UNIVERSITY OF LIVERPOOL

Three Essays on Exchange Traded Funds

Author:

María del Carmen Marí Clérigues

Supervisors:

Prof Ólan T. Henry

Prof Chris Florackis

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the University of Liverpool Management School

March 2018

Abstract

Three Essays on Exchange Traded Funds

by María del Carmen Marí Clérigues

This thesis studies different aspects of the Exchange Traded Funds' tracking performance. Chapter 1 introduces the topic and provides the motivation for each chapter. Chapter 2 relates the tracking ability of an Exchange Traded Fund to the optimal hedge ratio for a portfolio that is long one unit of the fund financed by a short position in the benchmark index. The sample employed contains a range of funds listed on the New York Stock Exchange that track fixed income and equity indices. Panel cointegration techniques are employed to estimate the long-run relationship between the funds and their benchmarks. Furthermore, a Monte-Carlo experiment is performed to illustrate the inefficiency of the tracking error estimates when cointegration is not allowed for. Tracking errors based on return matching regressions, which neglect the long run relationship between the fund and the benchmark, may yield misleading estimates of tracking performance and sub-optimal portfolio choices for investors. Chapter 3 investigates the determinants of the cross-sectional differences in tracking errors. We argue that specific proxies should be used to account for the special structure of the ETF. We distinguish between primary liquidity, which relates to the ETF's creation and redemption processes, and secondary liquidity, which is linked to the trading activity of ETFs, and construct a series of proxies that might explain the differences observed. The sample employed coincides with that of Chapter 2 in terms of funds but, due to data availability for the liquidity proxies, the time horizon in Chapter 3 is slightly shorter. The results attribute the differences in the cross-section of tracking errors to differences in liquidity arising from the creation-redemption processes in the primary market for ETFs, and to differences in the liquidity in the secondary markets. Chapter 4 focuses on the time series dimension of ETFs' tracking errors. We evaluate the tracking ability of the most traded ETF in world, SPY ETF, using the hedge ratio approach from Chapter 2. The model suggests that time variation in the optimal hedge ratio arises from two sources of new information; news about the ETF and news about the benchmark. Additionally, we allow the variance-covariance matrix to vary with time and for asymmetry in the response to shocks. Consequently, the hedge ratio and the key measures of ETF's performance may also display time variation and asymmetry. The results suggest that as news about the fund and the benchmark arrive to the market, the elements of the variance-covariance matrix respond causing the hedge ratio to vary accordingly. In the presence of asymmetry in response to news, the hedge ratio might also exhibit asymmetric response unless the quality of the tracking is extremely good. Chapter 5 provides the conclusions and the directions for future research.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor Professor Ólan T. Henry for the continuous support provided throughout my PhD. He has been the person who has walked alongside with me during the last three and a half years, supporting me academically and emotionally through the journey to finish this thesis. I have found a teacher, a colleague and a friend. He has given me the freedom to work on my own way while ensuring that my research was on track. I really appreciate all his challenges, discussions and ideas that turned my PhD into a fruitful and stimulating experience and brought this thesis towards completion. I could not have wished for a better supervisor.

Besides my supervisor, I would like to thank my second supervisor Prof. Chris Florakis and the rest of the members of my panel: Prof. Brendan McCabe, Dr. Jason Laws, Dr. Abhinav Goyal, and Dr. Michalis Stamatogiannis and the colleagues at the University of Liverpool for their insightful comments and encouragement, but also for the intricate questions which have broaden the scope of my research from various perspectives. I would also like to thank the ESRC for funding my research.

I would like to thank my family and my friends, particularly Angelica, Bill, Emma, Jess, Juane, Moeed, Sushil, Tim, Ximo, and my “olas” for their endless support during the thesis. Specially, I would like to thank my Mum for always being there for me. Without her patience and constant encouragement this thesis would not have been completed. Finally, I would like to dedicate this thesis to my wonderful Dad who sadly passed away while my PhD was only a thought. Wherever you are, I am sure that you are very proud of me!

Contents

Chapter 3: The Impact of Liquidity on Exchange Traded Funds' Tracking Performance

Chapter 4: Modelling Time Variation in Tracking Errors: Evidence from the US market

List of tables

List of figures

Appendices

Chapter 1

Introduction and Motivation

An Exchange Traded Fund, hereafter ETF, represents a portfolio of securities chosen to track the performance of an underlying benchmark. Consequently, its value and risk should be directly linked to the benchmark constituents. ETFs are considered a hybrid between mutual funds and stocks since they behave similarly to mutual funds but trade like stocks in secondary markets. Consequently, their prices are determined by demand and supply in secondary markets.

Typically, ETFs are passive rather than actively managed. Passively managed ETFs are intended to mirror movements in their benchmark, while actively managed ETFs are designed to outperform their benchmarks. Although the vast majority of ETFs mirror broad market indices, ETFs these days can also track commodities, fixed income, real estate and currencies. ETFs can replicate their underlying benchmark physically, by buying the constituents of the index, or synthetically, by entering into swap and derivate agreements. ETFs can replicate the underlying index using full replication or optimised samples. While the former requires the purchase of all the constituents of the index in the same proportions that they are included in the index, the latter requires purchasing only some of the index constituents, which generally are most representative assets in the benchmark or, in case of debt ETFs, the most liquid ones. Replicating the benchmark using optimised samples, rather than full replication, may reduce the overall costs of the ETF. However, the main drawback of this replication strategy is that departures from the benchmark indices can be significant.

Passive investing has many advantages. Particularly, passive ETFs allow the investors to benefit from the risk diversification, intraday liquidity and low trading costs provided by these funds. Another advantage is the ETFs' transparency, which means that investors know the securities included in the ETF at every point in time.

ETFs are nowadays used by portfolio managers, hedge funds, pension funds and individual investors, amongst others, to fulfil their investment objectives. While fund managers and hedge funds might be more interested in pursuing short-term dynamic strategies, individual investors usually are buy-and-hold investors looking for an easy and convenient way to achieve portfolio diversification. The core objective of an ETF is to match the performance of its benchmark, before expenses, as closely as possible. In principal, the benchmark symbolises a portfolio which can be passively replicated at minimum cost. In practice however, the replication of the benchmark is costly and might give rise to the so-called Tracking Error, henceforward TE. TE is a key concept in index management because it represents the effectiveness of a manager in replicating the performance of the relevant benchmark.

This thesis is divided in five chapters. Chapter 1 introduces the topic and provides the motivation of the work. The second chapter, using sample that includes equity and fixed income ETFs listed on the New York Stock Exchange, shows that the tracking ability of an ETF is related to the optimal weighting for a hedge portfolio that is long one unit of the ETF financed by a short position in the benchmark. Panel cointegration techniques are employed to estimate the long-run relationship between ETFs and their benchmark indices. TE based on return matching regressions employed by the previous literature, rather than the error correction models proposed, may yield misleading estimates of performance and sub-optimal investment choices for investors. This finding is particularly relevant for portfolios containing equity ETFs. A Monte-Carlo experiment and an out-of-sample analysis illustrate the sub-optimality of decisions made using TE estimates which do not allow for cointegration explicitly.

Chapter 3 uses the methods developed in Chapter 2 to estimate the TE for a panel of U.S. equity and debt ETFs. Chapter 3 studies what drives the cross-sectional differences in the TE in a panel of ETFs listed on the NYSE. Due to the special structure that characterises ETFs, we distinguish between primary liquidity, which originates in the ETF’s primary market, and secondary liquidity, which is related to the market trading activity of ETFs. We also include a series of variables that have been employed in the previous literature and might also have an impact on the ETF’s tracking performance (such as the expense ratio charged by the fund, the replication strategy followed by the fund and the duration of the bond portfolio held by the ETF, among others). Overall, the results show that the illiquidity resulting from the creation-redemption processes plays a key role on determining the tracking quality of ETFs, regardless of the asset class tracked by the fund. This result illustrates the difficulties experienced by the authorised participants in the creation-redemption processes due to disparities between the ETF's price and the net asset value of the underlying securities. These disparities eventually deteriorate the ETFs' tracking performance. Our results show that the illiquidity created in the creation-redemption processes remains a key determinant of the funds' tracking performance, regardless of the asset classes included in the benchmark index. In terms of the secondary liquidity, the fund turnover is negatively related to the tracking ability of the fund, for both equity and debt ETFs. The outcomes also provide evidence that spreads in equity ETFs are positively related to the TE in a statistically significant fashion. Finally, the duration of the bond portfolio is positively but marginally related to the tracking performance of the ETF. In terms of indirect measures of liquidity, none of the measures included in the analysis appear to be statistically significant at any level.

The sample periods used in Chapter 2 and Chapter 3 do not coincide exactly. This is because of a lack of availability of data for the proxies employed to gauge the effect of liquidity on the tracking performance. The results from Chapter 2 are robust to this change in the sample period. The results on cointegration are also robust to a change of sample, as we find similar outcomes for a sample of equity ETFs listed on the London Stock Exchange. We present these results in the appendix of Chapter 2.

In Chapter 4, we focus exclusively in one ETF included in the previous samples. This ETF, known as the SPDR S&P 500 ETF, or SPY ETF, constitutes the first ETF listed in the U.S. This fund has become a global leader in the indexing investment and played a significant role in the later developments of ETF's markets worldwide. Since the SPY ETF was listed in 1993, we could gather 24 years of daily data for the analysis. In this chapter, we evaluate the ETF's tracking performance using the hedge ratio approach outlined in Chapter 2. We model jointly the first and the second conditional moments of the distribution of returns for the fund and the benchmark. Since we let the variance-covariance matrix vary with time, the resulting optimal hedge ratio and ETF's tracking performance, measured by the Tracking Difference (TD) and TE, may also display time variation. The evidence indicates that ETFs and the benchmarks appear I (1) and cointegrated. Furthermore, a series of tests performed suggest that the hedge ratio might display asymmetry in response to news of the fund and the benchmark index. Given the results obtained, we employ a multivariate asymmetric VECM-GARCH (1,1) structure for the first and the second conditional moments of the return distributions. As a result, the hedge ratio updates according to the arrival of news about the ETF and news about the benchmark to the market. Overall, the results show that as new information arrives to the market the volatility of the ETF and the index change and therefore, the conditional hedge ratio varies accordingly. Moreover, in the presence of asymmetry there is the possibility that the hedge ratio will display asymmetric response to news unless any variance and covariance asymmetry observed in the data is offsetting. Chapter 5 the conclusions and the directions of future research.

Given the recent growth in passively managed index funds, this thesis might have implications for institutional and retail investors, who hold ETFs to benefit from the risk diversification, transparency, intraday liquidity and low trading costs provided by these funds.

Chapter 2

Tracking and Tracking Errors

2. 1. Introduction

An ETF constitutes a basket of securities designed to track the performance of a benchmark index. ETFs behave similarly to mutual funds but trade like stocks and offer investors an easy and convenient way to achieve portfolio diversification at a very low cost. In the 24 years since the introduction of the first ETF, these funds have become one of the fastest growing segments in the market. The ICI[1] monthly report shows that the assets under management held by the ETFs listed in the U.S. amount to $2,524 billion, as at December 2016. This represents an increase of 20.17% in assets under management with respect to the corresponding figure as at December 2015.

[1] Retrieved from https://www.ici.org/research/stats/etf/etfs_12_16. Last access on the 9th February 2016.

Generally, ETFs are passively managed and intended to mirror movements in their benchmarks. ETFs can mirror equity indices, commodities, debt indices, real estate and currencies.

Passive ETFs offer investors a tax efficient way to achieve portfolio diversification at a very low cost. The core objective of these funds is to match the performance of its benchmark, before expenses, as closely as possible. Implicitly, if the ETF matches the performance of the benchmark at all times, then any TE will be minimised. In theory, the benchmark portfolio can be passively replicated easily and without costs for each ETF. However, in practice, the replication of the benchmark is costly and far from easy giving rise to the potential for TEs. The TE is a central concept in index management because it measures the accuracy of the replication strategy pursued by the fund. The TE and the total expense ratio charged by the fund appear to be the main criteria used by investors and money managers when comparing competing ETFs. Given the importance of ETF's tracking performance for investors, this chapter focuses on the accurate calculation of the TE. While our focus is on the performance of a sample of ETFs, the methodology employed is valid for any index tracking product.

Essentially, this chapter makes two contributions. Firstly, given the importance of the TE to evaluate the performance of an ETF, we propose an alternative framework, based on a simple model of optimising behaviour, to compute the TE which also takes into account the stochastic nature of the data. We demonstrate that our approach yields better estimates and more reliable inference than those approaches currently employed in the literature. Secondly, since stocks and bonds seem to be the two main asset classes included in investors' portfolios, and the research studying the tracking efficiency of non-equity ETFs is still very limited, we include a mix of equity and debt funds in our sample.

This chapter estimates for the first time the TE employing panel data techniques which take careful account of the stochastic properties of the data, a consideration that has been almost completely overlooked in the existing literature. We argue that the tracking ability of an ETF is related to the optimal hedge ratio for a portfolio that is long one unit of the fund financed by a short position in the benchmark index. Since the ETF is designed to track the level, and hence the return, of the benchmark over time, the ETF and its benchmark should share a common stochastic trend. Two non-stationary variables, such as the level of the ETF and benchmark, which share a common stochastic trend are said to be cointegrated (Engle and Granger 1987). In the presence of a cointegrating relationship between the levels of the fund and those of the benchmark, an Error Correction Model, hereafter ECM, should be employed to compute the TE. Most of the existing literature obtains the TE by using a return matching regression, which basically regresses the returns of the ETF on those of the benchmark index (see Elton, Gruber, Comer and Li 2002, Rompotis 2009; Buetow and Henderson 2012; Drenovak, Urošević and Jelic 2014; Bertone, Paeglis and Ravi 2015, among others). In the presence of a cointegrating relationship, the return matching regression is misspecified and is likely to lead to inefficient TE estimates, and ultimately to misleading inference. Therefore, the method we propose is very useful for investors, fund managers and finance practitioners since it computes the TE appropriately and hence minimises the possibility of suboptimal investment decisions. Although we focus on the performance of a sample of U.S. ETFs, the methodology we propose is valid for any index tracking product[2].

Despite the wide range of the ETFs available in the market, such as fixed income, currency or real estate ETFs, most of the existing literature focuses on the TE of equity ETFs (see Frino and Gallagher 2001; Gastineau 2004; Chu 2011; Rompotis 2011; Bertone, Paeglis and Ravi 2015; among others). To date, the literature on TE for ETFs mirroring asset classes other than equity is still scarce (Houweling 2011; Buetow and Henderson 2012; Drenovak, Urošević, and Jelic 2014). The availability of a wide variety of ETFs in the market makes the selection process relatively complex for investors. Despite this, equity and fixed income seem to be the asset classes that most investors include in their portfolios. Taking this into account, together with fact that the existing literature on the tracking performance of debt ETFs is still limited, our sample is made up of passively managed ETFs listed on the NYSE that replicate the performance of various domestic equity and fixed income indices. Our main aim is to demonstrate that, regardless of the asset class that the fund tracks, the TE estimates obtained from our method are more efficient than those obtained by the previous literature as they account for the long run relationship between the fund and the benchmark. We demonstrate that our approach is both statistically and economically superior to the outcomes computed using a return matching regression. Finally, we provide two simple illustrations of the implications of employing return matching regressions to gauge the TE. The first is based on a small-scale Monte-Carlo experiment, the outcomes of which suggest that TE based

[2] We have applied the methodology to a sample of ETFs listed on the London Stock Exchange tracking the major UK stock indices. The results confirmed the superiority of our approach and are available in the Appendix (Tables A.2.3, A.2.4, A.2.5, A.2.6 and A.2.7)

on return matching regressions, rather than an ECM, yield misleading estimates of ETF's performance. The second is based on a very simple out-of-sample exercise which ranks ETFs based on the TE and construct portfolios based on those rankings. The results highlight the potential for inference based on the return matching regression to lead to sub-optimal decision making.

Our empirical results suggest that equity ETFs track their benchmarks more closely than comparable debt funds. Moreover, irrespective of the asset class underlying the ETF, the measures based on return matching regressions overstate the magnitude of the TEs. This overestimation appears most pronounced for those ETFs which track their benchmarks most accurately. The TEs obtained using the ECM are statistically significantly different from those based on return matching regressions. As a result, any investment decision based on TEs computed from misspecified models might result in suboptimal investment choices for investors who want to gain exposure to the U.S., particularly using equity ETFs.

The next section of this chapter presents the literature survey. The third section contains our approach to modelling tracking. The fourth section provides a description of our data and some preliminary tests. The fifth section evaluates the ETF's tracking process. The sixth section describes the different methods used to compute the TE and compares the TEs obtained from a return matching regression with those computed with the ECM. The seventh section presents a small-scale Monte-Carlo experiment designed to illustrate the generality of our results. The penultimate section illustrates the economic impact of incorrectly estimating the TEs for portfolio selection in an out-of-sample exercise. The final section provides a summary of the Chapter and some concluding comments.

2.2. Literature review

In the 24 years since the introduction of the first ETF, this asset class has become one of the fastest growing segments available in the market. The trading of ETFs relies on the creation and redemption processes and that results in high market efficiency (Deville 2008). Any difference that might arise between fund's Net Asset Value (NAV) and the ETF's price would present an arbitrage opportunity. The majority of the existing research concludes that the market price and NAV are very closely related in domestic ETFs, which is evidence in favour of the efficient pricing of ETFs, see, among others, Ackert and Tian (2000), Elton, Gruber, Comer and Li (2002), Curcio, Lipka and Thornton (2004), Engle and Sarkar (2006), Ackert and Tian (2008) and Petajisto (2017) for the USA, Kayali (2007) for Turkish market, and Gallagher and Segara (2006) for the Australian market. Conversely, significant mispricing has been evidenced in country ETFs trading in the US market. These ETFs, also called international ETFs, trade in one market but mirror equity indices in a foreign market. Jares and Lavin (2004), using a sample of Japanese and Hong-Kong ETFs traded in the U.S., find that international ETFs exhibit continuous deviations between the prices and NAVs. Ackert and Tian (2008) and Petajisto (2017) report comparable findings. The reason underlying the deviations observed in Jares and Lavin (2004) seems to be the non-synchronous trading hours between the ETF and the underlying stocks together with the information dissemination in the markets. As a result, exploitable inefficiencies arise. In this case, the non-overlapping trading hours also seem to play a role in the mispricing of country funds. However, market illiquidity, momentum and size effects seem to be the key drivers of the mispricing evidenced in these funds. Analogously, Aber Li and Can (2009), using a three domestic ETFs and one international ETF, link the greater premiums exhibit by the international ETFs to the asynchronous trading. In line with the previous authors, Levy and Lieberman (2013), using a sample of 20 country ETFs issued by iShares, argue that during the synchronized trading hours the prices trade very close to their NAV due to the arbitrage mechanisms. However, during the non-overlapping hours the S&P 500 seems to drive the ETFs' prices. On the other hand, Engle and Sarkar (2006) link the large and persistent mispricing of country funds to the slow response to economic news, high costs of creation and redemption of ETF's shares and the thin trading that characterise their home markets. In summary, while previous research report that domestic ETFs trade very close to their NAVs, research in county ETFs report significant differences between ETF's prices and NAVs. The main reasons underlying the mispricing evidenced seem to be the asynchronous trading hours and the market inefficiencies in the foreign markets.

The high demand for diversification of risk incentivises financial intermediaries to offer different types of vehicles to achieve both domestic and international diversification. An ETF is one of those vehicles. ETFs compete with other financial securities such as index futures, the securities included in the benchmark, options contracts and comparable products as conventional Index Mutual Funds (IMF) or closed-end funds (Deville 2008). ETFs constitute a hybrid between common stocks and traditional IMFs. Indeed, prior to the advent of ETFs, investors had to choose between the versatility of direct stock investment or the ease and diversification that characterise IMFs. Pioneer research on ETFs has focused mainly on describing the main features of these financial innovations (see Gastineau 2001; Deville 2008; Gastineau 2010) and their characteristics and performance relative to conventional IMFs (Poterba and Shoven 2002; Kostovetsky 2003; Gastineau 2004; Agapova 2011) and close-end funds (Harper, Madura and Schnusenberg 2006).

Since ETFs behave identically to mutual funds but trade intraday like stocks, they seem to provide an alternative to IMFs. In conventional IMFs, any order placed after the end of the trading day would have to be purchased or sold at the following day's closing NAV (Broms and Gastineau 2007). Conversely, ETF's shares can be sold in the market anytime. Using aggregate flows, Agapova (2011) claims that ETFs and IMFs are not perfect substitutes, and this is the reason underlying their coexistence in the market. Deville (2008) concludes that ETFs seem to have filled a gap in investors' needs since IMFs are not always available to all investors while ETFs are, coupled with the ability of ETFs to expand investor's allocation opportunities by investing in specific sectors, or even in markets, where IMFs do not exist.

One of the most important advantage of ETFs is their tax efficiency. The effect of tax on the returns of ETFs and conventional IMFs is investigated in Poterba and Shoven (2002). Tax-efficiency appears to be of the main advantages of the ETFs due to the in-kind redemption and creation processes that characterises them. However, when compared with IMFs, the authors find that pre-tax and after-tax returns are analogous. The reason underlying this similarity seem to be related to the low capital gains distributed by the Vanguard Index 500. However, the authors argue that this result probably does not hold for other mutual funds. Therefore, more evidence would be needed to corroborate the ETFs' tax efficiency. Other advantages of ETFs, compared to IMFs, could be their high level of liquidity and their trading features, which include the ability to sell ETFs short or at a margin (Miffre 2007).

An additional feature of ETFs is their low management expenses. The reason underlying the lower costs of ETFs, compared to IMFs, is that ETFs are typically passively managed funds which are designed to closely track the performance of a benchmark over time. As a result, ETFs only rebalance their portfolios when the benchmark does it, and therefore the fund expenses reduce considerably. In contrast, actively managed IMFs, which constantly rebalance their portfolios as they try to beat the benchmark, result in greater costs and consequently increase the management expenses charged by the ETF. Moreover, as ETFs trade in stocks markets, the purchase or sale of shares requires investors to pay a brokerage commission. Nevertheless, these commissions are usually lower than the costs that IMFs has to bear, such as the costs of bookkeeping and shareholder services, among others. Poterba and Shoven (2002) compared an ETF and an IMF that track the S&P 500 and conclude that, although the differences are of small size, the evidence suggests that IMFs seem to be more cost efficient than ETFs. Nevertheless, the authors argue that their conclusions are drawn from the comparison of one IMF and one ETF, therefore their findings cannot be extrapolated to the whole universe of ETFs and IMFs.

Index trackers, such IMFs and ETFs, aim to replicate the performance of their benchmarks as closely as possible. While this seems simple in theory, in practise the exact replication of a benchmark is very difficult, if not impossible. The replication process faces unavoidable market frictions, which are not present in the construction of the underlying index itself, that might give rise to TEs. Any divergence between the value of the fund and that of the benchmark is captured by the TE. Pope and Yadav (1994) argue that TEs can be used either as a benchmark to measure performance or as a constraint to portfolio rebalancing decisions. The recent growth of passively managed index funds motivates the use of TEs to evaluate their tracking ability. The methodology used to gauge the TEs for index trackers is analogous to the methods previously used to evaluate the performance of active managers. For instance, Frino and Gallagher (2001) compares the TEs of a sample of ETFs and

IMFs tracking the S&P 500 index using the three measures coined by Roll (1992), Pope and Yadav (1994) and Rudolf, Wolter and Zimmermann (1999). These three measures have been extensively used to evaluate the tracking performance of ETFs listed in the U.S. (see Rompotis 2009; Buetow and Henderson 2012; Bertone, Paeglis and Ravi 2015; among others). Elton, Gruber, Comer and Li (2002) employ a similar approach using the NAV net of dividends and expenses, instead of ETF's closing prices, to calculate the TEs of SPY ETF which mirrors the performance of the S&P 500 index. Similarly, Bertone, Paeglis and Ravi (2015) compute the TE of DIA, an ETF that tracks the performance of the Dow Jones Industrial Average, using data sampled at an intraday frequency. Although most of the literature on ETFs' TE has followed the methods developed to measure the performance of active managers, alternative measures to compute the TEs have also emerged. For instance, Johnson (2009) employs correlations between ETFs and the underlying benchmarks to gauge tracking performance, while Drenovak, Urošević and Jelic (2014) investigates the tracking accuracy of European bond ETFs using a range of measures including a vector error correction approach.

Despite the increasing importance of these financial innovations, research on ETFs' TE is still scarce. Moreover, the vast majority of studies largely neglect the time series properties of the data when computing the TEs and hence, TEs estimates are obtained from pairwise relationships which ignore the potential influence of the rest of the market on prices. In this chapter, we employ a panel cointegration and error correction approach to estimate the TEs of a sample of passively managed equity and debt ETFs which adequately accounts for the stochastic properties of the data and provides efficient estimates of the TE.

2.3. Modelling tracking

Let F_t and I_t be the prices of the ETF and benchmark index, respectively, with logarithms f_t and i_t, respectively. The actual return to holding the index for one period is computed as $\Delta I_t = \log(i_t / i_{t-1})$, similarly the actual return on the ETF is $\Delta F_t = \log(f_t / f_{t-1})$. Then, the expected return, *Et-1(Rt)*, for any period t of a portfolio that is long one unit of the ETF funded by a short position in $\lambda_{t/t-1}$ units of the index may be written as:

$$E_{t-1}(R_t) = E_{t-1}(\Delta F_t) - \lambda_{t|t-1} E_{t-1}\left(\Delta I_t\right) \tag{2.1}$$

Note here that the portfolio weight, $\lambda_{t/t-1}$, is determined using information up to and including time t-1 and is used to construct the portfolio for period t. If $\lambda_{t/t-1} = 1$ then any gain in Eq. (2.1) from an increase in the long position will be exactly offset by the loss in the short position. In other words, a one-unit movement in the underlying index is matched by a $-1/\lambda_{t/t-1}$ unit movement in the ETF. The variance, $h_{R,t}$, of the portfolio may be written as:

$$h_{R,t} = Var_t\left(E_{t-1}\left(\Delta F_t\right) - \lambda_{t|t-1} E_{t-1}\left(\Delta I_t\right)\right) \tag{2.2}$$

The square root of Eq. (2.2) evaluated at the unconditional variances of ΔF_t and ΔI_t for $\lambda_{t/t-1} = 1$ is typically defined as the TE in the prospectus for most ETF's offerings[3]. Expanding Eq. (2.2) yields:

[3]See the definition of the tracking error in the prospectus for Source Markets PLC. https://www.sourceetf.com/sites/default/files/documents/SOURCE_MARKETS_PLC_PROSPECTUS_EN.pdf for further details. Last accessed on 9th of February 2016.

$$h_{R,t} = h_{F,t} + \lambda_{t|t-1}^2 h_{I,t} - 2\lambda_{t|t-1} h_{FI,t} \tag{2.3}$$

Where $h_{R,t}$, $h_{F,t}$ and $h_{I,t}$ represent the conditional variances of the portfolio, the ETF and the benchmark, respectively. $h_{FI,t}$ represents the conditional covariance between the fund and index returns. If an agent has the two-moment utility function in:

$$U(E_t R_t h_{R,t}) = E_t(R_t) - \psi h_{R,t} \tag{2.4}$$

Then the risk averse utility maximising agent with degree of risk aversion ψ seeks to solve:

$$\max U(E_t, R_t, h_{R,t}) = \Delta F_t - \lambda_{t|t-1}\Delta I_t - \psi(h_{F,t} + \lambda_{t|t-1}^2 h_{I,t} - 2\lambda_{t|t-1} h_{FI,t}) \tag{2.5}$$

Solving Eq. (2.5) with respect to $\lambda_{t|t-1}$, under the assumption that ΔF_t and ΔI_t are martingale processes such that $E_t(\Delta F_t) = \varepsilon_{F,t}$ and $E_t(\Delta I_t) = \varepsilon_{I,t}$, yields $\lambda_{t|t-1}$, which is the optimal number of units of the ETF in the investor's portfolio:

$$\lambda_{t|t-1}^* = -\frac{h_{FI,t}}{h_{I,t}} \tag{2.6}$$

It is clear that $\lambda_{t|t-1}^*$, in this context, may be interpreted as a hedge ratio. That is, the utility maximising investor will choose to form a riskless portfolio by selling $\lambda_{t|t-1}^*$ units of the index and using the funds to purchase a single unit of the ETF. In most situations an estimate of $\lambda_{t|t-1}^*$, the constant optimal hedge ratio, may be obtained from the estimated slope coefficient b in the following regression:

$$\Delta F_{m,t} = a + b\Delta I_{n,t} + \epsilon_{mn,t} \tag{2.7}$$

Where m represents the m^{th} fund and n the n^{th} benchmark index.

Much of the existing research on ETFs and mutual funds' pricing adopts the methodology of Eq. (2.7), which we refer to as a return matching regression (see Frino and Gallagher 2001; Buetow and Henderson 2012; and Bertone, Paeglis and Ravi 2015; among others). By definition, an ETF is an index tracking security and hence, we may characterise index tracking using Eq. (2.8):

$$f_{m,t} = \varphi_0 + \varphi_1 i_{n,t} + u_{mn,t} \tag{2.8}$$

Where f_t and i_t represent the logarithms of the closing price on trading day t for the m^{th} ETF and the n^{th} underlying index, respectively. We define strong tracking as the situation where f_t and i_t are cointegrated with cointegrating coefficient equal to unity. Alternatively, an ETF is said to display relative or weak tracking when f_t and i_t are cointegrated with less than unit cointegrating coefficient. A third alternative is that of returns matching, which occurs when f_t and i_t are not cointegrated but the coefficient b in Eq. (2.7) is positive and statistically significant. Clearly, where the hypothesis H_o: $a = 0$, $b = 1$ in Eq. (2.7) is satisfied for the data, the expected returns to the two investments coincide and, the ETF is said to display strong return matching.

We may test whether the ETF displays strong tracking in Eq. (2.8) using:

$$H_0 : \varphi_1 = 1 \tag{2.9}$$

In general, regressions such as Eq. (2.8) will not lead to reliable inferences due to the presence of unit roots in the levels of the variables and so Eq. (2.7) should be used to avoid the problems associated with spurious regressions. However, an exception to this occurs when the index and ETF share a common stochastic trend, which generally occurs with index tracking strategies. In this case, the ETF and the benchmark are said to be cointegrated, and appropriate methods of estimation and inference for Eq. (2.8) are readily available. Under the Engle and Granger (1987), cointegration implies, and is implied by an ECM:

$$\Delta F_{m,t} = \kappa_0 + \kappa_1 \Delta I_{n,t} + \gamma_1 u_{mn,t-1} + v_{mn,t} \tag{2.10}$$

Where $u_{mn,t-1}$ is the equilibrium error from Eq. (2.8), γ_1 measures the rate of return to equilibrium and $v_{mn,t}$ is a white noise process. Unless there exists a cointegrating relationship between f_t and i_t, any inference based upon Eq. (2.8) is likely to be unreliable. Comparison of Eq. (2.7) and Eq. (2.10) reveals:

$$\in_{mn,t} = \gamma u_{mn,t-1} + v_{mn,t} \tag{2.11}$$

It follows that Eq. (2.7) will tend to provide inefficient estimates when f_t and i_t share a common stochastic trend (Kroner and Sultan 1993). Then, we have that:

$$\sigma_{\in}^2 = \gamma^2 \sigma_u^2 + \sigma_v^2 \tag{2.12}$$

The covariance term in Eq. (2.12) is suppressed as the residuals from the correctly specified ECM (2.10) will be uncorrelated with themselves or any other variable.

