HAC standard errors. P-values are reported in the brackets.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

D. Robustness tests

Time series regressions

Robustness test: Investment grade

	Dependent variable:						
	ΔVolDisp	ΔRetDisp	ΔYDisp	ΔLogYDisp	ΔRcorr	ΔYcorr	ΔLcorr
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ΔETFturnover	0.310**	0.033*	-0.097***	0.323***	-0.268	-0.124	-0.001
	(0.027)	(0.066)	(0.008)	(0.007)	(0.103)	(0.501)	(0.984)
ΔETFturnover, t-1	-0.353	-0.012	0.092**	-0.248***	-0.489**	-0.422**	0.041
	(0.150)	(0.510)	(0.034)	(0.004)	(0.028)	(0.011)	(0.284)
ΔETFturnover, t-2	0.150	-0.0002	-0.045	-0.054	0.742**	0.640**	-0.020
	(0.412)	(0.999)	(0.272)	(0.617)	(0.037)	(0.035)	(0.533)
ΔSDshares	0.069	-0.012**	0.025**	-0.013	0.031	-0.013	0.016
	(0.352)	(0.031)	(0.012)	(0.684)	(0.483)	(0.738)	(0.278)
ΔSDshares, t-1	-0.031	0.013*	-0.015*	0.062*	-0.008	0.022	-0.006
	(0.440)	(0.089)	(0.086)	(0.058)	(0.849)	(0.597)	(0.729)
ΔFlows	0.002	0.006	-0.031	-0.068	0.074	0.090	-0.059
	(0.992)	(0.733)	(0.438)	(0.464)	(0.619)	(0.552)	(0.116)
ΔFlows, t-1	-0.177	0.005	-0.032	-0.010	-0.084	-0.012	-0.012
	(0.312)	(0.727)	(0.176)	(0.884)	(0.547)	(0.911)	(0.676)
ΔFlows, t-2	0.126	-0.031***	0.019	0.002	0.210	0.132	0.039**
	(0.185)	(0.001)	(0.312)	(0.967)	(0.113)	(0.249)	(0.042)
ΔVolDisp, t-1	0.504***						
	(0.000)						
ΔRetDisp, t-1		0.536***					
		(0.000)					
ΔYDisp, t-1			0.532***				
			(0.000)				
ΔLogYDisp, t-1				0.480***			
				(0.000)			
ΔRcorr, t-11					0.089***		
					(0.004)		
ΔYcorr, t-11						0.079***	
						(0.006)	
ΔLcorr, t-11							0.047**
							(0.032)
VIX	0.016	0.003	0.001	0.018*	-0.004	-0.005	0.003
	(0.257)	(0.137)	(0.689)	(0.053)	(0.830)	(0.798)	(0.259)

UnexpInfl	-0.412	0.056	0.122	-0.327	-1.187	-1.065	-0.355
	(0.779)	(0.642)	(0.562)	(0.704)	(0.547)	(0.603)	(0.297)
IntRate	0.044*	0.004	0.003	0.043**	0.051**	0.044**	0.008*
	(0.053)	(0.365)	(0.543)	(0.015)	(0.017)	(0.043)	(0.057)
IGCDX	-0.031	-0.001	-0.005	0.005	-0.017	-0.034**	-0.002
	(0.120)	(0.603)	(0.368)	(0.560)	(0.344)	(0.040)	(0.576)
Observations	126	126	126	126	117	117	117
R^2	0.486	0.595	0.328	0.460	0.254	0.221	0.101
Adjusted R^2	0.427	0.548	0.251	0.397	0.161	0.124	-0.011
Residual Std. Error	0.036	0.005	0.008	0.022	0.042	0.041	0.009
F Statistic	8.222***	12.745***	4.247***	7.392***	2.726***	2.270**	0.900

Note: We use Newey-West HAC standard errors. All variables are on differenced form in order to obtain stationarity. Due to differencing, no constant is estimated. P-values are reported in the brackets.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Robustness test: High yield

	Dependent variable:						
	ΔVolDisp	ΔRetDisp	ΔYDisp	ΔLogYDisp	ΔVcorr	ΔYcorr	ΔLcorr
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ΔETFturnover	1.169**	0.143*	-0.014	0.295**	2.472***	2.525***	0.173
	(0.028)	(0.071)	(0.886)	(0.039)	(0.003)	(0.0003)	(0.189)
ΔETFturnover, t-1	-0.827	-0.258***	0.024	-0.553**	-0.487	-0.613	0.149
	(0.123)	(0.002)	(0.839)	(0.014)	(0.608)	(0.514)	(0.210)
ΔETFturnover, t-2	0.049	0.062	0.028	0.304***	0.261	0.344	-0.093
	(0.779)	(0.176)	(0.482)	(0.002)	(0.567)	(0.422)	(0.170)
ΔSDshares	-0.151*	0.040**	-0.034*	0.003	-0.395*	-0.429**	0.026
	(0.077)	(0.024)	(0.088)	(0.923)	(0.082)	(0.042)	(0.485)
ΔSDshares, t-1	0.229*	0.025	-0.002	0.039	0.035	0.079	-0.025
	(0.074)	(0.428)	(0.934)	(0.395)	(0.907)	(0.796)	(0.497)
ΔFlows	-0.123	0.009	-0.003	0.015	0.040	0.040	-0.003
	(0.121)	(0.336)	(0.684)	(0.489)	0.523	(0.480)	(0.724)
ΔFlows, t-1	0.044	-0.011	-0.001	-0.004	-0.062	-0.048	-0.001
	(0.484)	(0.324)	(0.988)	(0.847)	(0.373)	(0.445)	(0.946)
ΔFlows, t-2	-0.023	-0.009	0.008	-0.004	0.021	0.031	0.0001
	(0.566)	(0.274)	(0.321)	(0.828)	(0.693)	(0.548)	(0.990)
ΔVolDisp, t-1	0.487***						
	(0.000)						
ΔRetDisp, t-1		0.480***					
		(0.000)					
ΔYDisp, t-1			0.521***				