2.4. Data description and preliminary tests

Our sample contains a collection of passively managed ETFs listed on the New York Stock Exchange, hereafter NYSE, which track various equity and fixed income indices. The initial sample begins with the universe of ETFs listed on the NYSE. The sample is restricted to equity ETFs that track the major U.S. equity and fixed income indices. To ensure that our sample represents different liquidity categories, we select ETFs that track indices containing large-caps, mid-caps, small-caps, mixed-caps and micro-caps. Likewise, we include fixed income ETFs that track government, municipal and corporate bonds. We do not include those ETFs with missing data nor funds listed after the 14th of December 2009. Finally, we omit the leveraged and inverse ETFs since these funds are very risky and generally are used to pursue short-term strategies. The resulting sample contains 77 ETFs of which 59 track equity indices and 18 track debt indices. The name of the ETF, the Bloomberg ticker and the inception date are displayed in the Appendix (Table A.2.1 for equity ETFs and Table A.2.2 for debt ETFs). The mix of equity and debt ETFs in our sample allows us to compare the performance of ETFs tracking different asset classes.

Since the price of the ETF encapsulates in one figure the supply and demand of the ETF together with the NAV (Buetow and Henderson 2012), closing prices rather than NAVs are employed. Daily closing prices were collected from Bloomberg for the period 14th December 2009 to 09th December 2016. The closing prices from 14th December 2009 to 09th September 2016 are used for estimation of the TEs. The remaining data, spanning the period from 12th September 2016 to 12th December 2016, is used to perform a very simple out-of-sample raking exercise.

[Figure 2.1]

[Figure 2.2]

The graphs in Figure 2.1 illustrate the tracking performance of an equity ETF from our sample (Bloomberg Ticker: DIA) and its benchmark (Bloomberg Ticker: INDU).

Figure 2.2 displays the performance a debt ETF from our sample (Bloomberg Ticker: CSJ) and its underlying index (Bloomberg Ticker: LD01TRUU). These graphs have three common features; firstly, the log prices of the funds and underlying indices appear non-stationary, secondly the log prices of the funds and benchmarks appear to move together over time and finally, it is clear from the graphs that the replication process is far from perfect and, consequently, TEs arise. Further inspection of the graphs suggests that the deviations from the benchmark appear more pronounced for debt ETFs.

[Table 2.1]

A summary of the time series properties of the log prices of the ETFs and their underlying indices are displayed in Table 2.1. Panel A of Table 2.1 presents a series of panel unit root test that include Maddala and Wu (1999), Breitung (2001), and Im, Pesaran and Shin (2003). All these tests have the null hypothesis of unit root. Panel A also includes Hadri (2000), which is a panel test for the null of stationarity. While Breitung (2001) allows for a homogeneous unit root under the alternative hypothesis, Im, Pesaran and Shin (2003) allows for heterogeneity and Maddala and Wu (1999) is a Fisher-type test for unit root.

A common feature of these tests is the assumption of independence across the panel elements. Since this assumption may be slightly unrealistic, to alleviate the effect of cross-sectional dependence between the elements of the panel, we subtract the cross-sectional averages for each period following Levin, Lin and Chu (2002). As a robustness check, following Breitung and Das (2005), we include Breitung and Das test, which is a panel test for unit root that builds on Breitung (2001), and allows for contemporaneous cross-correlation between the units of the panel.

These various unit root tests agree that the ETFs and their underlying benchmarks are nonstationary. Hadri (2000) provides evidence against the null of stationarity for both the log prices of ETFs and the benchmarks. These results confirm the non-stationary behaviour of the log prices of the ETFs and benchmarks, even when the data has been demeaned to mitigate the potential cross-correlation among the elements of the panel or when such cross-correlation is allowed for.

2.5. Tracking versus return matching

The methodology proposed in this chapter to model ETFs' tracking is motivated by the fact that the prior literature has generally overlooked the dynamics of the data when computing the TEs. In the previous section we concluded that the levels of the index and those of the ETFs appear to be non-stationary. However, the evidence in Figures 2.1 and 2.2 suggests that there exists a degree of co-movement between the log prices of the ETFs and their benchmarks. Therefore, we formally test for the presence of a cointegrating relationship between the ETFs and their benchmarks using two approaches. The first approach is based on panel test for the null of cointegration following Pedroni (1999). Pedroni's approach employs four tests based on the within panel dimension, a non-parametric extension of the variance ratio statistic proposed by Phillips and Ouliaris (1990), which we denote as panel v-stat in Table 2.1, a panel version of the Phillips and Perron (1988) and Phillips and Ouliaris (1990) unit root tests, represented by panel rho-stat in Table 2.1, a semi-parametric adjustment of Phillips and Perron (1988) t-test statistic, represented by panel pp-stat in Table 2.1, and a panel form of the Augmented Dickey Fuller t-statistic (Fuller, 1976), which is denoted as aDF-stat in Table 2.1. Pedroni (1999) also employs three tests based on the between panel dimension, which can be seen as the group mean version of the within dimension statistics. Pedroni (1999) shows that the asymptotic distributions of these seven statistics follow a standard normal. The results of the first set of tests are displayed in Panel B of Table 2.1. Our second approach, which we employ for the purposes of robustness, is developed in Westerlund (2007) and tests the null of no-cointegration using a panel ECM. This approach tests the null of no cointegration by determining whether the estimates of the individual error correction terms in the panel are statistically significant. This test can accommodate heterogeneous dynamics, either in the long or in the short run, together with serial correlation and non-strictly exogenous repressors. The test is performed using error correction based cointegration tests for panel data (Persyn and Westerlund 2008). While the first two tests (Pt and Pa) in Panel C test the alternative hypothesis that the entire panel is cointegrated, the other two tests (Gt and Ga), are group mean statistics, which test the alternative that at least one series of the panel is cointegrated. To account for cross-correlation we bootstrap the critical values using 1000 replications. The results, which are displayed in Panel C of Table 2.1, are consistent with the view that f_t and i_t share a common stochastic trend.

Since the ETFs and their benchmark indices appear to be I (1) and share a common stochastic trend, we employ Fully Modified OLS (FMOLS) and Dynamic OLS (DOLS), following Pedroni (2001, 2004), to estimate the cointegrating relationship between the levels of the ETF and the underlying index defined in Eq. (2.8).

[Table 2.2]

[Table 2.3]

For the individual series, the coefficient estimates and t-stats of the FMOLS for the null hypothesis $H_0:\varphi_1 = 1$ are displayed in Table 2.2 (equity ETFs) and Table 2.3 (debt ETFs)[4]. When the individual series are analysed there are a couple of points worth mentioning. The first one is that there is a unique ETF, Powershrares Rusell Micap Pure Value (Bloomberg Ticker: PXMV), which fails to reject the null $H_0:\varphi_1 = 1$. This result suggests that PXMV ETF displays strong tracking. The second point is that the majority of the cointegrating coefficients for debt ETFs are consistently lower than those obtained for the equity ETFs, which are typically very close to unity.

[Table 2.4]

The panel coefficient estimates and the t-statistics for the null hypothesis $H_0:\varphi_1 = 1$ are displayed in Table 2.4. Both approaches, FMOLS and DOLS, provide analogous results. Regardless of the approach used to estimate the cointegrating relationship, the panel coefficient estimate of the slope is 0.74, and becomes closer to unity when we add time dummies. Both methodologies coincide in rejecting the null hypothesis $H_0:\varphi_1 = 1$ for the panel. Accordingly, we can conclude that the weak tracking, instead of strong tracking, is in place for most of the individual ETFs in our panel and for the panel as a whole.

Since the levels of the fund and the benchmark appear non-stationary and cointegrated, we construct an ECM following Engle and Granger (1987) to link the short-term dynamics with the long run equilibrium. We use two different approaches to estimate the panel ECM in Eq. (2.10) and present the results in Table 2.5.

4 Since the coefficient estimates and t-stats of the FMOLS are analogous to those provided by the DOLS, the latter is not displayed for brevity. These results are available upon request.

[Table 2.5]

The ECM essentially supports the idea that when asset prices share a common stochastic trend, even if they drift apart in the short run, they will eventually correct back to the cointegrating equilibrium. To account for heteroscedasticity, serial correlation and potential cross-correlation in the data, we estimate the ECM following Driscoll and Kraay (1998). We also estimate the ECM in Eq. (2.10) using fixed effects, as the Hausman test (see Hausman 1978) rejects the null hypothesis of random effects. We use two-way fixed effects to account for individual and time effects, since both appear to be significant in our sample. Further, to allow for potential cross-correlation in the data we bootstrap the coefficients using 1000 replications. Overall, the results of Table 2.5 verify the cointegrating relationship between the levels of the ETF and the benchmark, and at the same time, corroborate our previous findings. Therefore, we can conclude that, regardless of the estimation technique used, the coefficient estimates of the slope are near to unity and the rates of adjustment to the long run equilibrium are negative and statistically significant at all usual levels of confidence.

2.6. Tracking errors

The recent growth of passively managed index funds motivates the use of TEs to evaluate their tracking performance. The concept of TE was originally employed to assess the performance of active portfolio managers, which basically compared the performance of the manager with that of a benchmark. Any differences in performance generate a TE. There exist several definitions of the TE in the literature. Roll (1992) defines the TE as the variance of the deviation between a minimum variance portfolio and its benchmark. Pope and Yadav (1994) presents two measures of the TE. One of the measures defines the TE as the variance of the return differential between the fund and the benchmark, while the other computes the TE as the standard error of the residuals obtained by regressing the returns of an index tracking fund on those of its benchmark. TE has also been defined as the average of the absolute difference between the returns of the fund and the benchmark (Rudolf, Wolter and Zimmermann 1999).

Since passively managed ETFs are designed to closely match the performance of a benchmark over time, the definitions of the TE in the ETF's literature are analogous to the ones used to measure the performance of active fund managers. This study employs the three most widely used metrics in the TEs literature for ETFs (see Frino and Gallagher 2001; Rompotis 2009). The first measure defines the TE as the standard error of the return matching regression (2.7), and it is denoted as $\xi'_{1,mn}$. We note in passing, that the TEs obtained from Eq. (2.7) used in the previous literature are unlikely to be reliable. To overcome this problem, the standard error of the ECM regression (2.10), denoted $\xi_{1,mn}$, is computed. The TE from Eq. (2.7), represented by $\xi'_{1,mn}$, is included in Table 2.6 for completeness. The second method, $\xi_{2,mn}$, computes the TE as the average of the absolute difference between the continuously compounded returns of the ETF and those of the benchmark. Hence, the TE for the m^{th} fund and the n^{th} benchmark can be written as follows:

$$\xi_{2,mn} = \frac{\sum_{t=1}^{T} \left| e_{mn,t} \right|}{T} \tag{2.13}$$

Where $e_{mn,t} = \Delta F_{m,t} - \Delta I_{n,t}$ and T is the sample size. We compute $e_{mn,t}$ using the ECM, which assumes that weak tracking is put in place for most of the funds.

The third approach computes the TE as the standard deviation of the difference in continuously compounded returns between the ETF and its underlying index:

$$\xi_{3,mn} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left(e_{mn,t} - \bar{e}\right)^2} \tag{2.14}$$

Where, $e_{mn,t} = \Delta F_{m,t} - \Delta I_{n,t}$ and $\bar{e} = T^{-1}\sum_{t=1}^{T} e_{mn,t}$.

$\xi_{3,mn}$ is often referred to as the TE in the prospectus for ETF's and is computed as the square root of Eq. (2.2) under the assumption $\lambda^*_{t|t-1} = 1$.

[Table 2.6]

Table 2.6 displays the summary statistics of the TE using the ECM (2.10) together with those using the return matching regression (2.7), which are included for comparison purposes. The measures of the TE computed with the return matching regression ($\xi'_{1,mn}$, $\xi'_{2,mn}$ and $\xi'_{3,mn}$) will be unreliable in the presence of cointegration.

The TEs from the ECM, regardless of the metric employed, vary widely across funds. The first measure of the TE, $\xi_{1,mn}$, lies in the interval [0.00057, 0.00425] with average 0.00160, while $\xi_{3,mn}$ lies in a very similar interval [0.00424, 0.00057] and the average coincides with that of the first approach. The second metric, $\xi_{2,mn}$, lies in a narrower interval [0.00041, 0.00316]. In terms of debt ETFs, the TE gauged with the first and third metrics, $\xi_{1,mn}$ and $\xi_{3,mn}$, lie within the range [0.00012, 0.00589] and their average is 0.00241, while those computed with the second approach, $\xi_{2,mn}$, are contained in the interval [0.00009, 0.00355] and have an average value of 0.00162. In general, the magnitudes of the TE of debt ETFs are around 30% greater than those of equity ETFs, regardless of the method used to calculate them. This is in line with the outcomes of Table 2.2 and Table 2.3.

The three measures of the TE computed with the return matching regression behave in a similar fashion. However, regardless of the metric used to gauge the TE, the estimates based on the return matching regression (2.7) exceed those obtained using the ECM. This apparent overestimation of the TEs applies to all the ETFs in our sample, independently of the asset class they track. Comparison of the TE computed from the two models, for each metric, shows that the differences in TE are slightly greater for equity ETFs. On average, TEs based on Eq. (2.7) exceed those based on Eq. (2.10) by 3.22% across the three measures.

2.7. Monte-Carlo experiment

To illustrate the generality of our results about the impact of incorrectly ignoring the stochastic properties of the data when computing the TE, we present the outcomes from a small-scale Monte-Carlo experiment with 5000 replications. For ease of exposition, we generate a pair of cointegrated price series using the following DGP:

$$\begin{aligned} Y_t &= \mu + Y_{t-1} + \varsigma_{1,t} \\ X_t &= \omega Y_t + \varsigma_{2,t} \end{aligned} \tag{2.15}$$

Where the innovations $\varsigma_{j,t}$ are assumed to be independent and drawn as N (0,1). X represents an ETF and Y represents the benchmark index. By varying the value of ϖ

we can allow for differences in the ability of fund X to track index Y. We construct continuously compounded returns for both series and compute the three TE measures from the ECM:

$$\Delta y_t = \hat{\lambda}_0 + \hat{\lambda}_1 \Delta x_t + \rho u_{t-1} + z_{1t} \tag{2.16}$$

We also compute the three TE measures using the return matching regression:

$$\Delta y_t = \mu_0 + \mu_1 \Delta x_t + z_{2t} \tag{2.17}$$

We generate samples of size N=500, 1000 and 2000 for values of ϖ = 1, 0.90, 0.60 and 0.30.

[Table 2.7]

The outcomes of the Monte-Carlo experiment are presented in Table 2.7 and suggest that the three measures of the TE computed from the return matching approach overestimate the magnitude of the TE. The slope of the regression between the ETF and the benchmark in Eq. (2.15) is represented by ω. When ω = 1 the TE from the return matching regression, independent of the metric employed, overstates the TE by 15% on average. This difference remains the same regardless of the sample size. Conversely, when ω = 0.3 the difference between the two TE is still positive but it is remarkably smaller (1.8%). Although varying the sample size seems to have little impact upon the results, it worth noting that the largest and the smallest sample sizes in our experiment (2000 and 500 data points) provide slightly larger differences between the two approaches than the medium sample (1000 data points).

Comparing our results from the simulations with those obtained from our data is very informative. In our sample, since we include neither leveraged nor inverse ETFs, the coefficient of ω should only take values between zero and one. The closer is the coefficient of ω to unity, the greater the accuracy of the tracking in levels. Hence, further inspection of the outcomes in Table 2.7, shows that the differences between the TEs computed using the ECM and the return matching regression are inversely related to the accuracy of the tracking. Put simply, the nearer the value of ω to unity, the greater the differences between both TEs estimates.

In short, the outcomes from the Monte Carlo experiment corroborate our previous findings. The results confirm that computing the TEs using a return matching regression, instead of the ECM, exaggerates the magnitude of the TEs, independently of the metric employed. Moreover, the differences between the TEs computed with the ECM and those from the return matching regression are inversely related to the accuracy of the tracking. This implies that overestimation of the TE might have a greater impact on equity ETFs than on debt ETFs, since the former generally track their benchmark more precisely. Consequently, when investors and money managers select ETFs based on their tracking performance, overestimation of the TEs might result in erroneous investment choices. In other words, evaluating ETFs' tracking performance with the ECM proposed in this chapter is key to avoid misleading estimates and suboptimal investment decisions, particularly for portfolios including equity ETFs. While our focus is on the performance of a sample of U.S. ETFs, the methodology employed is valid for any passively managed index tracking product.

2.8. The economic consequences of invalid inferences

TEs are often employed as a tool to evaluate ETFs' performance by fund managers and investors. The evidence suggests that using return matching regressions, rather than the ECM, to evaluate ETFs' performance results in overestimation of the TEs and consequently, might result in erroneous investment choices. In this section we aim to illustrate the impact of incorrect estimation of the TEs in terms of portfolio selection. Hence, using daily data, we construct six equally weighted portfolios. Each portfolio includes the five ETFs with the lowest TE chosen by each model.

[Table 2.8]

The composition of the equally weighted portfolios is displayed in Panel A of Table 2.8. The Bloomberg ticker is used to identify the ETFs. Note that only equity funds are employed for the ranking since they track the benchmark indices more accurately and provide lower TEs than their debt counterparts.

The first three columns in the table are the measures of the TEs obtained with the ECM, while the last three columns display the three TE measures from the return matching regression. Essentially, we observe that there exist differences in terms of the funds included by each measure and also in the ordering of the ETFs in each portfolio. The outcomes show that first and the third measures of the TE computed with the ECM coincide. Similarly, the first and the third measures of the TE using the return matching regression provide similar rankings. However, the rankings provided by the ECM and the return matching regression for the first and the third metrics differ in both, the ordering and the constituents of the ETFs with the smallest TE. On the contrary, the second measure of TE from the ECM and the return matching regression selects the same ETFs with the lowest TE, but order of the funds differently.

In this very simple exercise, the metrics based on the return matching regression provide a different ranking of funds to the metrics based on the ECM, in terms of the constituents of the top five and the ordering therein. We note that the equal weighting scheme does not pay any attention to the ranking within the top five. Furthermore, comparison of the returns provided by the first and third metrics of the TEs obtained from the ECM and the return matching regression shows that the portfolio constructed using the ranking of the former results in a greater mean return (15%) and slightly lower risk (0.0081%) than that of the latter. The average return and risk of a portfolio constructed using the rankings of the second measure of the TE are analogous, independently of the model used to obtain them.

2.9. Conclusion and discussion

We examine the tracking ability for a sample of 59 passively managed equity ETFs and 18 passively managed debt ETFs listed on the NYSE over the period 14th December 2009 to 14th December 2016.

Most of the existing literature on ETFs' tracking performance computes the TE using a return matching regression, which regresses the returns of the ETF on those of the underlying index and that does not allow for the presence of cointegration. To avoid any misspecification when computing the TE, we propose a framework that considers the stochastic nature of the data and assesses the quality of the ETF's tracking accordingly.

The empirical outcomes suggest that levels of the ETFs and those of the benchmark indices appear to be unit root processes. Moreover, the various ETFs and their underlying indices share a common stochastic trend. This implies that the data uniformly reject the return matching hypothesis, which only obtains in the absence of cointegration.

In general, we observe that ETFs do not strongly track the underlying indices since the estimated cointegrating coefficient is significantly different to unity for most of the ETFs, and for the panel as a whole. This result suggests that a form of weak or relative tracking is in place and hence TEs arise. The differences observed in the magnitude of the cointegrating coefficients seem to indicate that equity ETFs track the underlying indices more precisely than their debt counterparts.

Our findings demonstrate that the TEs computed using a return matching regression, as opposed to the ECM, exaggerate the magnitude of the errors, regardless of the metric employed to gauge them or the type of security included in the benchmark index. Further, comparison of the TEs estimates shows that the differences between the two approaches increase with the tracking accuracy. As a result, the differences appear most pronounced for equity funds which generally track the benchmark more accurately than debt funds.

Overall, the results imply that omitting the cointegrating relationship between the fund and the underlying index tends to provide inferior estimates of a fund's ability to track its benchmark, which eventually can affect investors' decisions. The main implication for investors and money managers, who select ETFs based on the TE from return matching regressions, is that they are likely to make erroneous investment choices based on the estimates provided by the return matching approach. Hence, we argue that assessing ETFs' tracking performance should account for the stochastic nature of the data.

A series of Monte-Carlo simulations were employed to illustrate the consequences of ignoring the stochastic properties of the data and demonstrate the generality of our results. We illustrate the impact of incorrectly specifying the model used to estimate the TEs using a very simple out-of-sample exercise. We construct six equally weighted portfolios and show that the ordering and the constituents of the portfolio selected, which includes those ETF with the smallest TE, depends critically on the nature of the model used to compute the TE.

In short, our results demonstrate the economic and statistical importance of correctly accommodating the stochastic nature of the data when evaluating ETFs' tracking performance.

Figures

Figure 2.1. The tracking ability of DIA ETF and the benchmark index

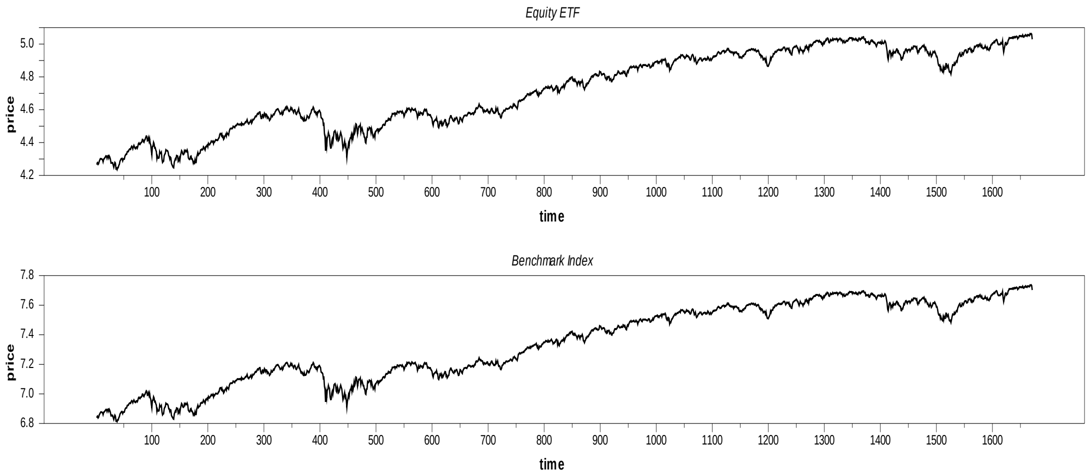


Figure 2.1. The former graph displays the logarithm of the closing prices of an equity ETF included in our sample, (Bloomberg Ticker: DIA), and the latter graph exhibits the logarithm of the closing prices of the underlying benchmark (Bloomberg Ticker: INDU).

Figure 2.2. The tracking ability of CSJ ETF and the benchmark index

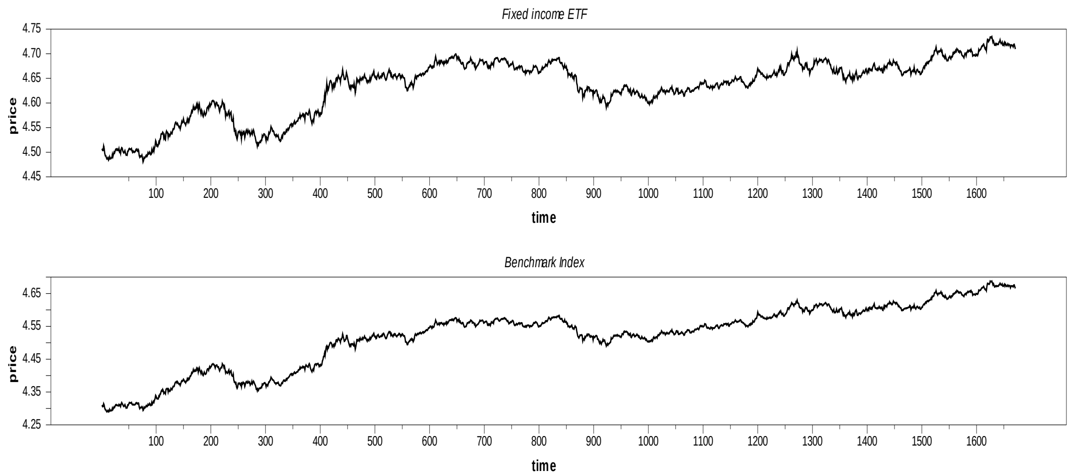


Figure 2.2. The former graph shows the logarithm of the closing prices of a debt ETF included in our sample, (Bloomberg Ticker: CSJ), the latter graph displays the logarithm of the closing prices of the underlying benchmark (Bloomberg Ticker: LD01TRUU).

Tables

Table 2.1. Panel unit root and panel cointegration tests

Panel A: Unit root tests					
Raw data	M-W	H	B	IPS	B-D
ETF	142.89	8000 [a]	4.78	0.14	0.69
Benchmark	75.71	9200 [a]	16.24	2.87	1.06
Demeaned	M-W	H	B	IPS	
ETF	162.99	6700 [a]	1.61	-0.97	
Benchmark	170.31	7000 [a]	-0.49	-1.22	

Panel B: Panel cointegration tests Pedroni (1999)			
Time dummies		No time dummies	
panel v-stat	-4.55	panel v-stat	0.56
panel rho-stat	-5.89 [a]	panel rho-stat	-19.10 [b]
panel pp-stat	-4.41 [a]	panel pp-stat	-11.64 [a]
panel aDF-stat	-5.24 [a]	panel aDF-stat	-11.46 [a]
group rho-stat	-60.13 [a]	group rho-stat	-67.31 [a]
group pp-stat	-26.04 [a]	group pp-stat	-22.25 [a]
group aDF-stat	-23.36 [b]	group aDF-stat	-18.72 [a]

Panel C: Panel cointegration tests Westerlund (2007)			
W-ECM Z-value		W-ECM Z-value bootstrapped	
Panel Gt	-1.79	Panel Gt	-1.89
Panel Ga	-15.24 [a]	Panel Ga	-15.61 [a]
Panel Pt	-18.36 [a]	Panel Pt	-4.23 [b]
Panel Pa	-7.26 [a]	Panel Pa	-4.09

Table 2.1. Panel A presents unit root test results for the tests: Maddala and Wu (1999) M-W; Hadri (2000) H; Breitung (2001) B, Levin, Lin and Chu (2002), LLC: Im, Pesaran and Shin (2003), IPS; Breitung and Das (2005), B-D. All the tests in Panel A have as the null hypothesis that the data contain a unit root except the Hadri test which has as the null hypothesis that the panel is stationary. All the tests are distributed as N (0,1) except M-W which follows a χ^2 with 2N degrees of freedom. Panel B presents tests of the null hypothesis of no cointegration following Pedroni (1999, 2004). Panel C presents test of the null of no cointegration following Westerlund (2007). All the cointegration tests are distributed as N(0,1). [a] Significant at 1% level, [b] Significant at 5% level, [c] Significant at 10% level.

Table 2.2: Fully modified least squares estimates for equity ETFs

Ticker	φ_1	t-stat	Ticker	φ_1	t-stat
SPY US	0.847521	-92.600957	RSP US	0.869942	-76.415912
MDY US	1.002916	11.897059	DVY US	0.736523	-122.230819
DIA US	0.998811	-4.350743	ITOT US	0.863005	-109.340367
IUSG US	0.893438	-104.331376	PFM US	0.783658	-96.076971
IVV US	0.851604	-87.416599	PXMG US	0.737067	-45.829419
IWB US	0.857072	-87.348900	PXMV US	1.005534	1.2644140
IJH US	0.895324	-69.445946	PXSV US	1.104616	16.998559
IJR US	0.909314	-68.589073	XLG US	0.975078	-34.429656
IVE US	0.816769	-72.046152	IWC US	0.894579	-50.692771
IVW US	0.879177	-113.0316	PEY US	0.709415	-129.358601
IWD US	0.825247	-70.795805	FDM US	0.969422	-42.766773
IWF US	0.887100	-106.251423	MDYG US	0.906815	-47.399655
IWM US	0.889706	-58.20916	SLY US	0.841854	-41.799026
IWV US	0.858627	-83.853977	SPHQ US	0.876647	-33.681424
IYY US	0.855786	-85.207197	PRF US	0.843351	-70.121298
IJJ US	0.864631	-64.533524	RFG US	0.945449	-44.906969
IJK US	0.923141	-68.010583	RFV US	0.878648	-48.770754
IJS US	0.885410	-63.682144	RPG US	0.939668	-74.311027
IJT US	0.930064	-71.083946	RPV US	1.002178	7.9384560
IWN US	0.824396	-52.306923	RZG US	0.946808	-54.575223
IWO US	0.943359	-63.486736	RZV US	0.920866	-40.798036
IUSV US	0.827481	-69.684361	VIG US	0.818965	-88.894303
SLYG US	0.889145	-24.079214	PKW US	0.905241	-68.806791
SLYV US	0.784924	-32.420331	IWY US	0.874239	-121.008456
SPYG US	0.887054	-88.018372	SCHA US	0.899707	-61.796120
SPYV US	0.815201	-71.783338	SCHB US	0.866275	-84.165449
OEF US	0.831814	-98.108723	SCHX US	0.857761	-88.258328
IWP US	0.915420	-73.499353	SCHG US	0.914064	-93.226839
IWR US	0.881014	-74.091571	SCHV US	0.795036	-83.232101
IWS US	0.850062	-70.006385			

Table 2.2: FMOLS estimates of φ_1 from Eq. (2.8) and t-stats for $H_0 : \varphi_1 = 1$ for equity ETFs.

Table 2.3: Fully modified least squares estimates for debt ETFs

Ticker	φ_1	t-stat	Ticker	φ_1	t-stat
LQD US	0.329233	-78.696149	MUB US	0.327828	-61.18127
LQD US	0.58853	-53.386028	MUB US	0.578938	-49.135592
IEF US	0.094983	-221.37472	PLW US	0.122725	-81.934733
CSJ US	0.248197	-78.998768	PHB US	-0.03101	-55.390666
CRED US	0.569219	-66.662193	JNK US	0.642777	-33.998779
IEI US	0.106443	-232.203589	EDV US	0.108417	-94.452838
SHV US	0.628023	-50.489069	SUB US	0.131914	-64.908941
TLH US	0.262902	-142.78314	HYD US	0.309179	-61.771487
SHY US	0.056546	-57.670126	LWC US	0.15128	-96.133285
HYG US	0.329233	-78.696149	TUZ US	0.327828	-61.18127

Table 2.3: FMOLS estimates of φ_1 from Eq. (2.8) and t-stats for $H_0 : \varphi_1 = 1$ for debt ETFs.

Table 2.4: Panel fully modified and dynamic least squares estimates

Panel cointegration estimates equity and debt ETFs				
	Time dummies		No time dummies	
	Coefficient	t-stat	Coefficient	t-stat
FMOLS	1.059	39.808	0.743	-624.987
DOLS	1.059	40.106	0.743	-628.153

Table 2.4. Panel FMOLS and DOLS estimates and t-stats from Eq. (2.8).

Table 2.5: Estimates of the error correction model

Panel error correction			
	κ_0	κ_1	γ_1
DKP	-0.00008 [a]	0.97027 [a]	-0.00008 [a]
	(0.0000)	(0.0019)	(0.0000)
TWFE	-0.00008 [a]	0.93729 [a]	-0.00010 [a]
	(0.0003)	(0.0147)	(0.0000)

Table 2.5. The table shows the outputs of the ECM regression Eq. (2.10) estimated using Driscoll and Kraay (1998), DKP, and two-way fixed effect with bootstrapped coefficients, TWFE. Standard inference applies. [a] Significant at 1% level, [b] Significant at 5% level, and [c] Significant at 10% level.

Table 2.6: Tracking errors

Panel A: Descriptive statistics						
	ξ_1	ξ_2	ξ_3	ξ'_1	ξ'_2	ξ'_3
Equity funds						
Average	0.001601	0.001024	0.001600	0.001644	0.001052	0.001674
Maximum	0.004250	0.003158	0.004240	0.004500	0.003258	0.004530
Minimum	0.000572	0.000405	0.000572	0.000581	0.000414	0.000583
S.Dev.	0.001132	0.000759	0.001131	0.001186	0.000784	0.001217
Debt funds						
Average	0.002411	0.001617	0.002409	0.002440	0.001672	0.002510
Maximum	0.005890	0.003548	0.005890	0.006070	0.003646	0.006130
Minimum	0.000120	0.000087	0.000120	0.000121	0.000090	0.000124
S.Dev.	0.001663	0.001063	0.001662	0.001707	0.001092	0.001729
Overall sample						
Average	0.001790	0.001163	0.001789	0.001830	0.001197	0.001869
Maximum	0.005890	0.003548	0.005890	0.006070	0.003646	0.006130
Minimum	0.000120	0.000087	0.000120	0.000121	0.000090	0.000124
S.Dev.	0.001310	0.000869	0.001309	0.001357	0.000898	0.001388

Table 2.6: Summary statistics of the tracking errors. ξ_1 is the TE defined as the standard error of the ECM regression $\xi_1^{'}$ the TE computed as the standard error of the return matching regression (2.7), ξ_2 and $\xi_2^{'}$ measure the extent to which the continuously compounded returns of the individual ETF diverge from those of the benchmark using ECM regression and Eq. (2.7), respectively. ξ_3 and $\xi_3^{'}$ are the standard deviation of the difference in continuously compounded returns between ETF and underlying index using ECM regression and Eq. (2.7), respectively. S.Dev. in the table represents the standard deviation.

Table 2.7: Monte-Carlo outcomes

Panel A: N=500						
ϖ	ξ_1	$\xi_1^{'}$	ξ_2	$\xi_2^{'}$	ξ_3	$\xi_3^{'}$
1.0	0.0038	0.0044	0.0028	0.0032	0.0038	0.0044
0.8	0.0042	0.0047	0.0031	0.0034	0.0042	0.0047
0.6	0.0047	0.0050	0.0034	0.0036	0.0047	0.0050
0.3	0.0053	0.0054	0.0038	0.0038	0.0053	0.0054
Panel B: N=1000						
ϖ	ξ_1	$\xi_1^{'}$	ξ_2	$\xi_2^{'}$	ξ_3	$\xi_3^{'}$
1.0	0.0024	0.0028	0.0016	0.0018	0.0024	0.0028
0.8	0.0027	0.0029	0.0017	0.0019	0.0027	0.0029
0.6	0.0029	0.0031	0.0019	0.0020	0.0029	0.0031
0.3	0.0032	0.0033	0.0021	0.0022	0.0032	0.0033
Panel C: N=2000						
ϖ	ξ_1	$\xi_1^{'}$	ξ_2	$\xi_2^{'}$	ξ_3	$\xi_3^{'}$
1.0	0.0021	0.0024	0.0013	0.0015	0.0021	0.0024
0.8	0.0023	0.0026	0.0015	0.0016	0.0023	0.0026
0.6	0.0025	0.0027	0.0016	0.0017	0.0025	0.0027
0.3	0.0028	0.0029	0.0018	0.0019	0.0028	0.0029

Table 2.7: Monte-Carlo simulations. ξ_1, ξ_2 and ξ_3 are the TE computed from Eq. (2.16). ξ'_1, ξ'_2 and ξ'_3 are the TE measures obtained from Eq. (2.17). The slope of the regression between the ETF and the benchmark following Eq. (2.15) is ϖ . Panel A, Panel B and Panel C display the results of the Monte-Carlo experiment for sample sizes N=500, 1000 and 2000, respectively.

Table 2.8: Out-of-sample exercise

Panel A

Ranking	ξ_1	ξ_2	ξ_3	ξ'_1	ξ'_2	ξ'_3
1	IJK US	IJK US	IJK US	IJK US	IJK US	IJK US
2	IWP US	IWP US	IWP US	IWP US	VIG US	IWP US
3	DIA US	VIG US	DIA US	RSP US	SPY US	RSP US
4	RSP US	SPY US	RSP US	DIA US	IWP US	DIA US
5	SCHB US	IVW US	SCHB US	IVW US	IVW US	IVW US

Panel B

Portfolio	ξ_1	ξ_2	ξ_3	ξ'_1	ξ'_2	ξ'_3
Mean	0.000797	0.000520	0.000797	0.000692	0.000520	0.000692
S. Dev.	0.006772	0.006627	0.006772	0.006828	0.006627	0.006828
Skew.	0.061614	0.071788	0.061614	0.081513	0.071788	0.081513
Kurt.	0.454536	0.571586	0.454536	0.370774	0.571586	0.370774

Table 2.8: Ranking of ETFs. ξ_1, ξ_2 and ξ_3 are the three measures of the TE computed with ECM and $\xi_1^{'}$, $\xi_2^{'}$ and $\xi_3^{'}$ are the three TE measures obtained from Eq. (2.16). Panel A displays the ETF ticker of the five ETFs with the smallest TE. The first three columns show the three definitions of the TE using ECM. The last three columns display the three measures of the TE computed with the return matching regression. Panel B present the summary statistics of an equally weighted portfolio containing the ETFs of Panel A, which are the ones with the smallest TE. S.Dev. in the table represents the standard deviation. Skew. and Kurt. represent the skewness and the kurtosis, respectively.

Appendix

Table A.2.1: Equity ETFs

Equity ETFs	Ticker	Inception
SPDR S&P500 ETF Trust	SPY US	22/01/1993
SPDR S&P500 ETF Trust	MDY US	04/05/1995
SPDR Dow Jones Ind. Average ETF	DIA US	14/01/1998
iShares Core US Growth ETF	IUSG US	24/04/2000
iShares Core S&P 500 ETF	IVV US	19/05/2000
iShares Russell 1000 ETF	IWB US	19/05/2000
iShares Core S&P Mid-Cap ETF	IJH US	26/05/2000
iShares Core S&P Small-Cap ETF	IJR US	26/05/2000
iShares S&P 500 Value ETF	IVE US	26/05/2000
iShares S&P 500 Growth ETF	IVW US	26/05/2000
iShares Russell 1000 Value ETF	IWD US	26/05/2000
iShares Russell 1000 Growth ET	IWF US	26/05/2000
iShares Russell 2000 ETF	IWM US	26/05/2000
iShares Russell 3000 ETF	IWV US	26/05/2000
iShares Dow Jones U.S. ETF	IYY US	16/06/2000
iShares S&P Mid-Cap 400 Value ETF	IJJ US	28/07/2000
iShares S&P Mid-Cap 400 Growth ETF	IJK US	28/07/2000
iShares S&P Small-Cap 600 Value ETF	IJS US	28/07/2000
iShares S&P Small-Cap 600 Growth ETF	IJT US	28/07/2000
iShares Russell 2000 Value ETF	IWN US	28/07/2000
iShares Russell 2000 Growth ETF	IWO US	28/07/2000
iShares Core US Value ETF	IUSV US	04/08/2000
SPDR S&P 600 Small Cap Growth ETF	SLYG US	29/09/2000
SPDR S&P 600 Small Cap Value ETF	SLYV US	29/09/2000
SPDR S&P 500 Growth ETF	SPYG US	29/09/2000
SPDR S&P 500 Value ETF	SPYV US	29/09/2000
iShares S&P 100 ETF	OEF US	27/10/2000
iShares Russell Mid-Cap Growth ETF	IWP US	01/01/2001
iShares Russell Mid-Cap ETF	IWR US	20/07/2001
iShares Russell Mid-Cap Value ETF	IWS US	24/07/2001

Guggenheim S&P 500 Equal Weight	RSP US	30/04/2003
iShares Select Dividend ETF	DVY US	07/11/2003
iShares Core S&P Total US Stock Market	ITOT US	23/01/2004
PowerShares Dividend Achievers ETF	PFM US	09/02/2004
PowerShares Russell Midcap Pure Growth	PXMG US	03/03/2005
PowerShares Russell Midcap Pure MidVal	PXMV US	03/03/2005
PowerShares Russell 2000 Pure Value	PXSV US	03/03/2005
Guggenheim S&P 500 Top 50 ETF	XLG US	20/05/2005
iShares Micro-Cap ETF	IWC US	16/08/2005
PowerShares High Yield Equity Div. Ach.	PEY US	15/09/2005
First Trust Dow Jones Select Microcap	FDM US	30/09/2005
SPDR S&P 400 Mid Cap Growth ETF	MDYG US	15/11/2005
SPDR S&P 600 Small Cap ETF	SLY US	15/11/2005
PowerShares S&P 500 Quality Portfolio	SPHQ US	06/12/2005
Powershares FTSE RAFI US 1000 ETF	PRF US	19/12/2005
Guggenheim S&P Midcap 400 Pure Growth	RFG US	07/03/2006
Guggenheim S&P Midcap 400 Pure Value	RFV US	07/03/2006
Guggenheim S&P 500 Pure Growth ETF	RPG US	07/03/2006
Guggenheim S&P 500 Pure Value ETF	RPV US	07/03/2006
Guggenheim S&P Smallcap 600 Pure Gr	RZG US	07/03/2006
Guggenheim S&P Smallcap 600 Pure Value	RZV US	07/03/2006
Vanguard Dividend Appreciation ETF	VIG US	27/04/2006
PowerShares Buyback Achievers Portfolio	PKW US	20/12/2006
iShares Russell Top 200 Growth ETF	IWY US	28/09/2009
Schwab US Small-Cap ETF	SCHA US	03/11/2009
Schwab US Broad Market ETF	SCHB US	03/11/2009
Schwab US Large-Cap ETF	SCHX US	03/11/2009
Schwab U.S. Large-Cap Growth ETF	SCHG US	11/12/2009
Schwab U.S. Large-Cap Value ETF	SCHV US	11/12/2009

Table A.2.1: The table includes the name of the ETF, the Bloomberg ticker and the inception date.

Table A.2.2: Debt ETFs

Debt ETFs	Ticker	Inception
iShares iBoxx $ Investment Grade Corp. Bond ETF	LQD US	26/07/2002
iShares 7-10 Year Treasury Bond ETF	IEF US	26/07/2002
iShares 1-3 Year Credit Bond ETF	CSJ US	11/01/2007
iShares Core US Credit Bond ETF	CRED US	11/01/2007
iShares 3-7 Year Treasury Bond ETF	IEI US	11/01/2007
iShares Short Treasury Bond ETF	SHV US	11/01/2007
iShares 10-20 Year Treasury Bond ETF	TLH US	11/01/2007
iShares 1-3 Year Treasury Bond ETF	SHY US	26/02/2007
iShares iBoxx $ High Yield Corporate Bond	HYG US	11/04/2007
iShares National Muni Bond ETF	MUB US	10/09/2007
PowerShares 1-30 Laddered Treasury Portfolio	PLW US	11/10/2007
PowerShares Fundam. High Yield Corp. Bond	PHB US	15/11/2007
SPDR Barclays High Yield Bond ETF	JNK US	04/12/2007
Vanguard Extended Duration Treasury ETF	EDV US	13/12/2007
iShares Short-Term National Municipal ETF	SUB US	07/11/2008
VanEck Vectors High-Yield Municipal ETF	HYD US	05/02/2009
SPDR Barclays Long Term Corporate Bond	LWC US	11/03/2009
PIMCO 1-3 Year U.S. Treasury ETF	TUZ US	02/06/2009

Table A.2.2: The table includes the name of the ETF, the Bloomberg ticker and the inception date.

Table A.2.3: ETFs listed on the London stock exchange and benchmarks

ETF	Benchmark index
Hsbc FTSE 100 ETF (HUKX)	FTSE 100 TRI (TUKXG)
Hsbc Ftse 250 ETF (HMCX)	FTSE 250 TRI (FTPTT250)
iShares Core Ftse 100 ETF (ISF)	FTSE 100 TRI (TUKXG)
iShares Ftse 250 ETF (MIDD)	FTSE 250 TRI (FTPTT250)
Source Ftse 100 ETF (S100)	FTSE 100 TRI (TUKXG)
Source Ftse 250 ETF (S250)	FTSE 250 TRI (FTPTT250)
Db xt Ftse 100 Short Daily ETF (XUKS)	FTSE 100 Short TRI (UKXS100)
PowerShares Ftse RAFI ETF (PSRU)	FTSE RAFI UK 100 TRI (TFRGB1NG)
Lyxor ETF Ftse 100 (L100)	FTSE 100 TRI (TUKXG)
Lyxor ETF Ftse 250 (L250)	FTSE 250 TRI (FTPTT250)
Lyxor ETF Ftse All Sh. (LFAS)	FTSE All Shares TRI (FTPTTALL)
Etfs Ftse 100 Lev. Daily 2x (LUK2)	FTSE 100 Daily Lev. TRI (TUKXL2G)
Db xt Ftse 100 ETF DR (XUKX)	FTSE 100 TRI (TUKXG)
Db xt Ftse 250 ETF (XMCX)	FTSE 250 TRI (FTPTT250)
Db xt Ftse All-Share ETF (XASX)	FTSE All Share TRI (FTPTTALL)
Etfs Ftse 100 Super Short Daily2x (SUK2)	FTSE 100 Daily Super Short 2Index (TUKXI2G)

Table A.2.3: Equity ETFs and benchmark indices. The table displays the name of the ETF and corresponding benchmarks. Bloomberg tickers appear in parenthesis. TRI denotes Total Return Index.

Table A.2.4: Panel unit root and cointegration tests

Panel A: Panel unit root tests

	M-W	H	B	LLC	IPS	B-D
ETF	22.144	1600.000[a]	2.442	-1.055	0.568	0.679
Benchmark	13.003	1600.000[a]	3.681	-1.196	1.6084	0.897

Panel B: Cointegration tests

Time dummies		No time dummies	
panel v-stat	3.700[a]	panel v-stat	4.840[a]
panel rho-stat	-23.170[a]	panel rho-stat	-30.000[a]
panel pp-stat	-11.120[a]	panel pp-stat	-14.250[a]
panel adf-stat	-6.220[a]	panel adf-stat	-6.320[a]
group rho-stat	-91.74[a]	group rho-stat	-198.62[a]
group pp-stat	-26.38[a]	group pp-stat	-42.00[a]
group adf-stat	-13.09[a]	group adf-stat	-16.07[a]

Panel C: Cointegration tests

W-ECM Z-value		W-ECM Z-value Bootstrap	
Panel Gt	-6.013[a]	Panel Gt	-5.902[a]
Panel Ga	-30.070[a]	Panel Ga	-27.246[a]
Panel Pt	-3.965[a]	Panel Pt	-3.970[b]
Panel Pa	-11.236[a]	Panel Pa	-10.459[b]

Table A.2.4. Panel unit root tests. Panel A presents unit root test results for the following tests: Maddala and Wu (1999) M-W; Hadri (2000) H; Breitung (2001) B, Levin, Lin and Chu (2002), LLC: Im, Pesaran and Shin (2003), IPS; Breitung and Das (2005), B-D. All the tests in Panel A have as the null hypothesis that the data contain a unit root except the Hadri test which has as the null hypothesis that the panel is stationary. All the tests are distributed as N (0,1) except M-W which follows a χ^2 with 2N degrees of freedom. Panel B presents tests of the null hypothesis of no cointegration following Pedroni (1999, 2004). Panel C presents test of the null of no cointegration following Westerlund (2007). All the cointegration tests are distributed as N(0,1). Significant at 1% level, [b] Significant at 5% level, and [c] Significant at 10% level.

Table A.2.5. Individual cointegration estimates

Panel A: Individual cointegration estimates				
ETF	φ_1 (FMOLS)	t-stat	φ_1 (DOLS)	t-stat
HUKX	0.6139	-42.8049	0.6119	-43.5590
HMCX	0.7987	-44.5169	0.7984	-44.4426
ISF	0.6250	-41.6769	0.6230	-42.4709
MIDD	0.8126	-38.1700	0.8124	-37.9913
S100	0.9624	-29.3653	0.9623	-29.1901
S250	0.9756	-33.3698	0.9756	-33.0539
XUKS	1.0838	40.0256	1.0843	41.5246
PSRU	0.5961	-37.0547	0.5948	-0.37266
L100	0.9757	-22.1039	0.9756	-22.0633
L250	0.9784	-26.7549	0.9785	-26.4977
LFAS	0.9623	-34.0297	0.9622	-33.9317
LUK2	0.8304	-32.5601	0.8293	-32.7654
XUKX	0.6499	-34.1621	0.6485	-34.1022
XMCX	0.8361	-34.9731	0.8360	-34.7054
XASX	0.6918	-34.8343	0.6912	-34.6973
SUK2	1.1063	52.8964	1.1067	53.9793

Panel B: Panel cointegration estimates				
	φ_1 (FMOLS)	t-stat	φ_1 (DOLS)	t-stat
No time dummies	0.8436	-98.3636	0.8432	-97.8084
Time dummies	0.9637	-18.3096	0.9633	-17.7529

Table A.2.5: Cointegration estimates. Panel A displays the coefficient estimates of Eq. (2.8) of Chapter 2 using Dynamic Least Squares (DOLS) and Fully Modified Least Squares (FMOLS) for the individual funds. Panel B shows the coefficient estimates of the same regression using DOLS and FMOLS for the panel. T-stats are for $H_0: \varphi_1 = 1$ and standard inference applies.

Table A.2.6. Error correction model for individual funds

ETF	κ_0	S.E.	κ_1	S.E.	γ_1	S.E.
HUKX	-0.0001[b]	(0.000)	0.9726[a]	(0.007)	-0.0068[b]	(0.003)
HMCX	-0.0001	(0.000)	0.9497[a]	(0.014)	-0.0176[a]	(0.006)
ISF	-0.0001 [a]	(0.000)	0.9941[a]	(0.008)	-0.0049[c]	(0.003)
MIDD	-0.0001[a]	(0.000)	0.9608[a]	(0.009)	-0.0129[a]	(0.002)
S100	0.0000	(0.000)	0.9194[a]	(0.022)	-0.4459[a]	(0.006)
S250	0.0000	(0.000)	0.9301[a]	(0.017)	-0.4586[a]	(0.079)
XUKS	-0.0001	(0.000)	0.8095[a]	(0.025)	-0.4804[a]	(0.037)
PSRU	-0.0001	(0.000)	0.9389[a]	(0.014)	-0.0113	(0.007)
L100	-0.0000	(0.000)	0.9906[a]	(0.005)	-0.2293[a]	(0.059)
L250	0.0000	(0.000)	0.9602[a]	(0.009)	-0.4530[a]	(0.081)
LFAS	-0.0000	(0.000)	0.9664[a]	(0.007)	-0.3486[a]	(0.079)
LUK2	-0.0001[a]	(0.000)	0.9741[a]	(0.005)	-0.0136[a]	(0.003)
XUKX	-0.0001	(0.000)	0.9578[a]	(0.021)	-0.0555	(0.046)
XMCX	-0.0001[b]	(0.000)	0.9807[a]	(0.007)	-0.0115[a]	(0.004)
XASX	-0.0001[b]	(0.000)	0.9831[a]	(0.006)	-0.0092[a]	(0.003)
SUK2	-0.0001	(0.000)	0.9884[a]	(0.011)	-0.2880[a]	(0.090)

Table A.2.6: Estimates of the Error correction model for individual funds as in (2.10). S.E represents the Standard Error. [a] Significant at 1% level, [b] 5% level, [c] 10% level.

Table A.2.7. Panel error correction model

Estimation Method	κ_0	S.E.	κ_1	S.E.	γ_1	S.E.
PBSE	-0.0001[a]	(0.000)	0.9507[a]	(0.016)	-0.0462[b]	(0.019)
DKP	-0.0001[c]	(0.000)	0.9507[a]	(0.006)	-0.0462[a]	(0.012)
MGW	0.48771	(0.386)	0.8737[a]	(0.036)	-0.0652[a]	(0.019)

Table A.2.7. Panel error correction model estimates of Eq. (2.10). PBSE, represents the pooled OLS regression with bootstrapped SE. DKP, is the Driscoll and Kraay (1998) method. MGW is the mean group estimator bootstrapped with 1000 replications following Westerlund (2007). S.E represents the Standard Error. [a] Significant at 1% level, [b] at 5% level, [c] 10% level.

Chapter 3

The Impact of Liquidity on Exchange Traded Fund's Tracking Performance

3.1. Introduction

3.1.1. Introduction and motivation of the chapter

In the 24 years since the introduction of the first ETF in the U.S. this asset class has become one of the fastest growing segments in the market. At the end of September 2017, the NYSE Arca, reports a total of 1,451 ETF listed with total assets under management of approximately $2.61 trillion[5].

[5] Retrieved from https://www.nyse.com/publicdocs/nyse/products/etp-funds/sept_2017_monthly_flash.pdf.

An ETF can be defined as a basket of securities set to track the performance of a benchmark index. Any departure from the benchmark might give rise to the so-called TE. The replication accuracy of the ETF is measured by the TE, which in turn is considered a key metric when assessing ETF's performance. ETFs have a unique structure because they combine the creation and redemption processes of mutual funds with the continuous trading in secondary markets of stocks.

Given the importance of the TE as a selection criterion, and the special structure of the ETF, the main aim of this chapter is to study the extent to which the cross-sectional differences on the ETF's tracking performance are related to ETF's illiquidity. Liquidity has been acknowledged by the prior literature as one the main factors explaining TEs (Buetow and Henderson 2012; Drenovak, Urošević and Jelic 2014; Bertone, Paeglis and Ravi 2015; among others). Since liquidity cannot be directly measured using a unique variable, the previous literature has mainly focused on generic proxies of market liquidity such as bid-ask spreads (Drenovak, Urošević and Jelic 2014; Broman 2016; Broman and Shum 2018) or trading volume (Buetow and Henderson 2012; Chu 2011). However, we conjecture that to evaluate the ETF's illiquidity requires more specific proxies which can account for the special structure of the ETF's contract. As a result, we distinguish between primary liquidity, which relates to ETF's creation-redemption processes, and secondary liquidity, which is linked to the trading activity in the market for the ETF, and we use eight alternative proxiess to capture the main aspects of ETF's liquidity.

This chapter is closely related to Rompotis (2009), Shin and Soydemir (2010), Chu (2011), Buetow and Henderson (2012) and Bertone, Paeglis and Ravi (2015). A common feature of these studies is that they compute the TE using a pairwise regression that ignores the stochastic properties of the data under consideration and might provide inefficient estimates of the TEs. To overcome this problem, the TEs in this study are calculated following the approach developed in Chapter 2, which examines the properties of a self-funding portfolio which should have zero risk and zero return under ideal tracking conditions. This approach uses an ECM to calculate the TEs, and accounts for the presence of a common stochastic trend in the prices of the fund and the benchmark index.

To date, most of the research studying the tracking efficiency for equity and debt ETFs has been done separately, and the research on the tracking ability of non-equity ETFs is still very limited. Following Chordia, Sarkar and Subrahmanyam (2004), we include a mix of equity and debt ETFs in our sample and assess the extent to which the impact of common ETF's liquidity factors on the TE depends on the asset class tracked by the ETF.

Our evidence suggests that, as the ETF and the benchmark appear to be cointegrated, an ECM should be used to compute appropriately the TEs. Comparison of the tracking performance of ETFs tracking equity and debt shows that the former track their benchmarks more closely than the latter, which agrees with the findings of Chapter 2, and may be due to the relatively lower liquidity in debt markets. Overall, the results demonstrate that the illiquidity resulting from the creation-redemption processes plays a key role in determining the tracking quality of the ETFs, regardless of the asset class they track. This outcome illustrates the difficulties experienced by the authorised participants during the creation and redemption processes when the ETF's price and the NAV of the underlying securities diverge. For example, in the case of very illiquid bond or small capitalisation stocks, which generally trade infrequently, disparities between price and NAV may arise and persist. In some cases, such deviations may take time to resolve and affect the accuracy of the fund's tracking performance. Moreover, the empirical evidence shows that the quality of the tracking improves as the turnover increases, but it decreases with the bid-ask spreads and the volatility of the ETF.

We also study whether the TEs are related to the asset class tracked by the fund. In short, the illiquidity originated in the creation and redemption processes remains a key determinant of the funds' tracking performance, independently of the asset class included in the benchmark index. In terms of the secondary liquidity however, the fund turnover is negatively related to the TE, for both equity and debt ETFs. For equity ETFs the evidence suggests that the bid-ask spread is positively related to the TEs in a statistically significant fashion. Furthermore, there is some evidence that the duration of the bond portfolio is positively related to the TE.

Given the recent growth in passively managed index funds, the results of this chapter might have implications investors who purchase ETFs to benefit from the risk diversification, transparency, intraday liquidity and low trading costs provided by these funds.

3.1.2. Creation-redemption mechanism

It is useful to consider the role of the creation and redemption mechanisms in understanding the general structure of an ETF.

-Figure 3.1-

Essentially, the creation and redemption mechanisms provide ETFs with the ability to increase the number of ETF's shares outstanding in response to demand pressures of the ETF in the secondary market or reduce the number of shares outstanding resulting from a selling pressure of the ETF. Figure 3.1 illustrates how the ETF's primary and secondary market work. ETF's shares must be created on the primary market by the so-called Authorised Participants, henceforth AP, before they can trade on secondary markets. Generally, AP are large instructional investors or market makers with an agreement with the ETF's sponsor to create and redeem ETF's shares. AP play a key role in the ETF's creation and redemption activities and consequently in the ETF's liquidity.

The ETF's creation process requires that the AP deposits the basket of securities included in the ETF in exchange for ETF's shares. Although most of the ETF's creation process is done in kind, some ETFs allow for cash creation and redemption. Under some circumstances, AP can exchange the ETF's shares for a combination of a cash and securities. In the in-kind creation process, the AP deposits a portfolio of securities into the ETF, which generally matches the holdings of the fund, in exchange for a specified number of ETF's creation units. A creation unit represents a block of a certain number of shares, which usually ranges between 25.000 and 100.000 shares. Analogously, ETF's shares can be redeemed in creation units. In the in-kind redemption process, the creation units are deposited in exchange for the basket of securities, which match the composition of the ETF's portfolio, together with cash. The cash payment is the compensation for undistributed income and/or any differences in the NAV, among other costs.

ETF's shares can only be created and redeemed when grouped in creation units. New shares are created and redeemed on an on-going basis depending on the demand and supply of ETF's shares in the secondary market. The prices at which creations and redemptions occur are based on the next calculation of NAV after a creation or redemption order is received. Generally, a standard fee for the creation-redemption transactions is charged to the AP on the day the transaction takes place. This standard fee is the same regardless of the units deposited or issued. In some cases, ETFs also charge a small percentage of the NAV per creation unit[6].

The next section contains the literature survey, the third section provides an overview of the methodology employed to assess the ETFs' tracking performance. The fourth section includes the time series description of the data and some preliminary tests. The fifth section describes the different methods used to compute the TE and evaluates the ETFs' tracking performance. The sixth section provides a cross-sectional analysis of the determinants of the TEs observed. The final section provides a summary of the findings and some concluding comments.

3.2. Literature survey

ETFs constitute a very easy and convenient way to achieve portfolio diversification. TE, represent any departure between the performance of the fund and that of the benchmark. From the theoretical point of view, the replication process of passively managed funds seems straightforward. In fact, in a frictionless world the index can be replicated without costs. In practice, however, perfect replication appears to be unachievable due to the market frictions that ETFs face during the replication process. TEs play a very relevant role in passive index management since it provides investors with information about the accuracy of the ETF's tracking performance. In effect, the smaller the size of TE the more accurate is the ETF's tracking ability.

[6] For instance, the prospectus of SUB ETF, which tracks a benchmark composed of investment-grade U.S. municipal bonds with maturities between 1-5 years, reads that the standard fee is $100 and the variable fee ranges between 2% and 3% of the NAV per creation unit, when the creation-redemption is settled in cash.

Numerous factors have been related to the TEs in the literature. Blume and Edelen (2004) claims that departure from full replication in S&P 500 index trackers might enhance returns, but in turn increases the magnitude of the TEs. Besides, the authors suggest that minimisation of the TEs could be achieved by synthetic replication of the underlying benchmark. Evidence of the relationship between TEs and the fund expense ratio was reported in Rompotis (2009). Using three different measures of the TE, the study finds that the management expenses charged by the fund appear to be the key determinant of the TEs exhibited by the equity ETFs and IMFs included in their sample. Using a correlation approach, Johnson (2009) reports that the number of overlapping hours between domestic and U.S. markets and the relative returns of the home index with respect to those of the S&P 500 index constitute the main determinants of the TEs observed in a sample of U.S. country ETFs. Shin and Soydemir (2010) investigates the tracking ability of a mixed sample of ETFs which includes U.S. county funds denominated in local currency and broad U.S. equity funds. The study identifies the management expenses charged by the fund, the volatility of ETFs' daily returns, the average of the logarithm daily volume, the annual dividend and the exchange rates as the potential factors underlying the TEs observed. The study shows that the main determinant underlying the TE appears to be the exchange rates. Chu (2011) studies the TEs of 18 equity ETFs listed in Hong Kong stock exchange and identifies a negative relationship between the fund size, measured by fund total assets in millions, and the TEs. On the other hand, the expense ratios charged by the ETFs appear to be insignificantly related to the magnitude of TEs, which seems to contradict the findings of Rompotis (2009). The main factors underlying the ETFs' TE were also investigated by Buetow and Henderson (2012). Interestingly, the paper uses a broad sample of ETFs listed in the

U.S that includes equity, fixed income, real estate ETFs, together with international and leveraged ETFs. The study finds that ETFs' liquidity, represented by fund size and volume traded (in dollars), is inversely related to the TEs observed mainly for equity, fixed income and real estate ETFs. Blitz, Huij and Swinkels (2012) constitutes a pioneer study on the tracking performance of equity ETFs in Europe. The paper includes a sample both IMFs and ETFs and links the TE detected to the expense ratio charged by the fund and to dividend taxes. The tracking performance of European ETFs is also examined in Drenovak, Urošević and Jelic (2014). The study focuses exclusively in bond ETFs. The authors relate the TEs to the volatility of the benchmark, the duration of the bond portfolio, the bid-ask spreads, the replication strategies and the size of the fund. More recently, Bertone, Paeglis and Ravi (2015) compute the TEs of the DIA, an equity ETF listed in the U.S. that tracks the Dow Jones Industrial Average index, using intraday data. The study reports a negative relationship between the trading volume and the TEs and a positive relationship between the volatility, of both the ETF and the index, and the quoted spreads.

We observe that liquidity has been acknowledged by the prior literature as one the main factors explaining TEs. As liquidity cannot be directly measured using a unique variable, the previous literature has employed generic proxies of market liquidity and linked them to the ETFs' tracking performance. However, in this chapter we conjecture that evaluation of the role of ETF's liquidity requires more specific proxies that can account for the special structure of the ETF's contract. The usage of more tailored proxies might shed light on the role played by ETFs' illiquidity on their tracking performance.

3.3. Tracking performance

To study the tracking ability of the ETFs in our sample, we follow the methodology introduced in the Chapter 2, which is based on the idea that the tracking ability of an ETF is related to the optimal hedge ratio for a portfolio which is long one unit of the ETF financed by a short position of λ units of the underlying benchmark.

Most existing research on TEs is based on the return matching regression in Eq. (3.1), (see Elton, Gruber, Comer and Li 2002, Milonas and Rompotis 2010; among others). If F_t and I_t are the prices of the ETF and benchmark index, respectively, with the corresponding logarithms f_t and i_t. The actual return to holding the index for one period is computed as $\Delta I_t = \log(i_t / i_{t-1})$, and that of the ETF is calculated as $\Delta F_t = \log(f_t / f_{t-1})$. Therefore, the return matching regression, for the mth ETF and the nth benchmark, can be written as follows:

$$\Delta F_{m,t} = a + b\Delta I_{n,t} + \in_{mn,t} \tag{3.1}$$

However, when a fund and its benchmark share a common stochastic trend, the return matching approach used by the prior literature is inappropriate for three reasons. Firstly, Eq. (3.1) is a regression in differences which ignores the long run relationship between the price of the levels of the fund and the benchmark (see Engle and Granger 1987). Secondly, any estimates of *b* from (3.1) will be biased downwards, (see Kroner and Sultan, 1993 for further discussion). Thirdly, Chapter 2 provides simulation and empirical evidence that the commonly used measures of the TEs based on Eq. (3.1) are at best inefficient and at worst will yield invalid and misleading inference.