			(0.000)				
ΔLogYDisp, t-1				0.449^{***}			
				(0.000)			
ΔVcorr, t-11					0.110^{***}		
					(0.004)		
ΔYcorr, t-11						0.085^{**}	
						(0.030)	
ΔLcorr, t-11							0.082^{***}
							(0.0001)
VIX	-0.029	0.002	0.001	0.008	-0.005	-0.009	-0.011^{***}
	(0.369)	(0.419)	(0.788)	(0.500)	(0.793)	(0.653)	(0.008)
UnexpInfl	1.147	-0.421	0.149	-0.176	-5.308^{**}	-4.980^{**}	-0.178
	(0.594)	(0.168)	(0.569)	(0.872)	(0.040)	(0.037)	(0.641)
IntRate	-0.066	-0.004	0.004	0.036^{**}	-0.002	-0.016	-0.010
	(0.114)	(0.470)	(0.442)	(0.033)	(0.933)	(0.586)	(0.118)
HYCDX	0.043	0.0003	0.003	0.029	0.007	0.029	0.009
	(0.512)	(0.973)	(0.733)	(0.200)	(0.875)	(0.512)	(0.351)
Observations	107	107	107	107	98	98	98
R^2	0.466	0.641	0.315	0.505	0.265	0.258	0.215
Adjusted R^2	0.393	0.591	0.220	0.437	0.152	0.145	0.095
Residual Std. Error	0.058	0.010	0.009	0.023	0.055	0.055	0.010
F Statistic	6.321^{***}	12.887^{***}	3.327^{***}	7.382^{***}	2.352^{***}	2.274^{**}	1.789^{*}

Note: We use Newey-West HAC standard errors. All variables are on differenced form in order to obtain stationarity. Due to differencing, no constant is estimated. P-values are reported in the brackets.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Panel regressions

Robustness test: Model 2 investment grade

	Dependent variable:					
	Y = $Corr_{i,t}$					
	18m rw	18, rw	18m rw	24m rw	24m rw	24m rw
	(1)	(2)	(3)	(4)	(5)	(6)
Turnover%	90.1853***			95.0880***		
	(0.0000)			(0.0000)		
SDshares%		51.3202***			48.1454**	
		(0.0000)			(0.0178)	
Flows%			2.5098			3.2587
			(0.1984)			(0.2187)
Rating	-0.0112***	-0.0115***	-0.0117***	-0.0127***	-0.0131***	-0.0133***
	(0.0002)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
Duration	0.0329***	0.0338***	0.0340***	0.0313***	0.0320***	0.0321***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Spread	-1.0924***	-1.0806***	-1.0897***	-0.8855***	-0.8724***	-0.8725***
	(0.0000)	(0.0000)	(0.0000)	(0.0002)	(0.0002)	(0.0002)
Observations	93,186	93,186	93,186	81,683	81,683	81,683
R^2	0.0430	0.0414	0.0405	0.0529	0.0508	0.0501
Adjusted R^2	0.0206	0.0190	0.0181	0.0292	0.0270	0.0263
F Statistic	1,023.6020***	984.1700***	962.0159***	1,113.6040***	1,065.8700***	1,051.0030***

Note: p-values in brackets. Standard errors are clustered at the individual level.
*p<0.1, **p<0.05, ***p<0.01

Robustness test: Model 2 high yield

	Dependent variable:					
	Y = $Corr_{i,t}$					
	18m rw	18, rw	18m rw	24m rw	24m rw	24m rw
	(1)	(2)	(3)	(4)	(5)	(6)
Turnover%	48.9794***			55.8138***		
	(0.0000)			(0.0000)		
SDshares%		34.5515***			44.4080***	
		(0.0000)			(0.0000)	
Flows%			3.8695***			4.7099***
			(0.0002)			(0.0000)
Rating	0.0098***	0.0101***	0.0101***	0.0052**	0.0055**	0.0055**
	(0.0001)	(0.0001)	(0.0001)	(0.0246)	(0.0203)	(0.0212)
Duration	0.0299***	0.0333***	0.0337***	0.0282***	0.0317***	0.0323***
	(0.0000)	(0.0000)	(0.0000)	(0.0002)	(0.0000)	(0.0000)
Spread	-0.3557*	-0.3758*	-0.3841*	-0.2778*	-0.2952**	-0.3031**
	(0.0803)	(0.0675)	(0.0647)	(0.0547)	(0.0462)	(0.0434)
Observations	38,968	38,968	38,968	32,080	32,080	32,080
R^2	0.0220	0.0159	0.0141	0.0249	0.0143	0.0109
Adjusted R^2	-0.0130	-0.0194	-0.0212	-0.0139	-0.0249	-0.0284
F Statistic	211.5201***	151.7532***	134.5284***	196.6253***	111.5329***	85.1005***

Note: p-values in brackets. Standard errors are clustered at the individual level.
* p < 0.1, ** p< 0.05, *** p< 0.01