In Chapter 2 we distinguish between tracking and return matching as competing alternatives, based on the presence of a cointegrating relationship between the ETF and the benchmark. Hence, the key point is to test whether the ETF and the benchmark are I (1) processes and share a common stochastic trend. In the presence of a cointegrating relationship between fund and the benchmark, an ECM as in Eq. (3.2) should be employed to compute the TE.

$$\Delta F_{m,t} = \kappa_0 + \kappa_1 \Delta I_{n,t} + \gamma_1 u_{mn,t-1} + v_{mn,t} \tag{3.2}$$

Where $u_{mn,t-1}$ is the equilibrium error from the cointegrating regression between the levels of the nth ETF and the mth benchmark, $v_{mn,t}$ is a white noise error term, and γ_1 measures the rate of return to equilibrium.

3.4. Data description and preliminary tests

The sample employed in this chapter, in terms of ETFs, coincides with that of Chapter 2. In short, our sample contains 77 passively managed ETFs listed on the NYSE. While 59 ETFs track equity indices, the rest mirror debt indices. The research question in this chapter requires the construction of some liquidity proxies. However, the bid and ask prices available on Bloomberg and DataStream exhibit some glitches between 2009-2011 for some ETFs from our sample. Therefore, to improve the quality of the data and so the inferences, we obtain further data for the same ETFs. The quality of the bid-ask spreads seems to improve from the end of 2011, hence closing prices and the rest of the variables for the 77 ETFs were collected, at the daily frequency, from Bloomberg for the period 1st of January 2012 to 17th of May 2017. Following the reasoning in Chapter 2, closing prices, rather than NAVs, are employed. The name of the ETF, the Bloomberg ticker and the inception date are displayed in the Appendix of Chapter 2 (Table A.2.1 and Table A.2.2, respectively).

[Table 3.1]

The time series properties of the log prices of the ETFs and their benchmarks have been studied in the Chapter 2. The outcomes of this chapter confirm the non-stationary behaviour of the log prices for ETFs and benchmarks, even when the data is demeaned to mitigate the potential cross-correlation among the elements of the panel or when such cross-correlation is allowed for. Table 3.1 shows two different panel unit root tests. Panel A includes Hadri (2000), which is a panel test for the null of stationarity and Breitung (2001) which allows for a homogeneous unit root under the alternative hypothesis. We formally test for the presence of a cointegrating relationship between the ETFs and their benchmarks using only one approach, since the conclusions of the Chapter 2 remained the same regardless of the cointegration test used. The approach employed is based on panel tests for the null of cointegration following Pedroni (1999). Pedroni's approach employs four tests based on the within panel dimension, a non-parametric extension of the variance ratio statistic proposed by Phillips and Ouliaris (1990), which we denote as panel v-stat in Table 3.1, a panel version of the Phillips and Perron (1988) and Phillips and Ouliaris (1990) unit root tests, which appears in Table 3.1 as panel pp-rho stat, a semi-parametric adjustment of Phillips and Perron (1988) t-test statistic, which is represented by panel pp-stat in Table 3.1, and a panel form of the augmented Dickey Fuller t-statistic (Fuller 1976), which appear in Table 3.1 as panel aDF-test. Pedroni (1999) also employs three tests based on the between panel dimension which can be seen as the group mean version of the within dimension statistics. Pedroni (1999) shows that the asymptotic distributions of these seven statistics follow a standard normal. The evidence shows that the log prices of the ETFs and the indices behave as an I (1) processes and appear to share a common stochastic trend. This result corroborates that findings of Chapter 2 and are robust to the change in the sample period.

Since the ETFs and their benchmark indices appear to be I (1) and share a common stochastic trend, we employ Fully Modified OLS (FMOLS) and Dynamic OLS (DOLS), following Pedroni (2001, 2004), to estimate the cointegrating relationship between the ETF and the underlying index in levels. The panel results shown in Panel C corroborate the findings of Chapter 2. Hence, we can conclude that weak tracking defines the equilibria for the ETFs in the sample. Following the conclusions of Chapter 2, and to avoid misleading and inefficient estimates, an ECM will be used to compute the TE.

To study the determinants of the ETFs' TEs, we construct a series of liquidity proxies that consider the special structure of the ETF. We gather the data required to to construct the liquidity proxies from the ETFs' factsheets and prospectuses, available in the issuers' websites, and the Bloomberg terminal. The ETFs' factsheets and prospectuses generally contain information about the size of the creation units, the total expense ratio, the replication strategy, the fund engagement into security lending and the duration of the bond portfolio held by the ETF. To construct the measures of ETFs' primary and secondary liquidity, we require the ETFs' NAV, fund flows, turnover, relative bid-ask spread, volume traded, which we gather from the Bloomberg terminal. The data is collected daily between the 1st of January 2012 and the 17th of May 2017. The ETFs' daily return and volatility are calculated using the closing prices from Bloomberg.

3.5. Tracking errors

The main conclusion of Chapter 2 is that, to avoid misleading estimates of the TEs and suboptimal portfolio choices, an ECM needs to be employed to calculate the TEs.

Despite the existence of numerous definitions of the TEs in the literature (see Roll 1992, Pope and Yadav 1994, and Rudolf, Wolter and Zimmermann 1999, among others) there are three approaches that have been extensively used by the existing literature for passively managed ETFs (see Rompotis 2009; Milonas and Rompotis 2010, Buetow and Henderson 2012; Bertone et al. 2015). Our first measure of the TE, $\xi_{1,mn}$, is the standard error of the residuals from OLS estimation of the ECM. The second measure, $\xi_{2,mn}$, is defined as the average absolute deviation between the log returns of the fund and those of the benchmark:

$$\xi_{2,mn} = \frac{\sum_{t=1}^{T} \left| e_{mn,t} \right|}{T} \tag{3.3}$$

Where $e_{mn,t} = \Delta F_{m,t} - \Delta I_{n,t}$ and T is the sample size.

The third approach, $\xi_{3,mn}$ computes the TE as follows:

$$\xi_{3,mn} = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} \left(e_{mn,t} - \bar{e} \right)^2} \tag{3.4}$$

Where, $e_{mn,t} = \Delta F_{m,t} - \Delta I_{n,t}$ and $\bar{e} = T^{-1} \sum_{t=1}^{T} e_{mn,t}$. T is the sample size.

Chapter 2 points out that $\xi_{3,mn}$ coincides with the definition of the TE used in the ETFs' prospectuses, which is computed as the standard deviation of the expected return of the long-short portfolio when the optimal hedge ratio equals unity ($\lambda^{*}=1$).

[Table 3.2]

Table 3.2 displays the summary statistics of the TE computed using the ECM. While Panel A displays the TE for equity funds, Panel B shows those of debt funds. Broadly speaking, the evidence shows that the TEs vary widely among funds. In line with the outcomes of Chapter 2, equity funds seem to track the underlying more accurately than debt ETFs. Moreover, we observe that equity ETFs provide the maximum value of the TE in the sample while debt funds exhibit the lowest TE. In terms of TE dispersion, equity ETFs provide higher dispersion than debt ETFs. Comparison of the average TE of equity and debt ETFs, suggests that the latter provide higher average TEs than the former (around 18%). The reason underlying this difference could be due to the replication strategies employed by the fund, the fund's management expenses, the usage of security lending strategies…etc.

Buetow and Henderson (2012) links the liquidity of the underlying assets and the magnitude of the TE, suggesting that there exists an indirect relationship between the liquidity of the securities tracked by the fund and the TE. In other words, the less liquid the underlying securities, the less accurate the fund's tracking performance. This finding is consistent with our results, shown in Table 3.2, since the maximum value of the TE for the overall sample is provided by an ETF that tracks an index that contains small-caps.

Analysis of the equity funds individually shows that the maximum value of the TE for equity ETFs is 0.010 while the minimum is 0.0005. The lowest TE, irrespective of the metric, is provided by iShares S&P Mid-Cap 400 Growth ETF (Bloomberg Ticker: IJK US). This fund mirrors the performance of the S&P Mid-Cap Growth index and hence, provides exposure to a range of U.S. mid-size companies that are expected to growth at a higher average rate than the market. Conversely, the maximum TE for equity ETFs is provided SPDR S&P 600 Small Cap Growth (Bloomberg Ticker: SLYG), which provides exposure to an index composed of small-capitalization U.S. equities that exhibit growth characteristics. Comparison of the TE of funds that mirror large and mid-caps with the average TE for all equity funds, shows that these funds provide a TEs 19% smaller than average. Conversely, funds that track small-caps, micro-caps and mixed-caps result in a TE which is 37% greater than the average TEs for equities. These findings suggest that indices that include large-caps seem easier to track than those containing small-caps.

Summary statistics of TE for bond ETFs are analogous. Individual analysis of bond ETFs shows that, regardless of the method and the metric used compute the TE, Vanguard Extended Duration Treasury ETF (Bloomberg Ticker: EDV US) exhibits the highest TE. This ETF targets the performance of the Bloomberg Barclays U.S. Treasury STRIPS 20–30 Year Equal Par Bond Index. Conversely, the lowest TE is delivered by iShares Short Treasury Bond ETF (Bloomberg Ticker: SHV US), which replicates the performance of ICE U.S. Treasury Short Bond Index and helps investors to gain exposure to the U.S. Treasury bonds that mature in less than 12 moths. It is worth noting on passing, that both ETFs belong to the Treasury Bonds category. A more comprehensive analysis shows that the TEs of corporate bond ETFs, regardless of the TE metric used, are 68% greater than the average TE of debt funds. These findings are analogous to those of Buetow and Henderson (2012), since the liquidity of the underlying securities seem to be related to the magnitude of the

TE. The differences in performance between bond ETFs and their benchmarks might be due to the replication strategy employed by the ETFs, the engagement in security lending or the expense ratios charged by the fund. We observe that SHV US ETF and EDV US ETF have the minimum and the maximum duration, respectively, which probably suggests that the magnitude of the TE can be related to the duration of the bond portfolio.

The outcomes show that the TEs vary widely across funds. Comparison of the average TEs of debt and equity ETFs suggests that equity ETF track their underlying benchmarks more accurately than their debt counterparts. The results support the idea that the tracking ability of equity ETFs might be closely related to the illiquidity of the underlying securities, as funds mirroring indices that include small-caps, micro-caps and mixed-caps provide larger TEs than the funds tracking large-caps and mid-caps. In terms of debt funds, the duration of the bond portfolio is likely to be one of the key determinants of the debt ETFs' tracking performance.

3.6. Analysis of the factors underlying the tracking errors

In this section, once we have properly computed the TE accounting for the stochastic nature of the data, the main question that we try to answer is which factors underlie the cross-sectional differences on the TEs observed. To answer this question, we divide the section into two subsections. In the first subsection, we define the different measures of liquidity used to approximate ETFs' primary and secondary liquidity, and we introduce a series of factors that might also have an impact on the ETFs' tracking performance. In the second subsection, we perform a cross-sectional analysis to study which are the main factors explaining the tracking ability of the

ETFs in the sample. Furthermore, we perform an additional cross-sectional analysis to identify whether the determinants of the ETFs' tracking performance depend on asset class tracked by the ETF.

3.6.1. Measures of liquidity

ETFs constitute a mixture of a mutual fund and a stock and combine the creation-redemption processes of the former with the continuous trading in secondary markets of the latter. To accommodate this special nature of the ETF, we distinguish between primary liquidity, which relates to ETF's creation-redemption processes, and secondary liquidity, which is linked to the ETF's trading activity.

The ETF creation process requires an AP to deposit the securities included in the benchmark to the ETF's provider in return of ETF's shares. Conversely, the redemption process takes place when the AP exchanges ETF's shares for the underlying securities. For highly liquid securities, the ETF's creation-redemption processes are straightforward. However, when the fund tracks highly illiquid securities, the AP may struggle to access these securities and, as a result, premiums and discounts arise and may persist over time. According to that, the creation-redemption processes are key to align the ETF's prices and the NAV of the underlying securities. Ackert and Tian (2008) argues that in a market without limits to arbitrage the ETF's price and the NAV should be always aligned. Any departure of the ETF's price from the NAV of the underlying securites would present an arbitrage opportunity. Since ETFs are traded like stocks on exchanges, the secondary liquidity relates to the ETF's trading activity. Secondary liquidity is intended to capture the effects of wrapping illiquid and over-the-counter fixed income securities and more liquid equities in the form of transparent and liquid ETFs. We construct a series proxies to estimate the effects of primary and secondary liquidity on the ETFs' tracking ability.

To approximate the overall effect of primary liquidity on tracking performance, we employ four proxies. The first, denominated PD, represents the value of the premiums/discounts that arise when a misalignment occurs between the price of the ETF and the NAV of the underlying securities. In line with Charupat and Miu (2011) we calculate the PD as follows:

$$PD_t = \frac{f_t - LNAV_t}{LNAV_t} * 100 \tag{3.5}$$

Where PD_t represents the relative Premium/Discount for day t. Moreover, f_t and $LNAV_t$ represent the logarithms of closing price of the fund and the logarithm of the NAV of the underlying securities on day t, respectively.

The second proxy, denoted CDF, is the illiquidity measure developed by Chacko, Das and Fan (2016). CDF is based on a portfolio that combines a long position of the ETF with a short position in the underlying securities. To the best of our knowledge this proxy has not been used before to assess the ETF's tracking performance. CDF can be seen as an indirect measure of the transaction costs of buying/selling the underlying securities and the execution risk underlying the creation and redemption processes. We compute CDF as follows:

$$CDF_t = -10000 * \log\left[\frac{1}{1 + \left[\frac{F_t}{NAV_t} - 1\right]}\right] \tag{3.6}$$

Where log represents the natural logarithm, F_t represents the ETF's closing price on day t, and NAV_t represents the NAV of the underlying securities on day t. Chacko, Das and Fan (2016) argue that the greater the difference between the ETF's closing price and the NAV, the greater the value of CDF and the less liquid the fund.

Prior to the commencement of trading on the stock exchange, the AP need to create the shares in the ETF. A creation unit represents the minimum number of shares required by the ETF's provider for the creation-redemption processes. The size of the creation unit is specified in the ETF's prospectus. When some illiquid ETFs, such funds mirroring corporate bonds or micro-caps, need to create ETF's shares the process of accumulating the required securities to form a creation unit in a timely fashion can be very complicated and, therefore, PDs may arise and persist until the creation unit is completed. In line with this, Acker and Tian (2008) report a persisting misalignment between ETF's prices and NAVs in a sample of country ETFs, which is related to the inherent complexity of the creation-redemption processes of these funds. Conversely, Elton, Gruber, Comer and Li (2002) report very small and short-lived differences between the price and the NAV of the SPY ETF, which one of the most liquid ETFs in the market, due to well-functioning arbitrage. To measure the extent to which the size of the creation unit proves to be a barrier to the arbitrage activities for illiquid ETFs and hence, has an impact on the tracking ability of the fund, we include CreatSize. This variable is equal to the size of the creation unit for each ETF. The fourth variable, Activity, uses daily data of ETF's flows to measure the total number of days within our sample period where creation and redemption activities occur for each ETF. A quick look to Table 3.5 shows that, while most of the ETFs that track blue chips have creation-redemption activities almost every day, ETFs tracking corporate bonds or small-caps have substantially less activity.

In terms of the secondary liquidity, we include four well-known factors that have been previously used the liquidity literature (Amihud and Mendelson 1986; Chordia, Roll, and Subrahmanyam 2001). Following Buetow and Henderson (2012), the first measure is Turnover. We compute the ETF's turnover as follows:

$$Turnover_t = \frac{1}{N_t}\sum_{t=1}^{Nt}\frac{Vol_t}{SO_t} \tag{3.7}$$

Where Vol_t represents the ETF trading volume on day t, and SO_t is the sum of the ETF's shares outstanding at time t. N_t is the number of trading days in the period under study. The authors conclude that the higher the dollar-volume traded, the more accurate the tracking performance, especially for equity and debt ETFs.

Amihud and Mendelson (1986) suggests the usage of bid-ask spreads as a measure of stock's liquidity. Therefore, following Bertone, Paeglis and Ravi (2015) and Broman (2016), we employ the ETF's relative bid-ask spread, denominated Spread, as a proxy of ETF's secondary liquidity. The spread is a measure of the transaction costs incurred when trading ETFs in a secondary market. Narrow spreads may reflect a high degree of liquidity whereas wider spreads are more likely to characterise illiquid securities.

The third measure, Amihud, is the illiquidity measure introduced by Amihud (2002), which represents the price change per dollar volume of trading and it is computed as follows:

$$Amihud_t = \frac{1}{N_t}\sum_{t=1}^{Nt}\frac{r_{F,t}}{Dvol_t} \tag{3.8}$$

Where $r_{F,t}$ represents the continuously compounded return of the ETF at time t, and $Dvol_t$ the dollar trading volume on day t. N_t is the number of trading days in the period under study. This proxy of the trade impact has been previously used in ETF's literature by Broman and Shum (2018) and Broman (2016), but not as a potential determinant of the TE. However, we include it in our analysis as it might capture a different dimension of the secondary market liquidity.

The fourth measure, following Bertone, Paeglis and Ravi (2015), is the risk of the fund, RiskF. We compute RiskF as the volatility of the continuously compounded returns of the ETFs.

Moreover, the prior literature relates TEs to the expense ratios charged by the fund, to the replication strategies employed to replicate the benchmark, and to the fund engagement into security lending. All these factors can be seen as indirect proxies of ETF's liquidity. For instance, higher expense ratios might reflect the difficulties experienced by the fund manager while tracking an illiquid benchmark index and the concomitant exposure to liquidity risk. In fact, the more illiquid the underlying securities the greater the trading costs and risks associated and consequently the less accurate the tracking, all else equal. Similar reasoning applies to the replication strategies and the security lending, which motivates the inclusion of these three control variables (TER, Replic and Lend) in the analysis. The ETFs' Total Expense Ratio (TER) summarises in one figure a variety of costs incurred in the ETF's management. Although all the ETFs in our sample replicate the benchmark physically, some funds replicate it using optimised samples, while others use full replication. The findings of Blume and Edelen (2004) indicate that any departure from full replication might result in less accurate tracking. Replic, is a categorical variable that takes value one when the ETF uses optimised sample and zero otherwise. Finally, since some ETFs in our sample engage into security lending with to enhance returns, we include a variable named Lend which is equal to one when the fund engages into security lending and zero otherwise.

[Table 3.3]

Table 3.3 provides a summary of the main measures of liquidity used in the cross-sectional analysis. The numerical variables are expressed in terms of their average.

3.6.2. Cross-sectional analysis

The cross-sectional analysis is divided in two parts. In the first part we study the impact on the TEs of the following variables: the PD, the illiquidity measure of Chacko, Das and Fan (2016), the creation size, the fund turnover, the relative bid-ask spreads, the Amihud illiquidity measure, the risk of the fund, the duration of the bond portfolio, the expense ratio charged by the fund, the replication strategy and the ETFs' engagement into security lending. In the second part of the analysis, we look at whether the determinants of the TEs depend on the asset class tracked by the fund. Following Drenovak, Urošević and Jelic (2014), in this part of the analysis we include the variable Duration, which takes the value of the duration of the ETF's bond portfolio measured in years if the fund tracks debt and zero otherwise. In theory, the longer the duration the more volatile the fund and the less accurate the tracking.

[Table 3.4]

Table 3.4 displays the correlation between the explanatory variables included in the first part of the cross-sectional analysis. The highest positive correlation is exhibited by the pair Activity and Turnover. Conversely, the highest negative correlation coefficient is displayed by the pair Turnover and Spread.

In the second part of the cross-sectional analysis, to assess wheter the impact of liquidity on the tracking performance depends on the asset class tracked by the fund, we create seven interaction dummies and perform the analysis again. Those interaction dummies ending with the letter E (B) will take the value of a particular variable when the ETF tracks equity (debt) and zero otherwise.

[Table 3.5]

Table 3.5 presents the correlation coefficients of the explanatory variables included in the second part of the cross-sectional analysis. The pair Activity-B and Turnover-B, displays the highest positive correlation coefficient. The variables Creation Size-E and Turnover-B display the highest negative correlation coefficient.

[Table 3.6]

Table 3.6 exhibits the results of the first part of the cross-sectional analysis. The evidence suggests that the main determinants of TEs for the three measures are very similar. The variables CDF, Activity, Turnover, RiskF and Spread appear statistically significant at 10% level.

The CDF illiquidity measure is positively related to the TE observed. This result illustrates the difficulties faced by the APs in arbitraging any disparities between the ETF's price and the NAV of the underlying securities, and their effect on the fund's performance. Activity is positively and statistically significantly related to TEs, which suggests that the costs generated by intense creation-redemption activities might deteriorate the ETF's tracking performance.

Turning to the proxies for secondary liquidity, we observe a negative relationship between the fund's turnover and the TEs. Finally, the bid-ask spread and the fund's volatility appear to be positively related to the TEs at 10% level. In other words, the wider the spreads and the higher the volatility, the less accurate the tracking performance and consequently the greater the TEs. The rest of the measures of indirect liquidity are not significant at the 10% level of confidence.

[Table 3.7]

Table 3.7 reports the coefficient estimates of the second part of the cross-section analysis. The results are analogous to those obtained in the first stage. In terms of primary liquidity, we observe that CDF-E and CDF-B are both positive and statistically significant. Comparison of the coefficient estimates suggests that the coefficient for equity funds is greater than that of debt funds. This finding might seem puzzling since, in general, illiquid and over-the-counter fixed income securities are more difficult to track than equities. However, the difference observed may be related to the inclusion of equity funds tracking benchmark indices that include small-caps and micro-caps in the sample, or perhaps to the fact that the sample contains several debt ETFs tracking government bonds. It may also be worth noting, that the funds tracking corporate bonds in the sample generally use optimised strategies to replicate a benchmark, which usually only include highly liquid bonds.

The coefficient estimates of the Activity-E and Activity-B variables, regardless of the metric employed, are positive and significant at 10% level, reflecting how the costs of issuing or redeeming ETFs' shares can increase the magnitude of the TE. Moreover, the fund turnover appears negatively related to the TE, regardless of whether the fund tracks equity or debt securities. Finally, the results show that the duration of the ETF's bond portfolio is positive and significantly related to the TE at 10% level of confidence. Since bonds with longer durations are very volatile and often quite illiquid, the fact that duration is positively related to the TE is consistent with the view that the greater the illiquidity of the bond portfolio, the harder it is to track the benchmark.

Overall, the outcomes demonstrate that the illiquidity resulting from the creation-redemption processes plays a significant role in determining the tracking quality of ETFs. This suggests that the difficulties experienced by the AP during the creation-redemption processes impedes the tracking process and the TEs increase. In terms of secondary liquidity, the evidence indicates that fund's turnover has a negative impact on the tracking ability of ETFs. In other words, the accuracy of the tracking seems to improve as the fund turnover increases. The evidence is consistent with the view that ETF's volatility is significantly related to the quality of the tracking, which in turn suggests that volatile funds are associated with larger TEs.

When the interaction dummies are added to the cross-sectional analysis, the results regarding the impact of primary liquidity remain qualitatively unchanged. The illiquidity originated in the creation-redemption processes remains an important determinant of the funds' tracking performance, independently on whether the fund tracks equity or debt securities. In contrast, when we distinguish between funds tracking equity and debt and include the duration of the bond portfolio, the only measure of secondary liquidity that remains significant, at 10% level, is the fund's turnover.

3.7. Conclusion and discussion

This paper examines the main determinants of the TEs using a sample of passively managed bond and equity ETFs that trade in the NYSE over the period 4th of January 2012 to 17^{h} May 2017. Given the growth in passively managed index funds in the last few years, our paper has implications for institutional and retail investors who purchase ETFs to benefit from the diversification, transparency, intraday liquidity and low trading costs provided by these funds.

Analysis of the nature of the data shows that the prices of ETFs and benchmarks appear to be non-stationary but cointegrated. Although the ETFs and the benchmarks share a common stochastic trend, in the short-term TEs may arise. We obtain three different metrics of the TE using an ECM, as recommended in Chapter 2. Empirical evidence shows that equity ETFs displays more accurate tracking than their debt counterparts.

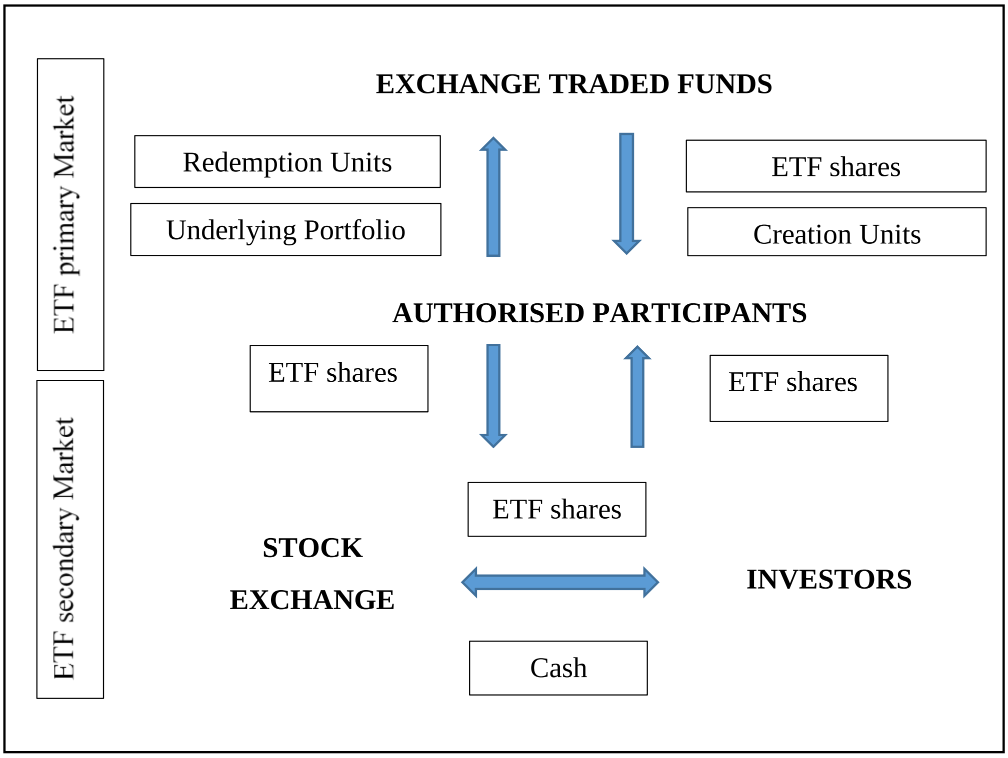
Overall, the results provide evidence of the important role played by ETF's liquidity on the ETFs' tracking ability. Since liquidity is an abstract concept, to capture the different aspects of it, we divide the total ETF's liquidity into ETF's primary liquidity, which arises during the creation and redemption processes, and ETF's secondary liquidity, related to the ETF's market activity. Additionally, we identify a series of factors (such as the expense ratios charged by the fund and the replication strategy of the fund, to name a few) which might represent indirect measures of ETF's liquidity.

The results from the cross-sectional analyses support the view that the illiquidity resulting from the creation and redemption processes, which has been overlooked by the prior literature, plays a significant role in determining the tracking performance of ETFs. Essentially, these outcomes illustrate the difficulties experienced by the APs during the creation and redemption of ETFs' shares, which eventually might deteriorate the ETFs' tracking performance. The results hold for equity and debt ETFs, independently of the metric used to quantify the TEs. In terms of secondary liquidity, the evidence demonstrates that the ETF's tracking ability is directly related to the ETF's turnover. In other words, the ETF's tracking performance improves as ETF's turnover increases. Moreover, the bid-ask spreads and the fund volatility appear to be positively related to the TEs.

The addition of interaction dummies to the cross-sectional analysis, to test whether the determinants of the TEs depend on the asset class tracked by the fund, yields comparable results. The illiquidity originated in the creation and redemption processes remains a key determinant of the funds' tracking performance, independently on whether the fund tracks equity or debt securities. In terms of the secondary liquidity, the fund turnover appears to be the main driver of the TEs, regardless of the asset class contained in the fund portfolio. Moreover, the duration of the bond portfolio held by the fund seems to play an important role in determining the tracking performance of the debt ETFs.

Figures

Figure 3.1. The creation-redemption mechanism.



Tables

Table 3.1. Panel unit root and cointegration tests

Panel A: Unit root tests		
Raw data	H	B
ETF	10620[a]	11.53
Index	13082[a]	1.63

Panel B: Panel cointegration tests			
Time Dummies		No Time Dummies	
panel v-stat	-2.79[a]	panel v-stat	-3.67[a]
panel rho-stat	-6.03[a]	panel rho-stat	-7.02[a]
panel pp-stat	-4.70[a]	panel pp-stat	-5.16[a]
panel aDF-stat	-3.67[a]	panel aDF-stat	-5.47[a]
group rho-stat	-87.17[a]	group rho-stat	-12.28[a]
group pp-stat	-28.55[a]	group pp-stat	-7.49[a]
group aDF-stat	-21.29[a]	group aDF-stat	-7.70[a]

Panel C: Fully modified and dynamic least squares panel estimates				
	Time Dummies		No Time Dummies	
	Coefficient	t-stat	Coefficient	t-stat
FMOLS	1.038	59.460	0.705	-410.641
DOLS	1.037	59.930	0.705	-410.537

Table 3.1: Panel A presents unit root test results for the following tests: Hadri (2000) H; Breitung (2001) B. The Breitung test in Panel A has the null hypothesis that the data contain a unit root while the Hadri test has as the null hypothesis that the panel is stationary. All the tests are asymptotically distributed as N (0,1). Panel B presents tests of the null hypothesis of no cointegration following Pedroni (1999, 2004). All the cointegration tests are distributed as N (0,1). Panel C displays the panel FMOLS and DOLS estimates. [a] Significant at 1% level, [b] Significant at 5% level, and [c] Significant at 10% level.

Table 3.2: Tracking errors

Panel A: Descriptive statistics				
Funds		ξ_1	ξ_2	ξ_3
Equity	Mean	0.001726	0.001113	0.001725
	Standard deviation	0.001621	0.001228	0.001620
	Maximum	0.010180	0.007645	0.010170
	Minimum	0.000523	0.000347	0.000523
Debt	Mean	0.001985	0.001369	0.001984
	Standard deviation	0.001354	0.000951	0.001353
	Maximum	0.004960	0.003336	0.004950
	Minimum	0.000119	0.000079	0.000119

Table 3.2: Tracking errors summary statistics. ξ_1 is the TE defined as the standard error of the ECM regression. ξ_2 measures the extent to which the continuously compounded returns of the ETF diverge from those of the benchmark index using the ECM regression. ξ_3 represents the standard deviation of the difference in continuously compounded returns between the ETF and underlying index using ECM.

Table 3.3. Summary of the liquidity proxies

Ticker	PD	CDF	CreatSize	Activiy	Spread	Turnover	Amihud	RiskF	TER	Dur	Lend	Replic	Equity
SPY US	-0.0003	2.2556	50000	1252	0.0056	23.7226	0.0005	0.0082	0.09	0	N	F	Y
MDY US	0.0021	3.4835	25000	932	0.0164	19.8638	0.0170	0.0095	0.25	0	Y	F	Y
DIA US	-0.0007	2.3853	50000	1066	0.0074	20.5175	0.0108	0.0078	0.17	0	N	F	Y
IUSG US	0.0030	5.1473	50000	167	0.1269	14.8196	5.9324	0.0085	0.05	0	Y	F	Y
IVV US	0.0016	2.7694	50000	984	0.0097	20.3343	0.0130	0.0082	0.04	0	Y	F	Y
IWB US	-0.0012	2.6182	50000	486	0.0177	18.1394	0.1559	0.0083	0.15	0	Y	F	Y
IJH US	0.0092	3.2775	50000	807	0.0208	18.6216	0.0659	0.0094	0.07	0	Y	F	Y
IJR US	0.0036	3.8205	50000	735	0.0324	18.3728	0.1000	0.0103	0.07	0	Y	F	Y
IVE US	-0.0019	2.4175	50000	475	0.0202	17.7417	0.1849	0.0084	0.18	0	Y	F	Y
IVW US	-0.0008	2.4747	50000	541	0.0207	17.9648	0.1689	0.0083	0.18	0	Y	O	Y
IWD US	0.0036	2.8821	50000	714	0.0132	18.8798	0.0550	0.0084	0.20	0	Y	F	Y
IWF US	0.0005	2.7851	50000	790	0.0145	18.8713	0.0596	0.0084	0.20	0	Y	O	Y
IWM US	-0.0021	4.5000	50000	1235	0.0098	22.0448	0.0027	0.0107	0.20	0	Y	F	Y
IWV US	0.0010	2.6942	50000	244	0.0410	16.9721	0.3407	0.0084	0.20	0	Y	F	Y
IYY US	-0.0025	4.7165	50000	76	0.1058	14.7046	3.3233	0.0082	0.20	0	Y	F	Y
IJJ US	0.0041	3.0697	50000	352	0.0623	16.3799	0.6094	0.0096	0.25	0	Y	O	Y
IJK US	0.0003	2.4208	50000	423	0.0597	16.6671	0.5025	0.0095	0.25	0	Y	F	Y
IJS US	0.0033	4.1062	50000	353	0.0810	16.3798	0.7219	0.0104	0.25	0	Y	F	Y
IJT US	-0.0020	3.9786	50000	383	0.0901	16.4177	0.7767	0.0104	0.25	0	Y	F	Y
IWN US	-0.0021	4.5154	50000	520	0.0233	18.4803	0.0881	0.0102	0.25	0	Y	F	Y
IWO US	-0.0021	4.3801	50000	666	0.0411	18.5708	0.1140	0.0115	0.25	0	Y	O	Y
IUSV US	0.0063	5.8436	50000	147	0.1237	14.7230	6.9646	0.0083	0.05	0	Y	O	Y
SLYG US	0.0916	83.1707	50000	87	0.4673	14.2298	12.8752	0.0105	0.15	0	Y	F	Y
SLYV US	0.0250	10.4714	50000	97	0.7158	14.0832	14.4762	0.0108	0.15	0	Y	F	Y
SPYG US	0.0128	7.2134	50000	107	0.2608	14.2055	10.1742	0.0083	0.15	0	Y	F	Y
SPYV US	0.0138	11.1960	50000	59	0.7059	13.2588	22.0816	0.0083	0.15	0	Y	F	Y
OEF US	-0.0011	2.6481	50000	711	0.0200	17.8416	0.2017	0.0081	0.20	0	Y	F	Y
IWP US	0.0036	2.7349	50000	371	0.0439	17.0011	0.3257	0.0092	0.25	0	Y	F	Y
IWR US	0.0070	2.9514	50000	383	0.0506	17.3535	0.2407	0.0089	0.20	0	Y	O	Y
IWS US	0.0052	3.0377	50000	505	0.0343	17.2689	0.3260	0.0088	0.25	0	Y	O	Y
RSP US	0.0042	2.7848	50000	492	0.0238	17.7694	0.2147	0.0088	0.40	0	N	F	Y
DVY US	0.0001	2.6299	50000	801	0.0238	17.9572	0.1032	0.0074	0.39	0	Y	F	Y
ITOT US	0.0188	4.4581	50000	324	0.0911	15.8486	2.0894	0.0083	0.03	0	Y	F	Y
PFM US	-0.0013	7.0751	50000	93	0.1683	13.5332	10.1019	0.0071	0.55	0	N	F	Y
PXMG US	-0.0554	15.9062	50000	39	0.3539	11.5294	100.0293	0.0097	0.39	0	N	F	Y
PXMV US	0.0007	14.3780	50000	44	0.2965	11.6508	47.5986	0.0093	0.39	0	Y	F	Y
PXSV US	-0.0334	14.4356	50000	51	0.4075	12.1389	72.5286	0.0109	0.39	0	Y	F	Y
XLG US	-0.0058	4.9426	50000	118	0.0931	14.6287	4.6357	0.0080	0.20	0	N	F	Y
IWC US	-0.0122	9.3969	50000	133	0.2624	15.3565	2.3072	0.0111	0.60	0	Y	F	Y
PEY US	0.0107	6.5456	50000	320	0.1246	14.5438	3.7201	0.0077	0.54	0	Y	F	Y
FDM US	-0.0211	16.9733	50000	44	0.3577	12.0786	121.0014	0.0108	0.60	0	N	O	Y
MDYG US	0.0322	12.9773	50000	71	0.6759	13.4067	52.9058	0.0094	0.15	0	N	F	Y
SLY US	0.0542	12.1623	50000	76	0.4949	14.0207	13.3764	0.0105	0.15	0	Y	F	Y

Ticker	PD	CDF	CreatSize	Activiy	Spread	Turnover	Amihud	RiskF	TER	Dur	Lend	Replic	Equity
SPHQ US	0.0302	6.8087	50000	280	0.1858	14.3041	10.6259	0.0074	0.29	0	N	F	Y
PRF US	0.0105	3.3358	50000	390	0.0818	16.1368	1.2376	0.0085	0.39	0	Y	F	Y
RFG US	-0.0090	5.8698	50000	136	0.1831	14.6717	2.8983	0.0104	0.35	0	N	F	Y
RFV US	0.0154	11.9605	50000	56	0.2353	12.7874	74.2602	0.0113	0.35	0	N	F	Y
RPG US	0.0183	4.6049	50000	269	0.0884	15.6295	3.1538	0.0096	0.35	0	N	F	Y
RPV US	0.0147	5.7988	50000	285	0.1319	15.2289	-1.2725	0.0103	0.35	0	N	F	Y
RZG US	-0.0078	12.9993	50000	98	0.2734	13.3814	32.5983	0.0113	0.35	0	N	F	Y
RZV US	-0.0050	13.0232	50000	59	0.2702	13.4417	15.4151	0.0131	0.35	0	N	F	Y
VIG US	0.0071	2.7404	25000	876	0.0217	17.9573	0.0979	0.0075	0.10	0	Y	F	Y
PKW US	0.0177	5.3819	50000	514	0.0987	15.7310	3.0873	0.0087	0.63	0	Y	F	Y
IWY US	0.0074	5.8417	50000	135	0.1237	13.9669	8.5769	0.0083	0.20	0	Y	F	Y
SCHA US	0.0229	5.2809	50000	400	0.1120	16.2704	1.0733	0.0101	0.05	0	Y	F	Y
SCHB US	0.0186	3.7136	50000	493	0.0671	16.7545	0.5388	0.0083	0.03	0	Y	O	Y
SCHX US	0.0192	3.6391	50000	619	0.0614	16.5970	0.7509	0.0082	0.03	0	Y	F	Y
SCHG US	0.0219	4.0544	50000	410	0.1227	15.7633	1.9382	0.0087	0.04	0	Y	F	Y
SCHV US	0.0254	4.2584	50000	410	0.1382	15.4718	1.7500	0.0079	0.04	0	Y	F	Y
LQD US	0.1756	21.9512	100000	1014	0.0143	19.4205	0.0101	0.0035	0.15	8.5	Y	O	N
IEF US	0.0180	5.5699	100000	950	0.0269	18.6420	0.0143	0.0038	0.15	7.63	Y	O	N
CSJ US	0.0960	10.2743	50000	612	0.1404	17.7464	0.0023	0.0008	0.20	1.96	Y	O	N
CRED US	0.1678	21.3805	50000	176	0.2052	15.5427	0.3702	0.0028	0.15	7.07	Y	O	N
IEI US	0.0181	3.1736	100000	508	0.1338	17.2736	0.0255	0.0020	0.15	4.5	Y	O	N
SHV US	0.0099	1.1308	100000	434	0.1337	17.6212	0.0013	0.0001	0.15	0.44	Y	O	N
TLH US	0.0174	7.6365	100000	189	0.1775	15.3988	1.0147	0.0049	0.15	10.2	Y	O	N
SHY US	0.0118	1.6384	100000	773	0.0426	18.3174	0.0016	0.0005	0.15	1.84	Y	O	N
HYG US	0.3111	38.2156	100000	831	0.0128	20.0043	0.0040	0.0044	0.50	4.1	Y	O	N
MUB US	0.0981	29.3155	100000	390	0.0534	17.2205	0.0532	0.0030	0.25	4.79	Y	O	N
PLW US	0.0134	8.9096	50000	145	0.3193	13.1446	17.2649	0.0050	0.25	11.3	N	F	N
PHB US	-0.0599	20.6662	100000	593	0.1310	15.7916	0.1120	0.0035	0.50	4.07	Y	F	N
JNK US	0.2288	29.3679	500000	1026	0.0265	19.3112	0.0077	0.0043	0.40	4.4	Y	F	N
EDV US	0.4542	46.1325	50000	157	0.2959	15.3469	1.0470	0.0127	0.07	24.8	N	F	N
SUB US	0.0786	13.5584	50000	160	0.2523	15.2209	0.0563	0.0012	0.25	1.97	Y	F	N
HYD US	-0.1993	42.7692	100000	421	0.0920	16.2253	0.2766	0.0047	0.35	6.97	Y	F	N
LWC US	0.4204	49.4583	100000	132	0.4317	13.9212	6.5258	0.0061	0.12	14.2	Y	F	N
TUZ US	-0.0090	2.9311	50000	72	0.4998	13.2765	-1.7851	0.0008	0.15	1.85	N	O	N

Table 3.3: Summary statistics and characteristics of the determinants. The variable "PD" is the ETF's premium/discount; "CDF" is the illiquidity developed by Chacko, Das and Fan (2016); "CreatSize" is the unit creation size for equity and fixed income ETFs; "Activity" ETF's creation-redemption activity in days; "Spread" the relative bid-ask spread; "Turnover" represents the ETF's turnover; "Amihud" is Amihud illiquidity measure; "RiskF" is volatility of the fund; "TER" is the ETF's total expense ratio; "Dur" is the duration of the bond portfolio; "Lend" is the security lending dummy which takes value "Y" when the fund engages into security lending and "Y" otherwise. "Replic" is the replication strategy dummy. This variable takes value "F" when the fund uses full replication and "O" when it uses optimised samples. "Equity" indicates whether the fund replicates equity (Y) or debt securities (N).

Table 3.4. Correlations matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	1.00										
(2)	0.54	1.00									
(3)	0.32	0.27	1.00								
(4)	0.02	-0.20	0.25	1.00							
(5)	0.06	-0.21	0.18	0.93	1.00						
(6)	0.13	0.36	-0.11	-0.67	-0.72	1.00					
(7)	-0.06	0.06	-0.03	-0.14	-0.20	0.13	1.00				
(8)	-0.14	-0.06	-0.34	-0.21	-0.19	0.12	0.12	1.00			
(9)	-0.13	0.11	0.13	-0.17	-0.26	0.04	0.28	0.09	1.00		
(10)	0.12	-0.03	0.09	0.14	0.19	-0.15	0.21	-0.48	-0.11	1.00	
(11)	0.02	-0.01	0.13	0.27	0.31	-0.26	-0.20	-0.21	-0.25	0.19	1.00

Table 3.4: The variables included in the correlation matrix are: (1) The average ETF's premium/discount; (2) The average of the illiquidity measure developed by Chacko, Das and Fan (2016); (3) The size of the creation unit; (4) The average number of days in which exists ETF's creation and redemption activities; (5) ETF's turnover; (6) The relative bid-ask spread; (7) The Amihud illiquidity measure; (8) The volatility of the fund; (9) The total expense ratio (10); The replication strategy dummy; (11) The security lending dummy.

Table 3.5. Correlations matrix extended

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
(1)	1.00																		
(2)	-0.08	1.00																	
(3)	0.51	-0.15	1.00																
(4)	-0.12	0.71	-0.23	1.00															
(5)	0.16	-0.48	0.32	-0.72	1.00														
(6)	-0.11	0.42	-0.21	0.59	-0.67	1.00													
(7)	0.02	-0.26	-0.13	-0.39	0.47	-0.36	1.00												
(8)	-0.13	0.37	-0.24	0.56	-0.78	0.78	-0.42	1.00											
(9)	0.15	-0.46	0.18	-0.70	0.92	-0.65	0.74	-0.76	1.00										
(10)	-0.16	0.49	-0.31	0.73	-0.97	0.72	-0.52	0.86	-0.94	1.00									
(11)	0.29	-0.19	0.57	-0.28	0.40	-0.26	-0.31	-0.30	0.17	-0.38	1.00								
(12)	-0.12	0.46	-0.22	0.55	-0.71	0.31	-0.38	0.22	-0.69	0.63	-0.28	1.00							
(13)	-0.28	-0.11	0.36	-0.16	0.23	-0.15	-0.28	-0.17	0.02	-0.21	0.58	-0.16	1.00						
(14)	-0.05	0.20	-0.09	0.24	-0.28	0.10	-0.15	0.03	-0.27	0.21	-0.11	0.45	-0.06	1.00					
(15)	0.10	-0.16	0.36	-0.38	0.80	-0.54	0.34	-0.69	0.74	-0.81	0.43	-0.52	0.30	-0.11	1.00				
(16)	-0.12	0.70	-0.22	0.80	-0.71	0.42	-0.38	0.44	-0.68	0.68	-0.28	0.67	-0.16	0.45	-0.28	1.00			
(17)	-0.31	-0.08	0.08	0.06	0.04	0.09	-0.22	0.07	-0.07	-0.01	0.12	-0.13	0.34	-0.02	0.09	-0.12	1.00		
(18)	-0.14	0.14	-0.19	0.13	-0.42	0.21	-0.20	0.43	-0.40	0.47	-0.27	0.20	-0.06	-0.10	-0.48	0.16	-0.11	1.00	
(19)	0.15	-0.01	-0.08	0.07	-0.10	0.14	0.11	0.20	-0.02	0.12	-0.19	-0.13	-0.38	-0.16	-0.21	-0.08	-0.25	0.19	1.00

Table 3.5: The variables included in the correlation matrix are: (1) and (2) the premium/discount for equity and debt ETFs, respectively; (3) and (4) The illiquidity developed by Chacko, Das and Fan (2016) for equity and debt ETFs, respectively; (5) and (6) The unit creation size for equity and debt ETFs, respectively; (7) and (8) The creation-redemption activities for equity and debt ETFs, respectively; (9) and (10) The ETF's turnover for equity and debt funds, respectively; (11) and (12) The relative bid-ask spread for equity and debt funds, respectively; (13) and (14) The Amihud illiquidity measure for equity and bonds, respectively; (15) The volatility of the fund; (16) The duration; (17) The total expense ratio; (18) The replication strategy dummy; (19) The security lending dummy.

Table 3.6. Cross-sectional analysis

	ξ_1	t-stat	ξ_2	t-stat	ξ_3	t-stat
Constant	0.001688400	0.96	0.001018000	0.74	0.001675800	0.95
PD	-0.002511200	-1.53	-0.001957600	-1.47	-0.002510900	-1.53
CDF	0.000089261	7.69[a]	0.000067510	6.50[a]	0.000089175	7.70[a]
CreatSize	0.000000001	1.37	0.000000001	1.50	0.000000001	1.37
Activity	0.000001675	3.28[a]	0.000001271	3.08[a]	0.000001670	3.28[a]
Turnover	-0.000194620	-2.09[b]	-0.000138340	-1.87[c]	-0.000193760	-2.08[b]
Spread	0.002952900	1.84[c]	0.002310500	1.90[c]	0.002955200	1.85[c]
Amihud	-0.000000110	-0.05	-0.000000593	-0.32	-0.000000074	-0.03
RiskF	0.129300000	5.13[a]	0.090800000	4.83[a]	0.129000000	5.13[a]
TER	0.000209020	0.51	0.000006978	0.02	0.000214250	0.53
Repl.	0.000062056	0.37	0.000078228	0.62	0.000060323	0.36
Lend.	0.000150040	0.74	0.000114800	0.77	0.000151500	0.75
R^2	0.7728		0.7623		0.7728	

Table 3.6: The variables included in the cross-sectional analysis are: (PD) The average ETF's premium/discount; (CDF) The average illiquidity measure developed by Chacko, Das and Fan (2016); (CreatSize) The size of the creation unit; (Activity) The creation-redemption activities; (Turnover) The ETF's turnover; (Spread) The relative bid-ask spread; (Amihud) The Amihud illiquidity measure; (RiskF) The volatility of the fund; (TER) The total expense ratio charged by the fund; (Repl.) The replication strategy dummy; (Lend.) The ecurity lending dummy. Standard inference applies. [a] Significant at 1% level, [b] Significant at 5% level, and [c] Significant at 10% level. ξ_1 is the TE defined as the standard error of the ECM regression. ξ_2 measures the extent to which the continuously compounded returns of the ETF diverge from those of the benchmark index using the ECM regression. ξ_3 represents the standard deviation of the difference in continuously compounded returns between the ETF and underlying index using ECM.

Table 3.7. Cross-sectional analysis extended

	ξ_1	t-stat	ξ_2	t-stat	ξ_3	t-stat
Constant	0.002155400	1.60	0.001087400	1.04	0.002153300	1.60
PD-E	-0.008269700	-1.37	-0.007774400	-1.70	-0.008331300	-1.38
PD-B	-0.000685360	-0.84	-0.000332290	-0.51	-0.000684660	-0.84
CDF-E	0.000107030	11.79[a]	0.000084257	12.12[a]	0.000106960	11.8[a]
CDF-B	0.000058490	7.42[a]	0.000041663	6.86[a]	0.000058547	7.43[a]
CreatSize-E	0.000000013	1.06	0.000000010	1.06	0.000000013	1.06
CreatSize-B	0.000000001	1.48	0.000000001	1.42	0.000000001	1.49
Activity-E	0.000001809	2.12[a]	0.000001271	1.90[c]	0.000001811	2.12[b]
Activity-B	0.000000848	1.81[c]	0.000000761	1.99[b]	0.000000843	1.80[c]
Turnover-E	-0.000216680	-2.12[a]	-0.000141460	-1.75[c]	-0.000216680	-2.12[a]
Turnover-B	-0.000139730	-1.83[c]	-0.000082262	-1.39	-0.000139540	-1.83[c]
Spread-E	0.003273900	1.62	0.002414200	1.57	0.003279700	1.62
Spread-B	-0.000872020	-0.68	-0.000100560	-0.11	-0.000872080	-0.68
Amihud E	-0.000003723	-0.60	-0.000002381	-0.52	-0.000003781	-0.61
Amihud B	-0.000015794	-0.70	-0.000000673	-0.05	-0.000015529	-0.69
RiskF	0.044700000	0.68	0.037100000	0.79	0.044600000	0.68
Duration	0.000069544	1.78[c]	0.000039087	1.41	0.000069095	1.77[c]
TER	0.000278830	0.72	-0.000018346	-0.06	0.000282020	0.73
Repl.	-0.000106370	-0.98	-0.000082010	-1.03	-0.000106350	-0.98
Lend.	0.000097828	0.59	0.000099299	0.86	0.000098537	0.60
R^2	0.7851		0.7885		0.7850	

Table 3.7: The variables included in the cross-sectional analysis are: (PD-E) and (PD-B) the ETF's premium/discount for equity and fixed debt, respectively; (CDF-E) and (CDF-B) the illiquidity developed by Chacko, Das and Fan (2016) for equity and debt respectively; (CreatSize-E) and (CreatSize-B) The unit creation size for equity and debt, respectively; (Activity-E) and (Activity-B) The ETF's creation-redemption activities for equity and debt, respectively; (Turnover-E) and (Turnover-B) The turnover for equity and debt funds, respectively; (Spread-E) and (Spread-B) The relative bid-ask spread for equity and bond funds, respectively; (Amihud-E) and (Amihud-B) The Amihud illiquidity measure for equity and bond funds, respectively; (RiskF) The volatility of the fund; (Duration) The duration; (TER) The total expense ratio; (Repl.) The replication strategy dummy; (Lend) The ecurity lending dummy. [a] Significant at 1% level, [b] Significant at 5% level, and [c] Significant at 10% level.

Chapter 4

Modelling Time Variation in Tracking Errors: Evidence from the U.S. Market

4.1. Introduction

An ETF constitutes a hybrid between mutual funds and equity contracts and provides investors with an easy and transparent way to achieve portfolio diversification at a very low cost. ETFs aim to replicate the performance of a benchmark index over time. The ETF’s tracking performance and management expenses are the key criteria for investors when comparing competitor ETFs.

A survey on the existing literature that evaluates the ETF’s tracking performance (see Frino and Gallagher 2001; Rompotis 2009; Rompotis 2011; Chu 2011; Bertone, Drenovak, Urošević and Jelic 2014; Paeglis and Ravi 2015; among others) shows that is common practice to assess the ETF's tracking performance by regressing the returns of the ETF on those of the benchmark index and, most importantly, assuming that the slope of this regression is constant over time. However, this regression would be inappropriate because it neglects the stochastic nature of financial data. The main purpose of Chapter 2 of this thesis was to provide an alternative method to assess tracking performance, which delivers a more efficient and accurate measure of the TE than that previously employed in the literature. This new approach considers the stochastic nature of the financial data when assessing ETF's performance and assumes a constant hedge ratio over time. In Chapter 2, we also demonstrate that the assumption of a hedge ratio equal to unity does not hold in practice and therefore, it should be relaxed to avoid misleading estimates of the TEs. In fact, any attempt to assess ETF's tracking performance assuming a hedge ratio equal to unity with the return matching regression might provide erroneous estimates of the ETF's performance. Therefore, in some way our approach constitutes the first valid test of this hypothesis. Finally, we show the consequences of using the TE from a misspecified regression as the main criterion to select competing ETFs. In Chapter 3 we investigate the core factors underlying the cross-sectional differences in ETF's performance. This fourth chapter constitutes a first step in the rather more difficult problem of evaluating what drives the time series differences in ETF's tracking performance. In this chapter, we relax the assumption of constant variance-covariance matrix and hence we allow the hedge ratio to vary over time accordingly. Basically, this approach let us model jointly the first and the second conditional moments of the distribution of returns to the ETF and the benchmark, while allowing for the possibility of a cointegrating relationship between them.

Very basic theory predicts that a risk averse investor will demand higher levels of return to compensate the exposure to increasing levels of risk. Consequently, in every financial decision investors face a time-varying trade-off between risk and expected return. Accounting for this time variation in risk and return is central to modern dynamic portfolio selection and risk management strategies.

In this chapter we focus on the tracking performance of S&P Depositary Receipts, known worldwide as SPY ETF, also called Spider. SPY ETF was launched on the NYSE in 1993 and it has traded continuously since then. We illustrate the in-sample and out-of-sample advantages of allowing for time variation and asymmetry in the variance-covariance matrix. To evaluate the SPY ETF's tracking performance over time, we concentrate on the Tracking Difference (TD), rather than the TE, which we define as the daily divergences that might appear between the performance of the fund and the benchmark, in terms of daily returns. The definition of the TE adopted in this chapter is closely related to the TD, however while the former focuses in variability, through the standard deviation, the latter measures performance in terms of returns.

We use three alternative definitions of the optimal hedge ratio, which will provide three different measures of the TD. The first definition considers a hedge ratio equal to the slope of the return matching regression obtained using OLS. The second approach employs a hedge ratio equal to unity. Both hedge ratios remain constant for the whole period. The third approach uses the hedge ratio obtained from the VECM-BEKK. Comparison of TDs obtained from the three hedging strategies, suggests that simpler models, which neglect the cointegrating relationship and time variation and asymmetry of the covariance matrix, may underestimate the TD. As a result, when portfolio decisions are based on the TD, miscalculation of the TD might lead to suboptimal portfolio choices. The same applies to the TE.

However, the results show that, the daily TD is small in magnitude. In other words, the benefits of allowing the hedge ratio to vary over time are minimal for SPY ETF, the most heavily traded ETF in the world, with more $302 billion in assets under management[7] at the end of January 2018. However, in future research this framework can be expanded to a wider sample of ETFs and to less frequently traded ETFs to assess the robustness of our results.

The next section of this paper presents the literature review, the third section of the paper provides the data description and some preliminary tests, and the fourth section describes our approach to modelling tracking allowing for time variation in the variance-covariance matrix. The fifth section presents the results and discussion. The sixth and the seventh sections present, respectively, the in-sample and out-of-sample estimates of the ETF's tracking performance. The final section provides a summary of the key outcomes and some concluding comments.

4.2. Literature review

A large inflow of funds in January 2018 has driven the assets held in ETFs globally above the $5tn mark[8]. The continuous development of the ETF market since the first ETF was launched has incentivised the research on the ETFs (see Charupat and Miu

[7] Retrieved from the company announcements section of the Financial Times journal. Last accessed on the 29th of January 2018 https://markets.ft.com/data/announce/detail?dockey=600-201801290945BIZWIRE_USPRX____BW5650-1.

[8] Retrieved from the Financial Times journal. Last accessed on the 11th of February 2018 https://www.ft.com/content/5cf7237e-0cdc-11e8-839d-41ca06376bf2

2011; Chu 2011; Buetow and Henderson 2012; Charupat and Miu 2013; Drenovak, Urošević and Jelic 2014; among others).

The literature on ETFs can be divided mainly in four areas. The first area evaluates ETFs' price efficiency. As the creation and redemption process permits the continuous trading of the ETFs' shares during the day at prices determined by demand and supply, differences between the NAV of the underlying securities and the fund's price give rise to premiums or discounts. In general, the literature shows that the price of the fund and the value of the underlying securities move closely over time, and if profitable arbitrage opportunities arise they disappear very quickly (Ackert and Tian 2000; Elton et al. 2002; Gallagher and Segara 2005; among others). In contrast, country ETFs, which are listed on a stock exchange but mirror the performance of indices that trade on a foreign exchange, seem to exhibit less pricing efficiency. The loss of pricing efficiency appears to be related to illiquidity of the to non-synchronous trading and market inefficiencies (Engle and Sarkar 2006; Jares and Lavin 2004; among others).

The second area studies the role of ETFs in the price discovery of the underlying index (Chu, Hseih and Tse 1999; Hasbrouck 2003; and Tse, Bandyopadhyay and Shen 2006; among others). Overall, the findings of the literature indicate that ETFs contribute to the price discovery, however in most cases their role is secondary. Hasbrouck (2003) studies the contributions to price discovery in the S&P 500, S&P Mid-Cap 400, and Nasdaq-100 indices of standard futures, ETFs and electronically traded mini futures. The evidence shows that the main contribution to the price discovery for the S&P 500 and Nasdaq-100 indices occurred through the electronically traded mini futures contracts, while the ETFs played a relatively minor role. In contrast, the contribution of the ETF to price discovery of S&P Mid-Cap 400 index, where the electronically traded mini future contracts did not exist, was key. Tse, Bandyopadhyay and Shen (2006) studies the role of information transmission between the Dow Jones Industrial Average index, the DJIA ETF listed on the Archipielago, the DJIA ETF listed on Amex, the floor-traded futures and the electronically traded mini futures and concludes that, although the main contribution to price discovery was provided by the electronically traded mini futures, the role of DJIA ETF listed on ArcaEx was also relevant.

Since ETFs constitute a mixture of common stocks and traditional index mutual funds, the third strand in the ETF's literature has focused mainly on describing the core features that characterise ETFs (Gastineau 2001; Deville 2008 and Gastineau, 2010), together with their performance relative to competitor funds such as IMFs and closed-end funds (Poterba and Shoven, 2002; Gastineau 2004 and Harper, Madura and Schnusenberg, 2006; Barnhart and Rosenstein 2010). Following a comprehensive comparison of ETFs and index mutual funds, Deville (2008) concludes that the ETFs seem to have contributed to satisfy investors' needs as IMFs are not always available to all investors. Moreover, the study highlights the ability of ETFs to expand investor's allocation opportunities by investing in specific sectors, or even in markets, where IMFs do not exist. ETFs and index mutual funds have also been compared in terms of tax and cost advantages (Poterba and Shoven 2002).

The fourth strand of the literature assesses the quality of the ETF's tracking. The exponential growth experienced by the ETF's market in the last few years is probably the reason underlying the increased interest on the ETF's tracking performance in the literature (Gastineau 2001; Gastineau 2004; Milonas and Rompotis 2010; Houweling 2011; Rompotis 2011; Bertone, Paeglis and Ravi 2015). Essentially, ETFs are investment funds listed on stock exchanges designed to replicate the performance of a benchmark index, and this implies that the performance of the fund and the benchmark should be analogous over time. In practice, however, departures between the managed portfolio and the benchmark, in the form of TE, seem unavoidable. Because of that, the TE has become a key tool in tracking performance evaluation. Despite its role on assessing performance of index tracker funds, there is not a unique definition of TE in the literature. Nevertheless, there are three approaches (see Rudolf, Wolter, and Zimmermann 1999; Roll 1992; Pope and Yadav 1994) that have been extensively used in the previous literature (Shin and Soydemir 2010; Milonas and Rompotis 2010; Buetow and Henderson 2012; Bertone, Paeglis and Ravi 2015; among others). A common practice in the literature has been to calculate the TE using a return matching regression while assuming a hedge ratio equal to unity (Elton, Gruber, Comer and Li 2002; Rompotis 2009; Buetow and Henderson 2012; Bertone, Paeglis and Ravi 2015; among others). However, in the presence of a cointegrating relationship between the ETF and the benchmark, the return matching regression is misspecified and might result in inefficient estimates of the TE and eventually in suboptimal portfolio choices for investors. Furthermore, the assumption of a hedge ratio equal to unity when computing the TE is tenuous at best. See Chapter 2 and Henry and Marí-Clérigues (2017 a,b) for further discussion.

By allowing the covariance matrix to be time varying and accounting for the cointegrating relationship between the ETF and the benchmark, Chapter 4 constitutes the first step in evaluating what drives the time series differences in ETF’s tracking performance. Since both the TE and the TD measure the precision of the ETF‘s replication process and play a crucial role for ETF’s buyers, the focus of this chapter is to model how the tracking performance varies with time and, whether neglecting such variation might impact upon the investment outcomes. Given our interest in the time series dimension of the ETF's tracking performance, this chapter concentrates on the TD, rather the TE. Nevertheless, since TD and the TE are closely related, the conclusions drawn for the TD will be analogous for the TE.

4.3. Data description and preliminary tests

In the early 1990's, S&P Depositary Receipts, known worldwide as SPY ETF was the very first equity ETF listed in the world. This ETF has become a global leader in the market for index investment. We use closing prices of the SPY ETF (Bloomberg ticker: SPY) and its underlying benchmark, the S&P 500 index (Bloomberg ticker: SPXT). Daily closing prices were collected from Bloomberg for the period 29th January 1993 to 21th March 2017. We estimate our models over the period 29th January 1993 to 2nd of April 2013 and retain the remaining 1000 observations for an out-of-sample analysis.

[Figure 4.1]

Figure 4.1 shows the tracking performance of SPY ETF and the S&P 500 index over time. The log prices of the series appear non-stationary and seem to comove over time. However, further inspection of the graph shows that generally the movements in the benchmark are not always equal to the movements of the ETF. Consequently, TEs and TDs arise. Note that we do not date the horizontal axis because of missing observations due to holidays, the 11th of September market closure, etc. which make the observation number a preferable y axis label.

[Table 4.1]

Given the well-documented non-stationary nature of stock prices, the first step is to test whether the prices of the ETF and the benchmark are I (1) processes. We employ two approaches, first we use a Dickey fuller test for the null of unit root and, for robustness, we use the KPSS test of Kwiatkowski, Phillips, Schmidt, and Shin (1992) for the null of stationarity. Not surprisingly, the results in Table 4.1 Panel A provide evidence in favour of the non-stationary behaviour for the levels of the ETF and the benchmark. This is in line with the findings of Chapter 2 and Chapter 3.

Two variables are said to be cointegrated if they are I (1) and there exists a linear combination of them that is stationary. Neglecting the cointegrating relationship might result in misspecified regressions and misleading conclusions. To avoid that, we test for cointegration using the Johansen (1988) test. The outcomes of Table 4.1 Panel B show that the log prices of the ETF and the index appear to share a common stochastic trend, and therefore an error correction term should be included in the mean model to avoid misspecified regressions and erroneous conclusions.

The summary statistics of the returns of the ETF and the benchmark, displayed in Table 4.1 panel C, indicate that the unconditional distribution of both return series, appear to be left skewed, leptokurtic and mean reverting.

[Figure 4.2]

Since most of financial decisions involve a trade-off between the risk and return, studying the dynamics of the second moment of stock returns has gained interest in the literature over the years. As Engle and Paton (2001) points out, among other regularities, there exist a series of stylised facts which characterise financial asset returns, such as volatility clustering, mean reversion and asymmetric response to news, among others. The plots in Figure 4.2 suggest that the returns exhibit volatility clustering, which motivates the calculation of a battery of statistics designed to formally detect the presence of heteroscedasticity in the returns of the ETF and the benchmark. The Engle (1982) LM test for ARCH, displayed in Table 4.1 panel D, provides evidence of fifth and tenth order ARCH effects in both, the ETF and the benchmark.

Standard symmetric GARCH models are not able to capture the asymmetric volatility effect. If a negative return shock has a greater impact on volatility than a positive return shock of equal magnitude, then a symmetric model will systematically under (over) forecast volatility following a negative (positive) return innovation. To determine whether negative and positive shocks and/or their sizes have an impact on the level of volatility, we follow Engle and Ng (1993). Firstly, we test if the ETF and the benchmark display own variance asymmetry, where the sign or the size of a shock in the ETF (index), or both, affect the unconditional variance of the ETF (index). Secondly, we test for cross-variance asymmetry, which implies that the sign or the size of a shock in the ETF (index), or both, affect the unconditional variance of the index (ETF). Thirdly, we test for the presence of covariance asymmetry to news about the ETF or index. Covariance asymmetry occurs when the covariance of returns between the ETF and the benchmark is affected by the sign or the size of a shock, or both, in the ETF or the benchmark.

The tests for own-variance asymmetry in variance are based on regressions of the type:

$$r_{Ft}^2 = \theta_0^{'} + \theta_1^{'}(S_{F,t-1}) + \theta_1^{''}(S_{F,t-1})r_{F,t-1} + \theta_1^{'''}(1-S_{F,t-1})r_{F,t-1} + \upsilon_t \tag{4.1}$$

Where the subscript F refers to the ETF. The Eq. (4.1) can be also written in terms of the benchmark by changing the subscript from F to I. Moreover, $S_{F,t-1} = 1$ if

$r_{F,t-1} < 0$ and zero otherwise. The actual return to holding the index for one period is computed as follows $r_{I,t} = \log(i_t / i_{t-1})$, similarly the return on the ETF is obtained using the following formula $r_{F,t} = \log(f_t / f_{t-1})$. A statistically significant estimate of $\theta_1^{'}$ provides evidence of own variance asymmetry in the form of negative sign bias. The t-statistics for $\theta_1^{''}$ and $\theta_1^{'''}$ can be used to test for negative and positive size bias, respectively. A joint test for the null of no asymmetry may be performed as a joint test of the null hypothesis: $H_0 : \theta_1^{'} = \theta_1^{''} = \theta_1^{'''} = 0$. This null hypothesis is tested using a Wald test.

Tests for cross-variance asymmetry may be obtained in an analogous fashion from:

$$r_{F,t}^2 = o_0^{'} + o_1^{'}(S_{F,t-1}) + o_1^{''}(S_{I,t-1})r_{I,t-1} + o_1^{'''}(1 - S_{F,t-1})r_{F,t-1} + \upsilon_t \qquad (4.2)$$

Where $S_{I,t-1} = 1$ if $r_{I,t-1} < 0$ and zero otherwise. The subscript F refers to the ETF. The Eq. (4.2) can be also written in terms of the benchmark by swapping the subscripts from F to I. To the best of the author's knowledge no tests of cross variance asymmetry have been reported in the existing literature on ETFs' TEs.

Finally, tests for covariance asymmetry may be based upon the rather more cumbersome regression:

$$r_{F,t}r_{I,t} = \iota_0^{'} + \iota^{'}(S_{F,t-1}) + \chi^{'}(S_{I,t-1}) + \iota^{''}(S_{F,t-1})r_{F,t-1} + \chi^{''}(S_{I,t-1})r_{I,t-1}$$

$$+\iota^{'''}(1 - S_{F,t-1})r_{F,t-1} + \chi^{'''}(1 - S_{I,t-1})r_{I,t-1} + \vartheta_t \qquad (4.3)$$

Again, the significance of an individual slope in Eq. (4.3) provides evidence of covariance asymmetry. For instance, a significant estimate of $\iota^{'}$ implies that the covariance responds asymmetrically to the sign of an innovation for the ETF, and a significant coefficient estimate of χ''' implies a covariance asymmetry arising from the size of positive shocks to the benchmark. A joint test for the null of no covariance asymmetry may be performed using an LM, F or Wald type tests for the null hypothesis that all the slopes are jointly insignificantly different from zero. To the best of the author's knowledge no tests of covariance asymmetry have been reported in the existing literature on ETFs' TEs.

[Table 4.2]

Panel A of Table 4.2 reports the results of the tests for own variance asymmetry. The evidence suggest that the size and the sign of the news appear to determine the direction and the magnitude of the volatility movements for both, the ETF and the benchmark. The results, shown in panel B of Table 4.2, suggest that the fund and the index also display cross-variance asymmetry. Finally, panel C of Table 4.2 reports the results of a series of tests for covariance asymmetry. The outcomes show that the magnitude of the shocks, for both the fund and the index, also affect the covariance. It is worth noting in passing, that the coefficient of the negative shocks on the fund is substantially larger than that of the index, which might indicate that negative shocks in the fund produce a greater movement in the covariance than positive shocks of equal size in the benchmark.

Since the ETF and the benchmark display own variance, cross-variance asymmetry and covariance asymmetry, when modelling the volatility, we need to allow for these asymmetries in our model. The resulting hedge ratio may also display asymmetry, except in the case that the asymmetry in the covariance between the ETF and the benchmark is exactly offset by the asymmetry in the variance of the index. In such an exceptional case, the hedge ratio will respond symmetrically to returns' shocks.

Taken together, the results displayed in Tables 4.1 and 4.2 suggest that any candidate model of the relationship between returns to the SPY ETF and S&P 500 index needs to allow for the presence of a cointegration vector between the levels of the ETF and the index, for ARCH effects and for own, cross and covariance asymmetry.

4.4. Modelling time-variation in ETF's tracking

The recent growth of passively managed index funds has motivated the usage of the TE to evaluate how accurate index tracker funds replicate the underlying benchmark. As we have already mentioned, there is no a unique definition of the TE in the literature (see Roll 1992; Pope and Yadav 1994; among others). Despite this, there is a measure of the TE that has been extensively used in the literature to assess ETF's performance. This measure, which coincides with the one adopted in this chapter, is defined as the volatility of the difference in returns between the fund and the benchmark. We denote F_t and I_t as the prices of the ETF and benchmark index, with logarithms f_t and i_t, respectively. The return series of the ETF and the benchmark index are computed as follows: $r_{F,t} = \log(f_t / f_{t-1})$ and $r_{I,t} = \log(i_t / i_{t-1})$, *res*pectively. Given this, we define the TE as:

$$\xi^{*}_{mn,t} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(TD_{mn,t} - \overline{TD})^2} \tag{4.4}$$

Where $TD_{mn,t} = r_{Fm,t} - \lambda_t r_{In,t}$ and $\overline{TD} = T^{-1}\sum_{t=1}^{T} TD_{mn,t}$. The subscript m and n refer to the mth fund and the nth benchmark index. T is the sample size.

The ETF's literature does not include λ_t in the definition of the TE because it assumes that λ_t is equal to unity and constant over time, therefore it computes the TE as:

$$\xi^{*'}_{mn,t} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left(e_{mn,t} - \bar{e}\right)^2} \quad (4.5)$$

Where, $e_{mn,t} = \Delta F_{m,t} - \Delta I_{n,t}$ and $\bar{e} = T^{-1}\sum_{t=1}^{T} e_{mn,t}$. If λ_t =1 and constant over time, then $TD_{mn,t} = e_{mn,t}$. The reason underlying the inclusion of λ_t in the definition of the TD, and hence in the Eq. (4.4), is that we make different assumptions about the hedge ratio through the chapter, and so we need to include it explicitly.

Comparison of ETFs can also be done be using the concept of TD, which is closely related to the TE as shown in Eq. (4.5) but seems to better capture the time series dimension of the ETFs' tracking performance. Therefore, given our interest on the time series dimension of the ETFs' performance, from now onwards we focus on the TD, rather than the TE. We define TD_t, which represents the tracking difference in period *t*, as:

$$TD_t = r_{F,t} - \lambda_t r_{I,t} \quad (4.6)$$

Again, the ETF literature assumes that λ_t in the definition of the TD is constant and equal to unity. When $\lambda_t = 1$ and the ETF tracks the benchmark perfectly such that $r_{F,t} = r_{I,t}$, then the TD and the TE are equal to zero. Given the close relationship between the TD and the TE, the conclusions drawn in this chapter for the TDs are also valid for the TEs.

To study the tracking ability of the ETFs in our sample we follow the methodology developed in Chapter 2, which models the ETF's tracking based on a simple model of optimising behaviour. The idea is to associate the fund tracking to an optimal hedge ratio for a portfolio which is long one unit of the ETF funded by a short position of lambda λ units of the underlying benchmark index. We assume that the maximising agent relies on a standard mean-variance utility function and that the fund and the benchmark follow a martingale process. Under this condition, the optimal number of units of the benchmark index is given by:

$$\lambda^{*} = -\frac{\sigma_{FI}}{\sigma_{I}^{2}} \tag{4.7}$$

See Kroner and Sultan (1993) for further details. In this framework, the portfolio return equals zero when the ETF perfectly replicates the index. Moreover, the resulting optimal weight in Eq. (4.7), under this scenario, will be unity since perfect tracking implies $\sigma_I = \sigma_F$. Kroner and Sultan (1993) argue that when the distributions of ETF and benchmark are assumed to be constant over time, this model could be extended to a multi-period model, if the utility function is time-separable. In this case, the hedge ratio is obtained from least squares estimation as follows:

$$r_F = a + br_I + \in \tag{4.8}$$

The expression in Eq. (4.8) constitutes the return matching regression, which obtains b using OLS and assumes that it remains constant over time.

Since the distribution of the ETF and the benchmark prices varies over time, we consider an alternative model for the optimal hedge ratio following Kroner and Sultan (1993). If we define $\lambda_{t/t-1}$ as the number of units of the index at time t-1 and

R_t as the payoff at time t of a portfolio that is long one unit of the ETF funded by a short position of $\lambda_{t/t-1}$ units of the benchmark:

$$R_t = r_{F,t} - \lambda_{t/t-1} r_{I,t} \tag{4.9}$$

Supposing again that the agent faces a two-moment expected utility function:

$$E(U(R_{t+1})) = E_t(R_{t+1}) - \psi \sigma^2{}_{R,t+1} \tag{4.10}$$

where ψ represents the degree of risk aversion and ψ>0. If the prices of the benchmark follow a martingale process, maximisation of Eq. (4.10) conditional on the information available at time t results in:

$$\lambda_t = -\frac{\sigma_t(F_{t+1,}I_{t+1})}{\sigma_t^2(I_{t+1})} \tag{4.11}$$

The hedge ratio in Eq. (4.11) is analogous to the one obtained in Eq. (4.7). However, in Eq. (4.11) the constant unconditional moments are substituted by conditional moments that vary over time. As a result, the hedge ratio will be updated as news of the fund and the benchmark arrive to the market. The conditional model can be considered a special case of the unconditional model if the joint distribution of the ETF and benchmark prices remains invariant over time.

Most existing research on ETFs' tracking performance is based on Eq. (4.8) which is the returns matching regression (Elton, Gruber, Comer and Li 2002; Rompotis 2009; Buetow and Henderson, 2012; Bertone, Paeglis and Ravi 2015, among others). In Chapter 2 we show that when the ETF and the underlying benchmark share a common stochastic trend, the return matching regression should be avoided since it neglects the long run relationship between the fund and the index and the estimates of the coefficient b will be biased downwards (Kroner and Sultan 1993).

Furthermore, when return matching regressions are employed in the literature to assess performance, the resulting measures might be at best inefficient, and at worst will yield invalid and misleading inference.

The approach used in this chapter is motivated by the previous research on dynamic hedging using futures (see Figlewski 1984; Cecchetti, Cumby and Figlewski 1988; Kroner and Sultan 1993; Brooks and Chong 2001; Engle 2002; Brooks, Henry and Persand 2002; among others). Cecchetti, Cumby and Figlewski (1988) provide evidence that the joint distribution of the spot and futures returns varies with time and so the optimal hedge ratio varies too. Likewise, Engle (2002) argues that when the correlations and volatilities change over time, the hedge ratio should be updated accordingly. Brooks, Henry and Persand (2002) demonstrate that in the presence of asymmetric dynamics in the conditional variance-covariance matrix, the hedge ratio will respond asymmetrically to positive and negative shocks unless the own and cross-variance asymmetry are exactly offsetting, a stochastic singularity that they do not observe in the data considered

The data description section suggests that the SPY ETF and the S&P 500 index share a common stochastic trend. There is also evidence of ARCH effects, own variance asymmetry, cross-variance asymmetry and covariance asymmetry in the ETF and the benchmark. Therefore, following the evidence found and the outcomes from previous research (see Cecchetti, Cumby and Figlewski 1988; Brooks, Henry and Persand 2002, Alexander and Barbosa 2008; among others) we fit an asymmetric VECM-GARCH model to the returns of the ETF and the benchmark. Since the BEKK form of Engle and Ng (1995) assumes a symmetric time-varying variance-covariance structure, following Brooks, Henry and Persand (2002) we define $\tau_{n,t} = \min(\varepsilon_t, 0)$ for n = ETFs, benchmarks, and extend the BEKK model to account for the asymmetric responses of past return innovations. The resulting model can be expressed as follows:

$$r_t = \mu + \sum_{s=1}^{n} \Gamma_s r_{t-s} + \Pi u_{t-1} + \varepsilon_t$$

$$r_t = \begin{bmatrix} r_{F,t} \\ r_{I,t} \end{bmatrix}; \mu = \begin{bmatrix} \mu_F \\ \mu_I \end{bmatrix}; \Gamma_s = \begin{bmatrix} \Gamma_{s,F}^{(F)} & \Gamma_{s,I}^{(F)} \\ \Gamma_{s,F}^{(I)} & \Gamma_{s,I}^{(I)} \end{bmatrix}; \pi = \begin{bmatrix} \pi_I \\ \pi_F \end{bmatrix}; \varepsilon_t = \begin{bmatrix} \varepsilon_{F,t} \\ \varepsilon_{I,t} \end{bmatrix} \tag{4.12}$$

Where Π represents the error correction term, u_t represents the residual of the cointegrating regression, and Ω_{t-1} the information set at time t-1. If we assume that $\varepsilon_t | \Omega_{t-1} \sim [0, H_t]$, where ε_t represents the innovation in Eq. (4.12), the bivariate asymmetric GARCH (1,1) model with the BEKK parametrisation can be written as follows:

$$H_t = \begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22}, t \end{bmatrix} \tag{4.13}$$

And

$$H_t = C_0^{*'} C_0^* + A_{11}^{*'} \varepsilon_{t-1} \varepsilon_{t-1}^{'} A_{11}^* + B_{11}^{*'} H_{t-1} B_{11}^* + D_{11}^{*'} \tau_{t-1} \tau_{t-1}^{'} D_{11}^* \tag{4.14}$$

Where,

$$C_0^* = \begin{bmatrix} c_{11}^* & c_{12}^* \\ 0 & c_{22}^* \end{bmatrix}; A_{11}^* = \begin{bmatrix} \alpha_{11}^* & \alpha_{12}^* \\ \alpha_{21}^* & \alpha_{22}^* \end{bmatrix}; B_{11}^* = \begin{bmatrix} \beta_{11}^* & \beta_{12}^* \\ \beta_{21}^* & \beta_{22}^* \end{bmatrix}; D_{11}^* = \begin{bmatrix} \delta_{11}^* & \delta_{12}^* \\ \delta_{21}^* & \delta_{22}^* \end{bmatrix}; \tau_t^2 = \begin{bmatrix} \tau_{F,t}^2 \\ \tau_{I,t}^2 \end{bmatrix}$$

Based on the features observed in the data description section, we employ Eq. (4.12) - (4.14) to model the returns of the ETF and the index. The quadratic form of this model ensures that the conditional covariance matrix is positive definite by definition, which represents an advantage compared to other multivariate GARCH specifications (see Bollerslev 1986). Moreover, this model allows us to test of the null hypothesis of no volatility spillover effects from the fund to the index, from the index to the fund, and in both directions. This represents an advantage to the dynamic conditional correlation model of Engle (2002), which assumes a diagonal conditional variance-covariance matrix and consequently, it does not allow for cross-volatility effects and spillovers between the fund and the benchmark. Another advantage of the asymmetric VECM-BEKK is that, when using a bivariate system, there are only fifteen parameters in the conditional variance-covariance matrix that need to be estimated, including a constant. The curse of dimensionality, discussed by Caporin and McAleer (2012), should not be an issue in this case.

4.5. Results and discussion

The heteroscedastic nature of stock returns has been extensively documented in the literature, and GARCH models have proved useful in explaining the distribution of stock prices (Bollerslev 1987; Bollerslev, Engle and Wooldridge 1988; Baillie and Myers 1990; Engle 2002; Jones and Olson 2013; Caporale, Ali and Spagnolo 2015; among others). Furthermore, there exists a large body of research that provided evidence of the adequacy of GARCH-BEKK models in modelling the variance and covariance of financial data (see Brooks, Henry and Persand 2002; Grier, Henry, Olekalns and Shields 2004; Caporale, Ali and Spagnolo 2015; Li and Giles 2015; Billio, Donadelli, Paradiso and Riedel 2017; among others). The main drawback of the BEKK model however, is that the parameter estimates are difficult to interpret individually. Allowing for time variation and asymmetry in the covariance matrix of returns seems advisable (Brooks, Henry and Persand 2002), and implies that the time-varying hedge ratio between the ETF and the index may be affected as news arrive to the market.

[Table 4.3]

To select the lag length of the VECM, we use Akaike's information criterion (AIC), Schwarz's Bayesian information criterion (BIC), and Hannan and Quinn information criterion (HQC), which are common approaches for model order selection. Results are displayed in Table 4.3. While the AIC criterion recommends a VECM with 8 lags, the other two criteria suggest 7 lags.

[Table 4.4]

The unconditional distribution of the return series, for both ETF and index, appear to be left skewed and leptokurtic. Since Gaussian GARCH models might not handle properly the leptokurtosis exhibited by the continuously compounded returns of the ETF and the benchmark, following Bollerslev (1987) we estimate Eq. (4.12) - (4.14) jointly using the method of likelihood, using 8 lags in the VECM, and assuming that the residual vector is jointly distributed as a conditional Student's t with unknow degrees of freedom. The parameter estimates are displayed in Table 4.4 along with their associated robust standard errors calculated using the method of Bollerslev and Wooldridge (1992). As a robustness measure, we also report the estimates for Eq. (4.12) – (4.14) assuming the same distribution for the residuals but including 7 lags in VECM. Furthermore, we estimate Eq. (4.12) - (4.14) with 7 and 8 lags, assuming normally distributed residuals and using quasi-maximum likelihood methods. These estimates are relegated to the appendix (Tables A.4.1, A.4.3, A.4.5) to avoid excessive repetition.

The results displayed in Table 4.4 show that coefficients of the error correction terms appear negative and significant, which suggests that both, the fund and the benchmark, seem to contribute to restore the equilibrium after a shock in the system. The first lag of the fund and the benchmark in the benchmark equation appear to be statistically significant. Conversely, in the fund equation the first four lags of the fund and the benchmark appear significant. In both cases the results suggest that news are aggregated via the ETF and the benchmark. In other words, there exists some evidence of a linear feedback relationship between the ETF and the underlying index.

Table 4.4 also displays the coefficient estimates of the second conditional moment. The coefficients α_{11}^{*} and α_{22}^{*} are statistically significant, which indicates that past return shocks on the ETF (index) at least influence current volatility in the ETF (index). The off-diagonal parameters are both statistically significant, which suggests the presence of a bidirectional shock transmission between the ETF and the index. The main diagonal elements of β^{*} matrix are also significant, which suggests that past ETF (index) volatility is an important determinant of current volatility, that is, volatility is persistent. The off-diagonal element β_{11}^{*} is positive and significant indicating the possible presence of persistent volatility transmission from the fund to the index. The coefficient of β_{22}^{*} appears marginally insignificant consequently, there is little evidence that historical ETF's return volatility impacts upon the current volatility of the returns of the benchmark index. The significance of the parameter δ_{22}^{*} is consistent with the presence of at least own variance asymmetry in the ETF's volatility. Conversely, the coefficient of δ_{11}^{*} appears to be insignificant and suggesting that, the index may not display own variance asymmetry. The parameter

δ_{12}^{*} appears significant, which provides evidence of possible cross-variance asymmetry and covariance asymmetry between the fund and the benchmark. These outcomes, together with the results of the sign and size bias tests, provide evidence of own variance, cross-variance and covariance asymmetry in the data.

The standardised residuals are defined as $z_{i,t} = \varepsilon_{i,t}\sqrt{h_{i,t}}$ for i = *F*, *I* and their squares. Panel C of Table 4.4 displays the summary statistics of the standardized residuals. Compared with the raw data, we observe that the coefficients of skewness and excess kurtosis have been reduced. Two further tests are conducted as residual diagnostics. The first test is Engle's Lagrange multiplier test to assess the significance of ARCH effects (see Engle 1982). The second test is the Ljung-Box Q-test for the null of no linear correlation in the residuals. Both tests are asymptotically distributed as a χ^2. The outcomes suggest that there is no evidence of neglected ARCH or autocorrelation in the standardised residuals.

[Table 4.5]

A range of diagnostic tests for the multivariate asymmetric BEKK-GARCH model are presented in Table 4.5 panel A. The first row displays the Lagrange multiplier test for multivariate ARCH effects. The null is rejected at 10% level. We also test whether the data supports the null hypothesis of diagonality in the variance structure of BEKK-GARCH using the null $H_3: \alpha_{12}^{*} = \alpha_{21}^{*} = \beta_{12}^{*} = \beta_{21}^{*} = \delta_{12}^{*} = \delta_{21}^{*} = 0$. The Wald test for the joint null is displayed in the second row of Table 4.5. The null of diagonality is strongly rejected and therefore, the model admits the possibility that news about the fund (index) volatility can help to forecast the index (fund) volatility. The third row of Table 4.5 presents the outcomes of a Wald test for the null hypothesis of no asymmetry in the GARCH-BEKK model, which presumes that the coefficients of δ_{11}^{*}, δ_{12}^{*}, δ_{21}^{*} and δ_{22}^{*} are jointly zero. The null of no asymmetry is rejected at 1% level of significance, which suggests that asymmetry should to be accounted for in the model.

The outcomes of the model provide evidence of a complex causality between the ETF and the index. In line with Granger (1988), we argue that when the ETF and the index share a common stochastic trend, at least causality in mean between the fund and the benchmark must exist, which will re-establish the long run equilibria after a shock in the system. We formally test for short run and long run linear Granger causality using several Wald tests. Basically, the null hypothesis for Granger causality in the index (fund) tests the joint significance of the lags of the returns of the fund (index) in the index (fund) equation of the VECM. The results, presented in the fourth and the fifth rows of Table 4.5, indicate that the null hypotheses are rejected at 5% level of confidence or better. This constitutes a formal evidence of a bidirectional feedback relationship between the ETF and the index. In other words, information about the returns of the fund might help us to forecast the returns of the index and vice versa.

Next, we test for long run Granger causality in mean. In this case, the null hypothesis constrains the coefficients of the lagged values of the index (fund) together with coefficient of the error correction term of the index (fund) equation of the VECM to be jointly zero. The sixth and the seventh rows of Table 4.5 exhibit the results of the Wald tests. The null hypothesis is rejected for both the index and the fund, which again implies the existence of a bidirectional relationship between the benchmark and the fund. More importantly, the evidence suggests that there is no evidence of leading or lagging. Consequently, price discovery occurs in both markets. As Brooks, Rew and Ritson (2001) points out, when the markets are internationally efficient, the spot and the future prices should reflect simultaneously the news as they arrive to the market to avoid arbitrage opportunities. Moreover, in the spot future relationship as Wahab and Lashgari (1993) argues, observing the true prices can become very difficult when stocks trade infrequently and, as a result, it might appear that future prices lead over spot prices. It can also happen that spot prices do not update the changes in the stocks included in the index, and hence they lag the developments in the stock market. Unlike the literature on futures pricing, in our case there is no evidence of leading or lagging. In other words, the outcomes from the short-run and long-run linear Granger causality tests are consistent and suggest that there is no evidence that price discovery occurs in the ETF (benchmark) market and spills over into the benchmark (ETF) market.

Given the feedback relationship observed in the mean and its implications in forecasting, we are interested in testing whether this causal relationship is also present in the variance. However, this is a non-trivial task in the case of the BEKK model. Our evidence suggests that simpler models such as the constant correlation model of Bollerslev (1990) or the Dynamic Constant Correlation model of Engle (2002) would not provide an adequate conditional characterisation. One approach to tackle this would be to estimate a full-VECH model however, this is difficult computationally and even harder in guaranteeing that the conditional variance-covariance matrix is positive definite for all t. On balance we chose to work with the BEKK structure.

To illustrate the difficulties that we face in testing for causality in variance, consider the conditional variance of the benchmark index $h_{11,t}$ given by:

$$h_{11,t} = c_{11}^{*} + \alpha_{11}^{*2}\varepsilon_{1,t-1}^{2} + 2\alpha_{11}^{*}\alpha_{21}^{*}\varepsilon_{2,t-1}\varepsilon_{1,t-1} + \alpha_{21}^{*2}\varepsilon_{2,t-1}^{2} + \beta_{11}^{*2}h_{11,t-1} + 2\beta_{11}^{*}\beta_{21}^{*}h_{12,t-1} +$$

$$\beta_{21}^{*2}h_{22,t-1} + \delta_{11}^{*2}\tau_{1,t-1}^{2} + 2\delta_{11}^{*}\delta_{21}^{*}\tau_{1,t-1}\tau_{2,t-1} + \delta_{21}^{*2}\tau_{2,t-1}^{2} \quad (4.15)$$

Hence, one way to test whether exists short run causality from the volatility of the ETF to the volatility of the index, is to test the null hypothesis $H_4 : \alpha_{21}^{*2} = \delta_{21}^{*2} = 0$. This hypothesis rules out news spillovers alone. Likewise, we test for long run causality from ETF's volatility to the volatility of the index with the null $H_6 : \alpha_{21}^{*2} = \beta_{21}^{*2} = \delta_{21}^{*2} = 0$. Similarly, we also test for short and long run causality in the variance of the fund and the index. Testing such joint non-linear restrictions is not feasible using Wald tests. We note that under the alternative null hypothesis $H_{4*} : \alpha_{21}^{*2} + \delta_{21}^{*2} = 0$, H_4 is satisfied since the terms in both hypotheses are positive definite. That is, given that both α_{21}^{*2} and δ_{21}^{*2} are not negative-definite, satisfaction of H_{4*} implies satisfaction of H_4.

Since the joint null of a series of squared coefficients cannot be tested directly, we set the null as the sum of a series of squared coefficients and we estimate the standard errors for the combined statistic using the delta method. The results of the tests are shown in rows 7th to 11th of panel A Table 4.5. The evidence suggests that there exists a feedback relationship between the variance of the ETF and that of the index. Nevertheless, the relationship between the variance of the fund in response to movements of the index is only marginally significant. Yet, it is worth noting on passing, that the evidence for causality from the index to the ETF seems to be stronger, both in the short run and in the long run, as the null is rejected at 1% level of confidence.

Since AIC and BIC and HQC provide different lag length for the VECM-GARCH, initially we model the conditional mean of the fund and the index using 8 lags, as suggested by the AIC criterion, and assuming that the residuals follow a conditional Student t distribution. As a robustness check, we model the first conditional moment of the ETF and the benchmark using the 7 lags recommended by BIC and HQC and assuming conditional Student t residuals. The outcomes are analogous hence, the conclusions drawn remain unchanged (see Tables A.4.1 and A.4.2 in the appendix). Again, for robustness we estimate the VECM-BEKK model assuming 7 and 8 lags in the VECM and gaussian residuals. The main results, displayed in Tables A.4.3, A.4.4, A.4.5 and A.4.6 in the appendix, as well as the main conclusions remain qualitatively unchanged.

Overall, the outcomes demonstrate that there exists a cointegrating relationship between the ETF and the benchmark. Because of that, we model the conditional mean and the conditional variance using a VECM-BEKK. Based on the features of the data observed, we test for causality in mean and causality in variance. Granger causality tests on the conditional mean and conditional variance provide evidence of a bidirectional feedback relationship between the ETF and the benchmark. The model specification tests and the residual diagnostics provide evidence of a well specified model. Put simply, the model was able to capture the dependence and asymmetry on past values in the conditional first and second moments for the ETF and the benchmark and therefore, may result in reliable forecasts.

The graphs in Figure 4.3 display the conditional volatility of the ETF and the benchmark index, which appear to overlap during the sample period. Figures 4.4 and 4.5 plot the conditional correlation and the conditional hedge ratio, respectively, which vary substantially over time, as expected. This suggests that the assumption of constant conditional volatility and hedge ratio assumed by the previous literature may be tenuous and could possibly yield invalid inference, unreliable forecasts, and most importantly, substandard investment outcomes. Figures 4.6 and 4.7 display the impact surface for the variances of the ETF and the underlying index, respectively. Finally, Figure 4.8 presents the impact surface for the covariance between the ETF and the benchmark.

4.6. In-sample analysis of the TD

Passively managed index tracker funds have grown at a very fast pace in the last few years. Consequently, TE have become an important tool to evaluate the extent to which index tracker funds replicate the performance of the benchmark index. Although many definitions of the TE are exist in the literature, this chapter focuses on a particular metric of the TE, $\xi^{*}_{mn,t}$, which measures the volatility of the differences in returns between the ETF and the benchmark. Accordingly, we define the TD as the differences between the continuously compounded returns of the fund and those of the index. To compute the TD is necessary to obtain the optimal hedge ratio beforehand. In line with Alexander and Barbosa (2008), we employ three different definitions of the optimal hedge ratio. These definitions of the hedge ratio will allow us to evaluate the advantages of permitting time variation and asymmetry in the variance-covariance matrix. In the first case, we define the hedge ratio as the lambda coefficient obtained from OLS estimation of the return matching regression in (4.8), which remains constant over time. In the second case, we set $\lambda = 1$. In the third case, we employ the time varying hedge ratio obtained from the VECM-BEKK.

[Table 4.6]

Table 4.6 exhibits the descriptive statistics of the TD computed using the three methods explained above. Comparison of the mean of the TDs, suggests that when simpler models, which neglect both the cointegrating relationship and the observed time variation and asymmetry in the variance-covariance matrix, are employed the mean of the TD becomes smaller. Consequently, the simpler models used to compute the tracking performance, on average, underestimate the TD, which in turn might have serious implications for investors who use the TD as selection criteria for competing funds. The results for the TE are analogous. We compute the annually equivalent TDs, and we observe that although the TDs become larger the conclusions previously drawn remain the same. Therefore, the benefits of allowing the hedge ratio to vary over time seem minimal for the most heavily traded ETF in the world, SPY ETF. However, in future research this framework can be expanded to a wider sample of ETFs and to less traded ETFs, which might change the outcomes dramatically.

4.7. Out-of-sample analysis of the TD

As mentioned in the data section, we estimate the in-sample TDs with data over the 29th January 1993 to 2nd of April 2013, and retained 1000 observations for an out-of-sample analysis. In this section we describe the outcomes of the out-of-sample exercise using the remaining 1000 observations, and compare the results obtained with the ones reported in the previous section.

-Table 4.7-

The descriptive statistics of the TDs, computed using the different definitions the hedge ratio, are displayed in Table 4.7. The mean of the TDs for the three methods is similar. Again, we observe that when we neglect the common stochastic trend between the ETF and the index and the time variation and asymmetry in the covariance matrix, the mean of the resulting TD reduces its magnitude. In other words, when simpler models are used to evaluate the tracking performance, on average, they miscalculate the TD, which again might be an issue when the TD is used by investors as a selection criterion. The results for the TE and for the annualized TD are comparable.

Overall, the outcomes of the in-sample and out-sample exercises are analogous. Simpler models that neglect the time series dynamics of the data and the time variation and asymmetry of the variance-covariance matrix, underestimate the TD and eventually might result in suboptimal portfolio choices when any of these metrics is employed to compare competing ETFs. Therefore, we can conclude that in this setting the BEKK model is advisable to avoid misleading outcomes and suboptimal portfolio choices.

4.8. Conclusion

Given the importance of the TEs for investors, we follow the approach proposed in Chapter 2 to assess the ETFs' tracking performance. This approach associates the tracking ability of an ETF to a hedge portfolio that is long one unit of the ETF funded by a short position in the benchmark index. This methodology is motivated by the fact that the prior literature has generally neglected the long run relationship between the fund and the benchmark index when computing the TEs.

Using daily closing prices of the first and most popular ETF in the USA, SPY ETF, and its benchmark, the S&P 500 index, from the 29th January 1993 till the 21th March

2017, we model jointly the first and the second moment of the distribution of returns for the fund and the benchmark index and relax the assumption of constant hedge ratio when computing the TD.

The results show that the ETF and the benchmark index share a common stochastic trend. Furthermore, GARCH class of models have proved useful in explaining the distribution of common stock returns when time variation is allowed (see Bollerslev 1987, Bollerslev, Engle and Wooldridge 1988, Baillie and Myers 1991). Therefore, we model the first conditional moment of the returns distribution, using a VECM, jointly with the second conditional moment, employing a multivariate asymmetric GARCH (1,1) structure. Essentially, our main aim is to study the impact of short-run departures from the long-run equilibrium between the ETF and the benchmark when the conditional second moments are time varying. Since the optimal hedge ratio is defined as the ratio of the covariance between the ETF and index and the variance of the index, we allow the hedge ratio to be time varying and display asymmetry in the response to news. We use a battery of diagnostic tests to ensure the adequacy of the model.

Our results suggest that there is a bidirectional causal relationship between the returns of the ETF and those of the benchmark, and their associated volatility. This implies that the two factors driving the hedge ratio updates are the news about the benchmark and the fund. Essentially, the outcomes indicate that as new information arrives to the market the volatility of the fund and the index change, and the conditional lambda coefficient varies accordingly. Moreover, in the presence of asymmetry, there is the possibility that the conditional lambda will display asymmetric response to news unless any variance and covariance asymmetry observed in the data is offsetting. In the case of SPY ETF, as the quality of the tracking is sufficiently high, there is little evidence of dynamic asymmetry in the hedge ratio.

To assess the advantages of allowing for time variation and asymmetry in the variance-covariance matrix, we employ three different definitions of the optimal hedge ratio. Overall, the outcomes indicate that when simpler models, which neglect the time series dynamics of the data such as cointegration and time variation and asymmetry in the covariance matrix, are used to assess the ETF's tracking performance the resulting measure of tracking performance, in the form of TD, is underestimated. The main implication of using the TD or the TE provided by these simpler models to evaluate ETF's tracking performance, is that they might result in suboptimal portfolio choices.

It is worth noting on passing, that in any case, the daily TDs are of small magnitude. Consequently, the benefits of allowing the hedge ratio over time might seem minimal for the most heavily traded ETF in the world, SPY ETF. However, in future research this framework can be expanded to a wider sample of ETFs and to less traded ETFs, which might change the outcomes dramatically.

Figures

Figure 4.1. The tracking performance of SPY ETF and the benchmark

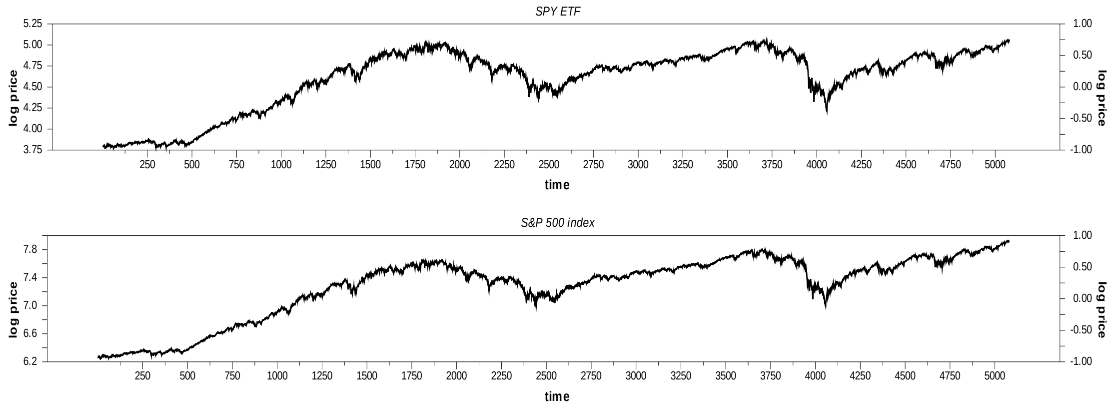


Figure 4.1: The former graph shows the logarithm of the closing prices of the SPY ETF, and the latter graph displays those of the underlying benchmark (S&P 500 index)

Figure 4.2. Returns of the ETF and the benchmark

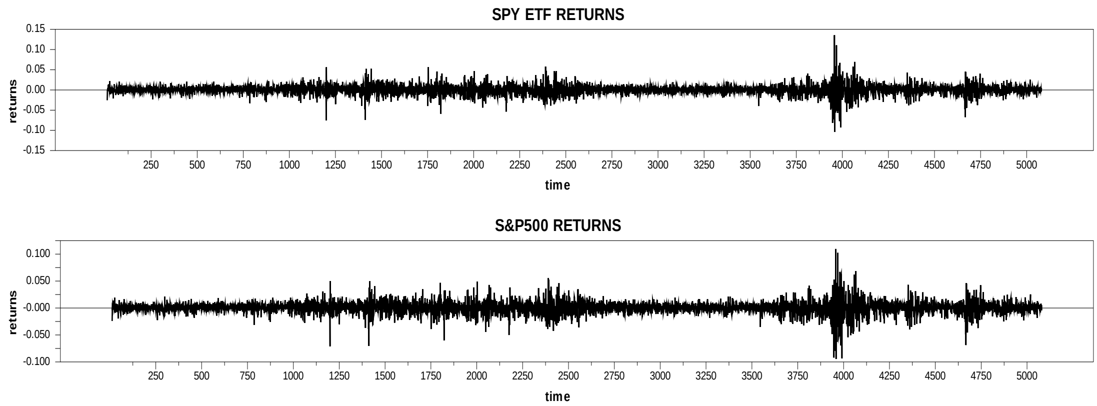


Figure 4.2: The first and the second graphs show the continuously compounded returns of the SPY ETF and the S&P 500 index, respectively.

Figure 4.3. Conditional variances

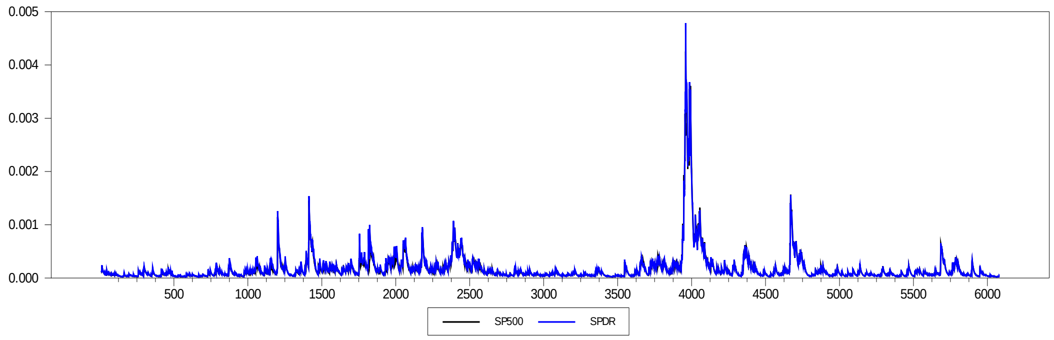


Figure 4.4. Conditional hedge ratio

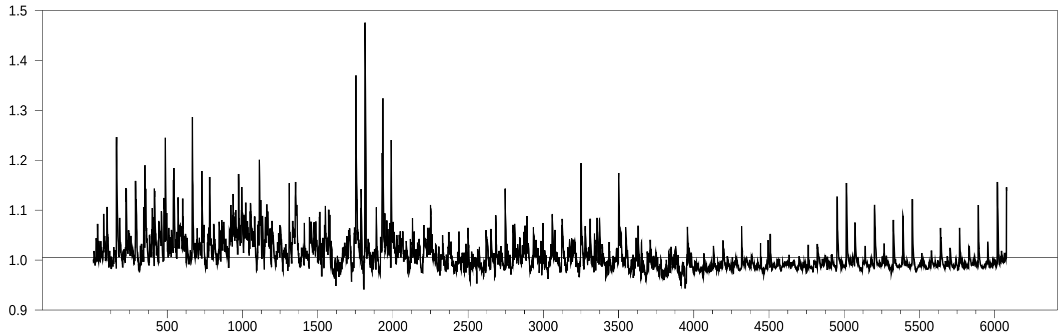


Figure 4.5. Conditional correlation

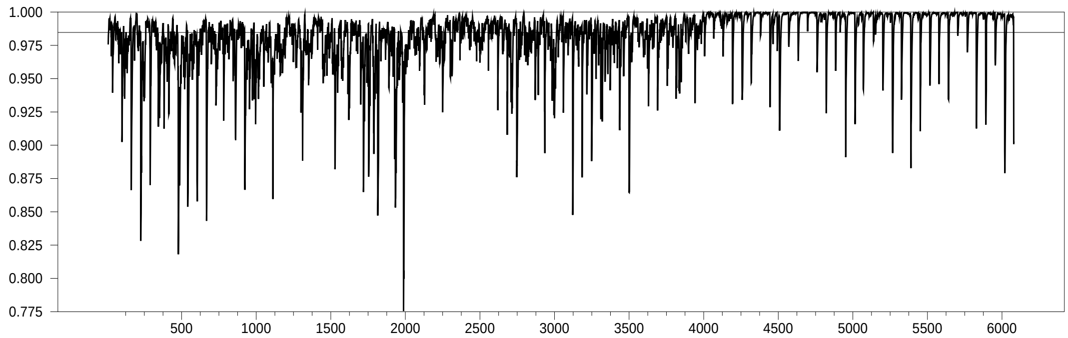


Figure 4.6. Impact surface for the ETF's variance

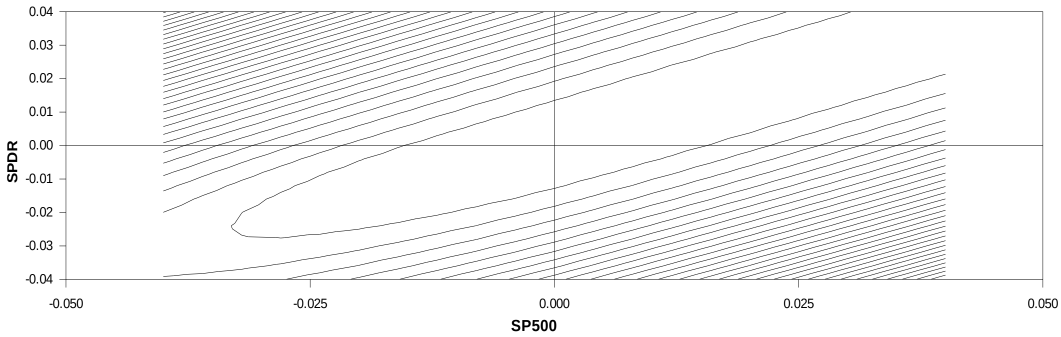


Figure 4.7. Impact surface the benchmark's variance

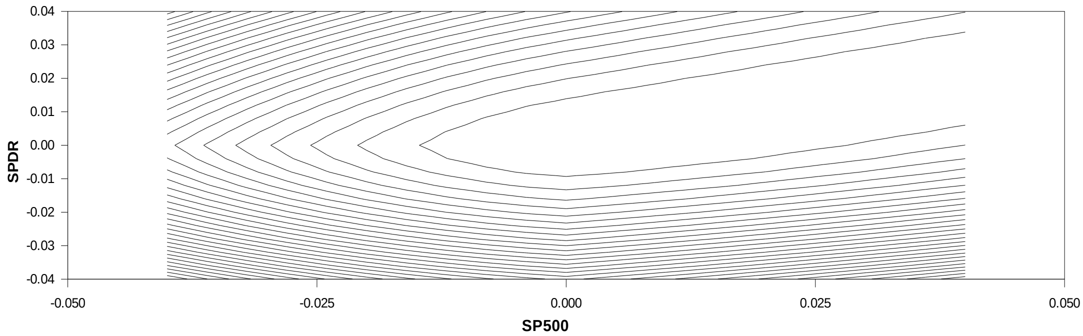
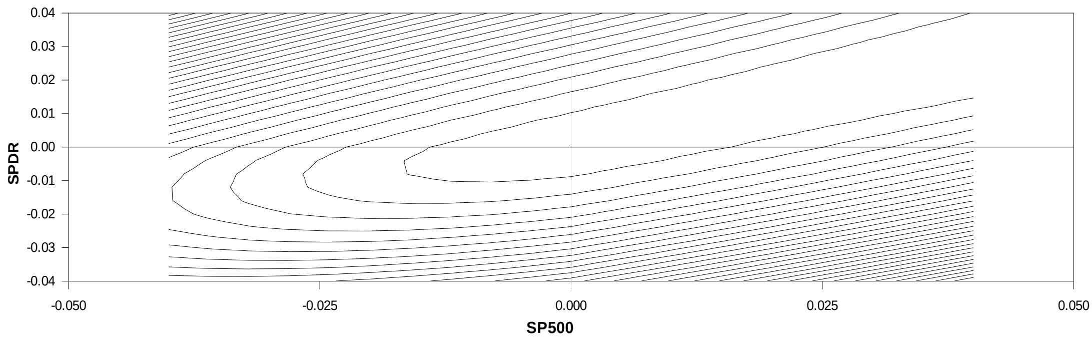


Figure 4.8. Impact surface for the covariance



Tables

Table 4.1. Unit root and cointegration tests

Panel A: Unit root tests					
	ADF	KPSS		ADF	KPSS
f_t	-1.8726	45.05862[a]	$r_{F,t}$	-31.5682[a]	0.0893
i_t	-1.7495	11.9103[a]	$r_{I,t}$	-54.7112[a]	0.0966

Panel B: Johansen cointegration test		
	λ_F	λ_I
Rank = 0	6.294	6.0295
Rank = 1	0.0001	0.0001

Panel C: Summary statistics of returns				
	Mean	Variance	Skew.	Ex. Kurt
$r_{F,t}$	0.0248	0.0153	-0.0928	9.6227
$r_{I,t}$	0.0327	0.0147	-0.2362	8.3487

Panel D: ARCH tests		
	ARCH - 5 lags stat	ARCH - 10 lags stat
$r_{F,t}$	1072.9780[a]	1219.6340[a]
$r_{I,t}$	1103.9980[a]	1257.9900[a]

Table 4.1: Panel A displays a series of unit root tests. ADF is the Augmented Dickey Fuller test for the null of unit root and KPSS is the Kwiatkowski, Philips, Schmidt, and Shin (1992) for the null of stationarity. Panel B shows the output of the cointegration test between the ETF and the benchmark using the Johansen (1988) test. Panel C contains the summary statistics of the returns series (Mean, Variance, Skewness and Excess Kurtosis). Panel D displays the outcomes of the Lagrange Multiplier test for ARCH of Engle (1992). [a] Significant at 1% level, [b] significant at 5% level and [c] significant at 10% level.

Table 4.2. Tests for own variance, cross-variance and covariance asymmetry

Panel A: Own variance

	$\theta_1^{'}$	t-stat	$\theta_1^{''}$	t-stat	$\theta_1^{'''}$	t-stat	Wald Test
$r_{F,t}$	-0.0001	-2.574	-0.0171	-15.703[a]	0.0050	4.513[a]	94.87[a]
$r_{I,t}$	-0.0001	-1.924[c]	-0.0157	-15.737[a]	0.0054	5.263[a]	98.47[a]

Panel B: Cross-variance Asymmetry

	$o_1^{'}$	t-stat	$o_1^{''}$	t-stat	$o_t^{'''}$	t-stat	Wald Test
$r_{F,t}$	-0.0001	-2.125[b]	-0.0161	-15.703[a]	0.0046	4.191[a]	88.17[a]
$r_{I,t}$	-0.0001	-2.0001[b]	-0.0153	-15.703[a]	0.0049	4.876[a]	94.94[a]

Panel C: Covariance Asymmetry

	$\iota_1^{'}$	$\chi_1^{'}$	$\iota_1^{''}$	$\chi_1^{''}$	$\iota_1^{'''}$	$\chi_1^{'''}$	Wald Test
$\sigma_{FI,t}$	-0.0001	0.0000	-0.0286	0.0125	0.0012	0.0039	53.84[a]
t-stat	-2.522[a]	0.701	-7.720[a]	3.319[a]	0.327	1.016	

Table 4.2. Tests for own variance, cross-variance and covariance asymmetry. To test for the effect of asymmetric volatility in the fund and the benchmark we employ Eq. (4.1). The null hypothesis tests the joint effects of shocks of different signs and sizes on the variance of that variable. A Wald test is used to test the joint null of no size and sign. We also test for cross-variance asymmetry using Eq. (4.2) and test for the joint effects using a Wald test. In a similar fashion, we test for covariance asymmetry using Eq. (4.3) and again test for the joint effects employing a Wald test. [a] significant at 1% level, [b] significant at 5% level and [c] significant at 10% level.

Table 4.3. VAR lag selection

lags	AIC	BIC	HQC
0	-1.2936	-1.2914	-1.2928
1	-15.2509	-15.2442	-15.2486
2	-15.4826	-15.4715	-15.4787
3	-15.5542	-15.5387	-15.5488
4	-15.5941	-15.5742	-15.5872
5	-15.6209	-15.5965	-15.6124
6	-15.6383	-15.6095	-15.6283
7	-15.6487	-15.6155*	-15.6371*
8	-15.6495*	-15.612	-15.6365
9	-15.6493	-15.6073	-15.6347

Table 4.3: The table includes three lag-order criteria. The first one is AIC which represents the Akaike Information, the second one is BIC, the Bayesian Information Criteria, and the third criteria is the HQC, the Hannan-Quinn Criterion.

Table 4.4. Estimates of the VECM-GARCH (1,1) model with 8 lags and t-distributed residuals

Panel A: Estimates of the conditional mean

$$r_t = \mu + \sum_{s=1}^{7} \Gamma_s r_{t-s} + \Pi u_{t-1} + \varepsilon_t$$

$$r_t = \begin{bmatrix} r_{F,t} \\ r_{I,t} \end{bmatrix}; \mu = \begin{bmatrix} \mu_F \\ \mu_I \end{bmatrix}; \Gamma_s = \begin{bmatrix} \Gamma_{s,F}^{(F)} & \Gamma_{s,I}^{(F)} \\ \Gamma_{s,F}^{(I)} & \Gamma_{s,I}^{(I)} \end{bmatrix}; \pi = \begin{bmatrix} \pi_I \\ \pi_F \end{bmatrix}; \varepsilon_t = \begin{bmatrix} \varepsilon_{F,t} \\ \varepsilon_{I,t} \end{bmatrix}$$

$$\Gamma_1 = \begin{bmatrix} -0.5509 & 0.5228 \\ (-9.274) & (9.084) \\ 0.0937 & -0.1212 \\ (1.529) & (-2.031) \end{bmatrix} \Gamma_2 = \begin{bmatrix} -0.3666 & 0.3467 \\ (-5.433) & (5.266) \\ 0.0437 & -0.0648 \\ (0.627) & (-0.951) \end{bmatrix} \Gamma_3 = \begin{bmatrix} -0.1928 & 0.1725 \\ (-2.625) & (2.397) \\ 0.0981 & -0.1184 \\ (1.289) & (-1.586) \end{bmatrix}$$

$$\Gamma_4 = \begin{bmatrix} -0.1776 & 0.1758 \\ (-2.469) & (2.478) \\ 0.0199 & -0.0231 \\ (0.269) & (-0.316) \end{bmatrix} \Gamma_5 = \begin{bmatrix} -0.1118 & 0.0849 \\ (-1.619) & (1.259) \\ 0.0241 & -0.0506 \\ (0.341) & (-0.732) \end{bmatrix} \Gamma_6 = \begin{bmatrix} -0.0522 & 0.0471 \\ (-0.767) & (0.702) \\ 0.0112 & -0.0177 \\ (0.162) & (-0.261) \end{bmatrix}$$

$$\Gamma_7 = \begin{bmatrix} -0.0620 & 0.0431 \\ (-1.109) & (0.779) \\ -0.0371 & 0.0189 \\ (-0.653) & (0.337) \end{bmatrix} \quad \mu = \begin{bmatrix} 0.0000 \\ (0.009) \\ 0.0000 \\ (0.097) \end{bmatrix} \quad \Pi = \begin{bmatrix} -0.0001 \\ (-2.316) \\ -0.0001 \\ (-2.190) \end{bmatrix}$$

Panel B: Estimates of the conditional variance

$$H_t = C_0^{*'}C_0^{*} + A_{11}^{*'}\varepsilon_{t-1}\varepsilon_{t-1}^{'}A_{11}^{*} + B_{11}^{*'}H_{t-1}B_{11}^{*} + D_{11}^{*'}\tau_{t-1}\tau_{t-1}^{'}D_{11}^{*}$$

$$C_0^* = \begin{bmatrix} c_{11}^* & c_{12}^* \\ 0 & c_{22}^* \end{bmatrix};\ A_{11}^* = \begin{bmatrix} \alpha_{11}^* & \alpha_{12}^* \\ \alpha_{21}^* & \alpha_{22}^* \end{bmatrix};\ B_{11}^* = \begin{bmatrix} \beta_{11}^* & \beta_{12}^* \\ \beta_{21}^* & \beta_{22}^* \end{bmatrix}; D_{11}^* = \begin{bmatrix} \delta_{11}^* & \delta_{12}^* \\ \delta_{21}^* & \delta_{22}^* \end{bmatrix}; \tau_t^2 = \begin{bmatrix} \tau_{F,t}^2 \\ \tau_{I,t}^2 \end{bmatrix}$$

$$C_0^* = \begin{bmatrix} 0.0011 & 0.0012 \\ (8.098) & (8.389) \\ & 0.0001 \\ & (6.949) \end{bmatrix} \quad A_{11}^* = \begin{bmatrix} -0.1059 & -0.7391 \\ (-0.994) & (-6.218) \\ 0.2128 & 0.8493 \\ (1.970) & (7.110) \end{bmatrix}$$

$$B_{11}^* = \begin{bmatrix} 0.9865 & 0.1403 \\ (22.041) & (2.908) \\ -0.0311 & 0.8125 \\ (-0.704) & (16.999) \end{bmatrix} \quad D_{11}^* = \begin{bmatrix} 0.1711 & 0.1171 \\ (1.488) & (0.883) \\ 0.2313 & 0.2872 \\ (2.011) & (2.141) \end{bmatrix}$$

Shape = 3.8524
(27.012)

Panel C: Standardised residuals tests

	Mean	Var.	Skew.	Ex.Kurt	J-B	ARCH	L-Box
Fund	-0.066	0.771	-0.526	1.972	1264.164[a]	10.567	2.031
Index	-0.005	0.766	-0.495	1.758	1030.915[a]	1.797	1.785

Table 4.4: Panels A and B of table display the estimates of the VECM-GARCH (1,1) model with 8 lags and conditional Student t distributed residuals. Panel C shows the summary statistics of the standardised residuals for the index and the fund. Var. represents the variance, Skew. represents the skewness, Ex. Kurt characterises the excess kurtosis, J-B is the Jarque-Bera test for normality. ARCH is the Engle's test for ARCH and L-Box is the Ljung-Box test for autocorrelation in the residuals [a] significant at 1% level, [b] significant at 5% level and [c] significant at 10% level.

Table 4.5. Model specification tests

Tests	Test-stat	P-value
Multivariate GARCH	60.0200	0.0663
BEKK cross effects	38.5274	0.0000
Asymmetry	10922.9889	0.0000
Short run causality in mean index-fund	22.6945	0.0000
Short run causality in mean fund-index	14.2029	0.0000
Short and long run causality in mean index-fund	19.5571	0.0000
Short and long run causality in mean fund-index	12.5267	0.0000
Short run NLGC in volatility fund-index	0.3527	0.7242
Short run NLGC in volatility index-fund	1.4545	0.1457
Short and long run NLGC in volatility fund-index	0.95385	0.3401
Short and long run NLGC in volatility index-fund	3.0122	0.0025

Table 4.5: Displays a series of model specification tests. The first test is the Lagrange multiplier test for multivariate ARCH effects. The second test is a Wald test for the null hypothesis of diagonality in the variance structure of the BEKK-GARCH model. The third test is a Wald test for the null of no asymmetry in the GARCH-BEKK model. The fourth and the fifth rows display the results of the Wald tests for short run Granger causality in mean for the index and the ETF. Analogously, he fifth and the sixth rows show the outcomes of the Wald tests for the joint null of short run and long run causality in mean for the index and the ETF. Rows seventh to eleventh exhibit the results of the Wald tests for short run causality in variance and for the short and long run causality in variance for the index and the fund.

Table 4.6. Daily tracking difference statistics: In-sample

Tracking Difference	Mean	Variance	Skewness	Kurtosis
TD_{OLS}	-0.000079	0.000006	0.299633	19.469014
TD_{Unity}	-0.000080	0.000006	0.296966	19.339472
$TD_{VECM-GARCH}$	-0.000081	0.000006	0.510244	24.083213

Table 4.6: TD is the Tracking Difference which we define as the daily divergences that might appear between the performance of the fund and the benchmark, in terms of daily returns. We compute the TD defined in Eq. (4.6) using the three different definitions of the hedge ratio.

Table 4.7. Daily tracking difference statistics: Out-of-sample

Tracking Difference	Mean	Variance	Skewness	Kurtosis
TD_{OLS}	-0.000084	0.00000058	1.77801	30.960578
TD_{Unity}	-0.000086	0.00000058	1.74887	30.737525
$TD_{VECM-GARCH}$	-0.000080	0.00000060	1.64630	29.949412

Table 4.7: TD is the Tracking Difference which we define as the daily divergences that might appear between the performance of the fund and the benchmark, in terms of daily returns. We compute the TD defined in Eq. (4.6) using the three different definitions of the hedge ratio.

Appendix

Table A.4.1 Estimates of the VECM-GARCH (1, 1) model with 7 lags and t-distributed residuals

Panel A: Estimates of the conditional mean

$$r_t = \mu + \sum_{s=1}^{6} \Gamma_s r_{t-s} + \Pi u_{t-1} + \varepsilon_t$$

$$r_t = \begin{bmatrix} r_{F,t} \\ r_{I,t} \end{bmatrix}; \mu = \begin{bmatrix} \mu_F \\ \mu_I \end{bmatrix}; \Gamma_s = \begin{bmatrix} \Gamma_{s,F}^{(F)} & \Gamma_{s,I}^{(F)} \\ \Gamma_{s,F}^{(I)} & \Gamma_{s,I}^{(I)} \end{bmatrix}; \pi = \begin{bmatrix} \pi_I \\ \pi_F \end{bmatrix}; \varepsilon_t = \begin{bmatrix} \varepsilon_{F,t} \\ \varepsilon_{I,t} \end{bmatrix}$$

$$\Gamma_1 = \begin{bmatrix} -0.5464 & 0.5191 \\ (-10.450) & (10.220) \\ 0.0950 & -0.1218 \\ (1.768) & (-2.321) \end{bmatrix} \quad \Gamma_2 = \begin{bmatrix} -0.3621 & 0.3433 \\ (-6.486) & (6.364) \\ 0.0429 & -0.0629 \\ (0.741) & (-1.213) \end{bmatrix}.$$

$$\Gamma_3 = \begin{bmatrix} -0.1881 & 0.1686 \\ (-2.966) & (2.683) \\ 0.0959 & -0.1154 \\ (1.456) & (-1.761) \end{bmatrix} \quad \Gamma_4 = \begin{bmatrix} -0.1716 & 0.1706 \\ (-2.634) & (2.651) \\ -0.0179 & -0.0203 \\ (-0.268) & (-0.305) \end{bmatrix} \quad \Gamma_5 = \begin{bmatrix} -0.1031 & 0.0776 \\ (-1.880) & (1.433) \\ 0.0227 & -0.0479 \\ (0.405) & (-0.860) \end{bmatrix}$$

$$\Gamma_6 = \begin{bmatrix} -0.0344 & 0.0306 \\ (-0.786) & (0.700) \\ 0.0145 & -0.1978 \\ (0.328) & (-0.445) \end{bmatrix} \quad \mu = \begin{bmatrix} -0.0025 \\ (-0.123) \\ 0.0000 \\ (0.056) \end{bmatrix} \quad \Pi = \begin{bmatrix} -0.0001 \\ (-4.721) \\ -0.0001 \\ (-4.362) \end{bmatrix}$$

Panel B: Estimates of the conditional variance

$$H_t = C_0^{*'}C_0^* + A_{11}^{*'}\varepsilon_{t-1}\varepsilon_{t-1}^{'}A_{11}^* + B_{11}^{*'}H_{t-1}B_{11}^* + D_{11}^{*'}\tau_{t-1}\tau_{t-1}^{'}D_{11}^*$$

$$C_0^* = \begin{bmatrix} c_{11}^* & c_{12}^* \\ 0 & c_{22}^* \end{bmatrix};\ A_{11}^* = \begin{bmatrix} \alpha_{11}^* & \alpha_{12}^* \\ \alpha_{21}^* & \alpha_{22}^* \end{bmatrix};\ B_{11}^* = \begin{bmatrix} \beta_{11}^* & \beta_{12}^* \\ \beta_{21}^* & \beta_{22}^* \end{bmatrix};D_{11}^* = \begin{bmatrix} \delta_{11}^* & \delta_{12}^* \\ \delta_{21}^* & \delta_{22}^* \end{bmatrix};\tau_t^2 = \begin{bmatrix} \tau_{F,t}^2 \\ \tau_{I,t}^2 \end{bmatrix}$$

$$C_0^* = \begin{bmatrix} 0.011 & 0.0012 \\ (7.283) & (7.517) \\ & -0.0001 \\ & (-7.324) \end{bmatrix} \quad A_0^* = \begin{bmatrix} -0.1042 & -0.7361 \\ (-1.717) & (-11.725) \\ 0.2097 & 0.8449 \\ (3.494) & (13.649) \end{bmatrix}$$

$$B_{11}^* = \begin{bmatrix} 0.9854 & 0.1389 \\ (34.109) & (4.733) \\ -0.0299 & 0.8139 \\ (-1.093) & (28.822) \end{bmatrix} \quad D_{11}^* = \begin{bmatrix} 0.1706 & 0.1117 \\ (1.609) & (0.920) \\ 0.2332 & 0.2942 \\ (2.189) & (2.383) \end{bmatrix}$$

Shape = 3.8604
(28.4234)

Table A.4.1: Panels A and B of table display the estimates of the VECM-GARCH (1,1) model with 7 lags and conditional Student t distributed residuals.

Table A.4.2. Model specification tests

Model specification tests		
	Test-stat	P-value
Multivariate GARCH	60.3000	0.0633
BEKK cross effects	49.3286	0.0000
Asymmetry	46.1196	0.0000
Short run causality in mean index-fund	83.8823	0.0000
Short run causality in mean fund-index	3.5966	0.0014
Short and long run causality in mean index-fund	98.7603	0.0000
Short and long run causality in mean fund-index	21.2503	0.0000
Short run NLGC in volatility index-fund	5.4125	0.0000
Short run NLGC in volatility fund-index	1.8854	0.0593
Short and long run NLGC in volatility index-fund	5.5148	0.0000
Short and long run NLGC in volatility fund-index	1.8928	0.0593

Table A.4.2: Displays a series of model specification tests. The first test is the Lagrange multiplier test for multivariate ARCH effects. The second test is a Wald test for the null hypothesis of diagonality in the variance structure of the BEKK-GARCH model. The third test is a Wald test for the null of no asymmetry in the GARCH-BEKK model. The fourth and the fifth rows display the results of the Wald tests for short run Granger causality in mean for the index and the ETF. Analogously, he fifth and the sixth rows show the outcomes of the Wald tests for the joint null of short run and long run causality in mean for the index and the ETF. Rows seventh to eleventh exhibit the results of the Wald tests for short run causality in variance and for the short and long run causality in variance for the index and the fund.

Table A.4.3. Estimates of VECM-GARCH (1,1) model with 7 lags and normally distributed residuals

Panel A: Estimates of the conditional mean

$$r_t = \mu + \sum_{s=1}^{6} \Gamma_s r_{t-s} + \Pi u_{t-1} + \varepsilon_t$$

$$r_t = \begin{bmatrix} r_{F,t} \\ r_{I,t} \end{bmatrix}; \mu = \begin{bmatrix} \mu_F \\ \mu_I \end{bmatrix}; \Gamma_s = \begin{bmatrix} \Gamma_{s,F}^{(F)} & \Gamma_{s,I}^{(F)} \\ \Gamma_{s,F}^{(I)} & \Gamma_{s,I}^{(I)} \end{bmatrix}; \pi = \begin{bmatrix} \pi_I \\ \pi_F \end{bmatrix}; \varepsilon_t = \begin{bmatrix} \varepsilon_{F,t} \\ \varepsilon_{I,t} \end{bmatrix}$$

$$\Gamma_1 = \begin{bmatrix} -0.5329 & 0.5174 \\ (-8.957) & (8.978) \\ 0.0969 & -0.1119 \\ (1.479) & (-1.734) \end{bmatrix} \quad \Gamma_2 = \begin{bmatrix} -0.3844 & 0.3809 \\ (-5.329) & (5.352) \\ 0.0225 & -0.0272 \\ (0.282) & (-0.345) \end{bmatrix} \quad \Gamma_3 = \begin{bmatrix} -0.2345 & 0.2251 \\ (-3.103) & (3.017) \\ 0.0664 & -0.0772 \\ (0.779) & (-0.917) \end{bmatrix}$$

$$\Gamma_4 = \begin{bmatrix} -0.1692 & 0.1644 \\ (-2.121) & (2.107) \\ 0.0308 & -0.0388 \\ (0.356) & (-0.456) \end{bmatrix} \quad \Gamma_5 = \begin{bmatrix} -0.1278 & 0.0928 \\ (-1.538) & (1.137) \\ -0.0074 & -0.0272 \\ (-0.085) & (-0.318) \end{bmatrix} \quad \Gamma_6 = \begin{bmatrix} -0.0419 & 0.0384 \\ (-0.659) & (0.614) \\ 0.0272 & -0.0339 \\ (0.399) & (-0.505) \end{bmatrix}$$

$$\mu = \begin{bmatrix} 0.0002 \\ (0.914) \\ 0.0001 \\ (0.405) \end{bmatrix} \quad \Pi = \begin{bmatrix} -0.0004 \\ (-1.934) \\ -0.00004 \\ (-1.509) \end{bmatrix}$$

Panel B: Estimates of the conditional variance

$$H_t = C_0^{*'}C_0^* + A_{11}^{*'}\varepsilon_{t-1}\varepsilon_{t-1}^{'}A_{11}^* + B_{11}^{*'}H_{t-1}B_{11}^* + D_{11}^{*'}\tau_{t-1}\tau_{t-1}^{'}D_{11}^*$$

$$C_0^* = \begin{bmatrix} c_{11}^* & c_{12}^* \\ 0 & c_{22}^* \end{bmatrix}; A_{11}^* = \begin{bmatrix} \alpha_{11}^* & \alpha_{12}^* \\ \alpha_{21}^* & \alpha_{22}^* \end{bmatrix}; B_{11}^* = \begin{bmatrix} \beta_{11}^* & \beta_{12}^* \\ \beta_{21}^* & \beta_{22}^* \end{bmatrix}; D_{11}^* = \begin{bmatrix} \delta_{11}^* & \delta_{12}^* \\ \delta_{21}^* & \delta_{22}^* \end{bmatrix}; \tau_t^2 = \begin{bmatrix} \tau_{F,t}^2 \\ \tau_{I,t}^2 \end{bmatrix}$$

$$C_0^* = \begin{bmatrix} 0.0016 & 0.0013 \\ (9.951) & (9.590) \\ & 0.0004 \\ & (9.148) \end{bmatrix} \quad A_{11}^* = \begin{bmatrix} -0.1033 & -0.6002 \\ (-1.026) & (-5.408) \\ 0.1858 & 0.6842 \\ (0.185) & (6.272) \end{bmatrix}$$

$$B_{11}^* = \begin{bmatrix} 0.9488 & 0.0905 \\ (14.801) & (1.391) \\ 0.0019 & 0.8574 \\ (0.031) & (13.426) \end{bmatrix} \quad D_{11}^* = \begin{bmatrix} 0.2825 & 0.2282 \\ (1.837) & (1.318) \\ 0.0910 & 0.1448 \\ (0.593) & (0.835) \end{bmatrix}$$

Table A.4.3: Panels A and B of table display the estimates of the VECM-GARCH (1,1) model with 7 lags and conditional normally distributed residuals.

Table A.4.4. Model specification tests

Tests	Test-stat	P-value
Multivariate GARCH	76.7800	0.0022
BEKK cross effects	23.3456	0.0000
Asymmetry	44.7820	0.0000
Short run causality in mean index-fund	13.8771	0.0000
Short run causality in mean fund-index	0.6050	0.7266
Short and long run in mean causality index-fund	11.8995	0.0000
Short and long run causality in mean fund-index	1.4551	0.1783
Short run NLGC in volatility index-fund	2.8405	0.0045
Short run NLGC in volatility fund-index	0.8566	0.3937
Short and long run NLGC in volatility index-fund	2.7741	0.0055
Short and long run NLGC in volatility fund-index	0.8566	0.3916

Table A.4.4: Displays a series of model specification tests. The first test is the Lagrange multiplier test for multivariate ARCH effects. The second test is a Wald test for the null hypothesis of diagonality in the variance structure of the BEKK-GARCH model. The third test is a Wald test for the null of no asymmetry in the GARCH-BEKK model. The fourth and the fifth rows display the results of the Wald tests for short run Granger causality in mean for the index and the ETF. Analogously, he fifth and the sixth rows show the outcomes of the Wald tests for the joint null of short run and long run causality in mean for the index and the ETF. Rows seventh to eleventh exhibit the results of the Wald tests for short run causality in variance and for the short and long run causality in variance for the index and the fund.

Table A.4.5. Estimates of the VECM-GARCH (1,1) model with 8 lags with normally distributed residuals

Panel A: Estimates of the Conditional Mean model

$$r_t = \mu + \sum_{s=1}^{7} \Gamma_s r_{t-s} + \Pi u_{t-1} + \varepsilon_t$$

$$r_t = \begin{bmatrix} r_{F,t} \\ r_{I,t} \end{bmatrix}; \mu = \begin{bmatrix} \mu_F \\ \mu_I \end{bmatrix}; \Gamma_s = \begin{bmatrix} \Gamma^{(F)}_{s,F} & \Gamma^{(F)}_{s,I} \\ \Gamma^{(I)}_{s,F} & \Gamma^{(I)}_{s,I} \end{bmatrix}; \pi = \begin{bmatrix} \pi_I \\ \pi_F \end{bmatrix}; \varepsilon_t = \begin{bmatrix} \varepsilon_{F,t} \\ \varepsilon_{I,t} \end{bmatrix}$$

$$\Gamma_1 = \begin{bmatrix} -0.5357 & 0.5198 \\ (-8.341) & (8.371) \\ 0.0973 & -0.1128 \\ (1.355) & (-1.596) \end{bmatrix} \quad \Gamma_2 = \begin{bmatrix} -0.3901 & 0.3862 \\ (-5.467) & (5.520) \\ 0.0215 & -0.0264 \\ (0.271) & (-0.339) \end{bmatrix} \quad \Gamma_3 = \begin{bmatrix} -0.2424 & 0.2325 \\ (-3.152) & (3.048) \\ 0.0651 & -0.0762 \\ (0.745) & (-0.880) \end{bmatrix}$$

$$\Gamma_4 = \begin{bmatrix} -0.1797 & 0.1746 \\ (-2.233) & (2.220) \\ 0.0291 & -0.0374 \\ (0.334) & (-0.436) \end{bmatrix} \quad \Gamma_5 = \begin{bmatrix} -0.1456 & 0.1098 \\ (-1.623) & (1.249) \\ -0.0135 & -0.0217 \\ (-0.143) & (-0.234) \end{bmatrix} \quad \Gamma_6 = \begin{bmatrix} -0.0691 & 0.0644 \\ (-0.894) & (0.837) \\ 0.0158 & -0.0237 \\ (0.195) & (-0.293) \end{bmatrix}$$

$$\Gamma_7 = \begin{bmatrix} -0.0520 & 0.0465 \\ (-0.743) & (0.069) \\ -0.0265 & 0.0201 \\ (-0.363) & (0.2769) \end{bmatrix} \quad \mu = \begin{bmatrix} 0.0002 \\ (0.482) \\ 0.0001 \\ (0.287) \end{bmatrix} \quad \Pi = \begin{bmatrix} -0.0003 \\ (-2.302) \\ -0.00003 \\ (-0.832) \end{bmatrix}$$

Panel B: Estimates of the conditional variance

$$H_t = C_0^{*'}C_0^* + A_{11}^{*'}\varepsilon_{t-1}\varepsilon_{t-1}^{'}A_{11}^* + B_{11}^{*'}H_{t-1}B_{11}^* + D_{11}^{*'}\tau_{t-1}\tau_{t-1}^{'}D_{11}^*$$

$$C_0^* = \begin{bmatrix} c_{11}^* & c_{12}^* \\ 0 & c_{22}^* \end{bmatrix};\ A_{11}^* = \begin{bmatrix} \alpha_{11}^* & \alpha_{12}^* \\ \alpha_{21}^* & \alpha_{22}^* \end{bmatrix};\ B_{11}^* = \begin{bmatrix} \beta_{11}^* & \beta_{12}^* \\ \beta_{21}^* & \beta_{22}^* \end{bmatrix}; D_{11}^* = \begin{bmatrix} \delta_{11}^* & \delta_{12}^* \\ \delta_{21}^* & \delta_{22}^* \end{bmatrix}; \tau_t^2 = \begin{bmatrix} \tau_{F,t}^2 \\ \tau_{I,t}^2 \end{bmatrix}$$

$$C_0^* = \begin{bmatrix} 0.0013 & 0.0014 \\ (9.856) & (9.166) \\ & 0.0004 \\ & (10.494) \end{bmatrix} \quad A_{11}^* = \begin{bmatrix} -0.1097 & -0.6087 \\ (-1.083) & (-5.405) \\ 0.1916 & 0.6919 \\ (1.935) & (6.272) \end{bmatrix}$$

$$B_{11}^* = \begin{bmatrix} 0.9534 & 0.0962 \\ (14.527) & (1.475) \\ -0.0026 & 0.8517 \\ (-0.039) & (13.273) \end{bmatrix} \quad D_{11}^* = \begin{bmatrix} 0.2775 & 0.2220 \\ (1.711) & (1.214) \\ 0.0958 & 0.1508 \\ (0.594) & (0.826) \end{bmatrix}$$

Shape = 3.8604
(28.4234)

Table A.4.5: Panels A and B of table display the estimates of the VECM-GARCH (1,1) model with 8 lags and conditional normally distributed residuals.

Table A.4.6. Model specification tests

Tests	Test-stat	P-value
Multivariate GARCH	76.9700	0.0021
Short run causality in mean index-fund	7.4198	0.0000
Short run causality in mean fund-index	12.1215	0.0000
Short and long run causality in mean index-fund	6.3599	0.0000
Short and long run causality in mean fund-index	10.3978	0.0000
BEKK cross effects	30.6422	0.0000
Asymmetry	7817.6944	0.0000
Short and long run NLGC in volatility fund-index	0.9603	0.3368
Short and long run NLGC in volatility index-fund	2.5688	0.0102
Short run NLGC in volatility fund-index	0.0255	0.9796
Short run NLGC in volatility index-fund	0.73755	0.4607

Table A.4.6: Displays a series of model specification tests. The first test is the Lagrange multiplier test for multivariate ARCH effects. The second test is a Wald test for the null hypothesis of diagonality in the variance structure of the BEKK-GARCH model. The third test is a Wald test for the null of no asymmetry in the GARCH-BEKK model. The fourth and the fifth rows display the results of the Wald tests for short run Granger causality in mean for the index and the ETF. Analogously, he fifth and the sixth rows show the outcomes of the Wald tests for the joint null of short run and long run causality in mean for the index and the ETF. Rows seventh to eleventh exhibit the results of the Wald tests for short run causality in variance and for the short and long run causality in variance for the index and the fund.

Chapter 5

Conclusions and Directions for Future Research

5.1. Conclusions

An ETF is an investment fund set to track the performance of a benchmark index. In theory, the ETF's tracking process should be easy and costless. However, in practice, the fund's replication process faces market frictions which potentially give rise to tracking errors. Since the tracking error measures how accurately the fund manager replicates the performance of the benchmark index, it constitutes an important metric for investors and money managers when evaluating ETF's performance.

In this thesis, we studied the ETF's tracking performance in three different perspectives, which at the same time constitute the three main chapters of the thesis. In Chapter 2, we developed a framework to assess the quality of the ETF's tracking that accounts for the stochastic nature that characterises financial data. Essentially, this approach consists of computing the TE using an error correction model that accounts for the cointegrating relationship between the ETF and the underlying benchmark. Using panel techniques, we concluded that the passively managed U.S. equity and debt ETFs included in our sample and their underlying benchmarks appear to be cointegrated. We also observed that, in general, equity ETFs track their underlying indices more precisely than their debt counterparts. Overall, the results imply that omitting the cointegrating relationship between the ETF and the index tends to result in inferior estimates of the tracking errors. Since it is common practice for investors and money managers to select ETFs based on their tracking performance, we argue that, to avoid erroneous investment choices, the tracking errors should be based on methods that take into consideration the nature of the data. We illustrated the consequences of incorrectly specifying the model used to estimate the TE using a series of Monte-Carlo simulations, which show that the ordering and the constituents of the portfolio selected to include those funds with the smallest tracking error varies with the model used to gauge them.

Using the appropriately computed tracking error, Chapter 3 focused on the role of liquidity on the ETF's tracking performance. To study the impact of liquidity on the tracking errors, the previous literature has mainly employed generic proxies of market liquidity. However, in this chapter we conjectured that an evaluation of ETF's liquidity requires more specific proxies which account for the special structure of the ETF. We distinguished between primary liquidity, which relates to ETF's creation-redemption processes, and secondary liquidity, which is linked to the market activity of the ETF. As a result, we constructed eight alternative proxies to capture the main aspects of ETF's liquidity. Our findings suggested that the illiquidity resulting from the creation-redemption processes plays a key role in determining the tracking quality of the ETF, regardless of the asset class tracked by the fund. This outcome illustrates the difficulties experienced by the authorised participants during the creation and redemption of ETF’s shares due to the disparities between the price of the ETF and the net asset value of the underlying securities. Moreover, the outcomes of the cross-sectional analysis indicated that the accuracy of the ETF’s tracking improves as the turnover increases, and it decreases as the bid-ask spreads widen. The volatility of the fund appeared positively related to the tracking errors. We also studied whether the tracking performance is related to the asset class tracked by the fund, finding evidence that the illiquidity originated in the creation-redemption processes remains a key determinant, regardless of the underlying securities tracked by the ETF. In terms of the secondary liquidity however, the fund turnover appears to be negatively related to the tracking error, regardless of the asset class tracked by the fund. For equity ETFs the evidence suggested that the bid-ask spread is positively related to the tracking error in a statistically significant fashion. There is some evidence that the duration of the bond portfolio is positively related to the tracking performance of debt ETFs.

Chapter 4 studied the time series dimension of the ETF’s tracking performance, assessing the impact of short-run departures from the long-run equilibrium between the ETF and the benchmark when the conditional second moments are time varying. Since we allowed the variance-covariance matrix to vary with time, the optimal hedge ratio may also display time variation. Consequently, the ETF’s tracking performance, measured by the tracking difference and the tracking error, may also display time variation and asymmetry in the response to news. We used a battery of diagnostic tests to ensure the adequacy of the model. Our results suggested that there exists a bidirectional causal relationship between the returns of the ETF and those of the benchmark, and their associated volatilities, which implies that news about the benchmark and the fund might drive the movements in the hedge ratio. Moreover, in the presence of asymmetry, there is the possibility that the hedge ratio will display asymmetric response to news unless any variance and covariance asymmetry observed in the data is offsetting.

To assess the advantages of allowing for time variation and asymmetry in the variance-covariance matrix, we employed three different definitions of the optimal hedge ratio. The evidence suggests that models that omit the time series dynamics of the data, such as cointegration, time variation and asymmetry in the covariance matrix, result in underestimation of the tracking performance, measured in terms of the tracking difference. Consequently, using these simpler models to assess the tracking performance might lead investors to suboptimal portfolio choices. It is worth noting that, in any case, the daily tracking differences observed were relatively small in magnitude, which implies that the benefits of allowing the hedge ratio over time seems minimal for the most heavily traded ETF in the world, SPY ETF. However, in future research this framework can be expanded to a wider sample of ETFs and to less traded ETFs, which might change the outcomes dramatically.

Given the recent growth in passively managed index funds this research has implications for institutional and retail investors, money managers, sovereign wealth funds and insurance companies amongst others, who hold ETFs to benefit from the risk diversification, transparency, intraday liquidity and low trading costs provided by these funds.

5.2. Directions for future research

5.2.1. Benefits of international diversification

I am currently engaged in a project with my supervisor Professor Ólan Henry and Professor Nilss Olekalns from The University of Melbourne. This project explores the use of ETFs as a vehicle to simplify the diversification of risk. The working title of the project is "Achieving the Benefits of International Diversification using Domestically Listed Securities". The project exploits modern asset pricing theory to compare and contrast the outcomes from diversification using a global physical position via the international markets with the outcomes achieved using ETFs listed only in the domestic economy of the investor. The key to the success of the ETF's approach is to avoid the costs and risks associated with international diversification via more typical multi-country asset investments.

5.2.1. Purchasing Power Parity

Based on the paper "Purchasing Power Parity Tests in Cointegrated Panels" by Pedroni (2001), the idea is to test whether the Purchasing Power Parity theory, hereafter PPP, holds in the ETF's market. Initially, we will use various ETFs listed in different stock markets (developed countries) and with a common currency. The second step will be to test the PPP using the ETFs listed in different developed markets but with different currencies. Since it has been documented that stock markets in emerging countries are less efficient than those of developed countries, we could also test the PPP using ETFs listed in both markets and compare the outcomes.

Bibliography

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

Ackert, L.F., and Tian, Y.S. (2000). Arbitrage and valuation in the market for Standard and Poor's Depositary Receipts. *Financial Management* (2000): 71-87.

Ackert, L. F., and Tian, Y. S. (2008). Arbitrage, liquidity, and the valuation of exchange traded funds. *Financial markets, institutions and instruments*, 17(5), 331-362.

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

Alexander, C., and Barbosa, A. (2008). Hedging index exchange traded funds. *Journal of Banking & Finance*, *32*(2), 326-337.

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of financial Economics*, *17*(2), 223-249.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets, 5(1), 31-56.*

Baillie, R. T., and Myers, R. J. (1991). Bivariate GARCH estimation of the optimal commodity futures hedge. *Journal of Applied Econometrics*, *6*(2), 109-124.

Barnhart, S. W., and Rosenstein, S. (2010). Exchange-Traded Fund Introductions and Closed-End Fund Discounts and Volume. *Financial Review*, *45*(4), 973-994.

Bertone, S., Paeglis, I., and Ravi, R. (2015). (How) has the market become more efficient? *Journal of Banking and Finance*, 54, 72-86.

Billio, M., Donadelli, M., Paradiso, A., and Riedel, M. (2017). Which market integration measure?. *Journal of Banking & Finance,* 76, 150-174.

Blitz, D., Huij, J., and Swinkels, L. (2012). The Performance of European Index Funds and Exchange-Traded Funds. *European Financial Management*, 18(4), 649-662.

Blume, M. E., and Edelen, R. M. (2004). S&P 500 indexers, tracking error, and liquidity. The Journal of Portfolio Management, 30(3), 37-46.

Breitung, J. (2001). The local power of some unit root tests for panel data. In *Nonstationary panels, panel cointegration, and dynamic panels* (pp. 161-177). Emerald Group Publishing Limited.

Breitung, J., and Das, S. (2005). Panel unit root tests under cross-sectional dependence. *Statistica Neerlandica*, 59(4), 414-433.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307-327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, 542-547.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of political Economy*, *96*(1), 116-131.

Bollerslev, T., and Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews*, *11*(2), 143-172.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The review of economics and statistics*, 498-505.

Broman, M. S. (2016). Liquidity, style investing and excess comovement of exchange-traded fund returns. *Journal of Financial Markets, 30*, 27-53.

Broman, M. S., and Shum, P. (2018). Relative Liquidity, Fund Flows and Short-Term Demand: Evidence from Exchange-Traded Funds. *Financial Review*, *53*(1), 87-115.

Broms, T. J., and Gastineau, G. L. (2007). The Development of Improved Exchange-Traded Funds in the United States. *The Brookings Institution*, 193-209. Institution, Washington, D.C.September 12, 2006.

Brooks, C., Rew, A. G., and Ritson, S. (2001). A trading strategy based on the lead–lag relationship between the spot index and futures contract for the FTSE 100. *International Journal of Forecasting*, *17*(1), 31-44.

Brooks, C., and Chong, J. (2001). The cross-currency hedging performance of implied versus statistical forecasting models. *Journal of Futures Markets*, *21*(11), 1043-1069.

Brooks, C., Henry, O. T., and Persand, G. (2002). The effect of asymmetries on optimal hedge ratios. *The Journal of Business*, *75*(2), 333-352.

Buetow, G. W., and Henderson, B. J. (2012). An empirical analysis of exchange-traded funds. *The Journal of Portfolio Management*, 38(4), 112-127.

Chacko, G., Das, S., and Fan, R. (2016). An index-based measure of liquidity. *Journal of Banking & Finance*, 68, 162-178.

Caporale, G. M., Ali, F. M., and Spagnolo, N. (2015). Exchange rate uncertainty and international portfolio flows: A multivariate GARCH-in-mean approach. *Journal of International Money and Finance*, *54*, 70-92.

Caporin, M., and McAleer, M. (2012). Do we really need both BEKK and DCC? A tale of two multivariate GARCH models. *Journal of Economic Surveys*, *26*(4), 736-751.

Cecchetti, S. G., Cumby, R. E., and Figlewski, S. (1988). Estimation of the optimal futures hedge. *The Review of Economics and Statistics*, 623-630.

Charupat, N., and Miu, P. (2011). The pricing and performance of leveraged exchange-traded funds. *Journal of Banking & Finance*, *35*(4), 966-977.

Charupat, N., and Miu, P. (2013). The pricing efficiency of leveraged exchange-traded funds: evidence from the US market. *Journal of Financial Research*, *36*(2), 253-278.

Chordia, T., Roll, R., and Subrahmanyam, A. (2008). Liquidity and market efficiency. *Journal of Financial Economics*, *87*(2), 249-268.

Chordia, T., Sarkar, A., and Subrahmanyam, A. (2004). An empirical analysis of stock and bond market liquidity. *The Review of Financial Studies*, *18*(1), 85-129.

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

Chu, Q. C., Hsieh, W. L. G., and Tse, Y. (1999). Price discovery on the S&P 500 index markets: An analysis of spot index, index futures, and SPDRs. *International Review of Financial Analysis*, *8*(1), 21-34.

Cotter, J., and Hanly, J. (2006). Reevaluating hedging performance. *Journal of Futures Markets*, *26*(7), 677-702.

Curcio, R. J., J.M. Lipka, and J.H. Thornton Jr. (2004). Cubes and individual investors. *Financial Services Review*. 13(2), 123-139.

Deville, L. (2008), Exchange traded funds: History, trading, and research. *In Handbook of Financial Engineering* (pp. 67-98). Springer US.

Drenovak, M., Urošević, B., and Jelic, R. (2014). European fixed income ETFs: tracking errors and the sovereign debt crisis. *European Financial Management*, 20(5), 958-994.

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

Elton, E., M. Gruber, G. Comer and K. Li (2002), Spiders: where are the bugs? *Journal of Business*, 75 (3), 453–473.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.

Engle, R. F., and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251-276.

Engle, R. F., and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric theory*, *11*(1), 122-150.

Engle, R. F., and Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The journal of finance, 48*(5), 1749-1778.

Engle, R. F., and Patton, A. J. (2001). What good is a volatility model. *Quantitative finance*, *1*(2), 237-245.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339-350.

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

Figlewski, S. (1984). Hedging performance and basis risk in stock index futures. *The Journal of Finance*, *39*(3), 657-669.

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

Fuller, W.A. (1976). Introduction to Statistical Time Series. New York: John Wiley.

Gallagher, D.R, and R. Segara, (2006). The performance and trading characteristics of exchange-traded funds. *Journal of Investment Strategy*, 1 (1), 47-58.

Gastineau, G. L. (2001). Exchange-Traded Funds. *Handbook of finance*.

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

Gastineau, G. L. (2010). The exchange-traded funds manual (Vol. 186). *John Wiley and Sons*.

Granger, C. W. (1988). Causality, cointegration, and control. *Journal of Economic Dynamics and Control*, *12*(2-3), 551-559.

Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *The Econometrics Journal*, 3(2), 148-161.

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

Hasbrouck, J. (2003). Intraday price formation in US equity index markets. *The Journal of Finance*, *58*(6), 2375-2400.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1251-1271.

Grier, K. B., Henry, Ó. T., Olekalns, N., and Shields, K. (2004). The asymmetric effects of uncertainty on inflation and output growth. *Journal of Applied Econometrics*, *19*(5), 551-565.

Henry, Ó.T. and M.C. Mari-Clérigues (2017a). Tracking or return matching: Evidence from the NYSE. *Working paper, The University of Liverpool.*

Henry, Ó.T. and M.C. Mari-Clérigues (2017b). The determinants of Tracking Errors in Equity and Fixed Debt Exchange Traded Funds: Evidence from the NYSE. *Working paper, The University of Liverpool.*

Houweling, P. (2011). On the performance of fixed income exchange traded funds.

Im, K. S., Pesaran, M. H., and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of econometrics*, *115*(1), 53-74.

Jares, T.E. and A.M. Lavin (2004). Japan and Hong Kong exchange-traded funds (ETFs): Discounts, returns, and trading strategies. *Journal of Financial Services Research* 25:1 57-69.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, *12*(2-3), 231-254.

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

Jones, P. M., and Olson, E. (2013). The time-varying correlation between uncertainty, output, and inflation: Evidence from a DCC-GARCH model. *Economics Letters*, *118*(1), 33-37.

Kayali, M. M. (2007). Pricing efficiency of exchange traded funds in Turkey: early evidence from the Dow Jones Istanbul 20. *International Research Journal of Finance and Economics*, *10*, 14-23.

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

Kroner, K. F., and Sultan, J. (1993). Time-varying distributions and dynamic hedging with foreign currency futures. *Journal of Financial and Quantitative Analysis*, 28(4), 535-551.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. *Journal of econometrics*, *54*(1-3), 159-178.

Levy, A. and O. Lieberman (2013). Overreaction of country ETFs to US market returns: Intraday vs. daily horizons and the role of synchronized trading. *Journal of Banking and Finance* 37 (2013) 1412–1421.

Levin, A., C. Lin and C. Chu (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics*, 108, 1–24.

Li, Y., and Giles, D. E. (2015). Modelling volatility spillover effects between developed stock markets and asian emerging stock markets. *International Journal of Finance & Economics*, *20*(2), 155-177.

Maddala, G. S., and Wu, S. (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and statistics*, 61(S1), 631-652.

Miffre, J. (2007). Country-specific ETFs: An efficient approach to global asset allocation. *Journal of Asset Management*, 8(2), 112-122.

Milonas, N. T., and Rompotis, G. G. (2010). Dual offerings of ETFs on the same stock index: US vs. Swiss ETFs. *The Journal of Alternative Investments*, *12*(4), 97.

Pedroni P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics*, 61, 653-670.

Pedroni, P. (2001). Purchasing power parity tests in cointegrated panels. *Review of Economics and Statistics* 83, 727–731.

Pedroni, P. (2004). Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. *Econometric Theory*, 20(03), 597-625.

Persyn, D. and J. Westerlund. 2008. Error Correction Based cointegration Tests for Panel Data. *Stata Journal 8* (2), 232-241.

Petajisto, A. (2017). Inefficiencies in the pricing of exchange-traded funds. *Financial Analysts Journal*, *73*(1), 24-54.

Phillips, P. and P. Perron (1988). Testing for a Unit Root in Time Series Regressions. *Biometrika*, 75, 335-346.

Phillips, P. C., and Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. Econometrica*: Journal of the Econometric Society*, 165-193.

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

Poterba, J. M., and Shoven, J. B. (2002). Exchange traded funds: A new investment option for taxable investors. (No. w8781). *National bureau of economic research*.

Roll, R. (1992). A mean/variance analysis of tracking error. *The Journal of Portfolio Management, 18(4), 13-22*.

Rompotis, G. G. (2009). Interfamily competition on index tracking: The case of the vanguard ETFs and index funds. *Journal of Asset Management*, 10(4), 263-278.

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

Rudolf, M., Wolter, H. J., and Zimmermann, H. (1999). A linear model for tracking error minimization. *Journal of Banking and Finance*, 23(1), 85-103.

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

Wahab, M., and Lashgari, M. (1993). Price dynamics and error correction in stock index and stock index futures markets: A cointegration approach. *Journal of Futures Markets*, *13*(7), 711-742.

Westerlund, J. 2007. Testing for error correction in panel data. *Oxford Bulletin of Economics and Statistics* 69: 709–748.

Tse, Y., Bandyopadhyay, P., and Shen, Y. P. (2006). Intraday price discovery in the DJIA index markets. *Journal of Business Finance & Accounting*, *33*(9-10), 1572-1585.

www.ft.com

www.ici.org

www.nyse.com

www.sourceetf.com