GEE

GEE Papers

Number 109

September 2018

Do Exchange Traded Funds (ETFs) Outperform the Market? Evidence from the Portuguese Stock Index

Carlos Manuel Pinheiro, Hugo Hilário Varela

Do Exchange Traded Funds (ETFs) Outperform the Market? Evidence from the Portuguese Stock Index

Carlos Manuel Pinheiro[1], Hugo Hilário Varela[2]

Abstract

We investigate the performance of exchange-traded funds (ETFs) that track the Portuguese stock index PSI20 since 2012. ETFs have contributed to the competitiveness of the Portuguese financial market. To test our hypothesis of whether ETFs generate significant abnormal returns as compared to the market, we use daily closing prices from December 2012 to June 2017. Using the risk-return model and analyzing Jensen's alpha we do not find evidence that ETFs tracking the PSI20 outperform the market, unlike other ETFs tracking major European and U.S. stock indices. Our results should interest investors and the financial industry.

JEL Classification: G11, G23, G12

Keywords: *Exchange-Traded Funds; Performance; Abnormal returns; Capital Asset Pricing model (CAPM)*

Note: This article is sole responsibility of the authors and do not necessarily reflect the positions of GEE or the Portuguese Ministry of Economy.

[1] Caixa Geral de Depósitos S.A.; Lisbon Accounting and Business School (LABS-ISCAL); Instituto Universitário de Lisboa (ISCTE-IUL) Business Research Unit (BRU-IUL), Lisboa, Portugal; Universidade Europeia, Laureate International Universities, Portugal

[2] Caixa Económica Montepio Geral, Portuguese Bank

* Address for correspondence: Hugo Hilário Varela, Rua Peixinho Júnior, 3 - Paço de Arcos, Lisbon, Portugal. Phone: +(351)933853664. E-mail addresses: hugo.hilario.varela@montepio.pt (Hugo H. Varela); carlos.manuel.pinheiro@cgd.pt (C. M. Pinheiro).

1. Introduction

Over the last 20 years, since their inception in the 1990s, Exchange Traded Funds (henceforth 'ETFs') have been drawing investors' appetite for highly profitable, less expensive tradable financial products combining the features of stocks and mutual funds. Mutual funds are commonly pictured as a "train", composed by many locomotives (investors), which constitute the dynamic force of the "train".

The train is constantly backed up by the locomotives to ensure it runs its normal course – stop at every defined destination "dropping off and picking up passengers" (purchase of securities), who are later transported to the "land of positive returns". This succinct allegory might explain the success of the ETF industry in the last few years with assets under management (AUM) exceeding $3tn, whilst 20 years ago AUM was negligible (see, Authers & Newlands, 2016).

The number of ETF products traded worldwide, including equity, commodities, and fixed income ETFs, reached massive proportions. In 2000, there were approximately less than 300 ETF products available worldwide, while in 2016, the number of ETF products was surprisingly close to 5,000.

ETFs are a relatively recent phenomenon. It all started in 1993, with the successful launch of SPDR (Standard & Poor's Depository Receipt), a unit investment trust holding portfolio poised to match and track the performance of S&P 500 (Standard and Poor's index of the 500 largest trading companies). SPDR's incredible success hails from its special and revolutionary trading characteristics – investors can trade SPDR throughout the day, unlike more traditional mutual funds that can only be traded at the end of the day when their net asset value is calculated. Another important feature of ETFs is their ability to be created or redeemed to meet and satisfy large-scale investors' needs. In a summarized way, the creation and redemption mechanisms of ETFs reflect how ETFs gain exposure to the market, which confer ETFs with special agility and efficient responsiveness mechanisms to deal properly with sudden changes in the market. The creation and redemption mechanism allows ETFs to be more flexible and capable of providing additional liquidity and less expensive transactions to investors.

ETFs promise higher returns at a lower cost as compared to holding the corresponding portfolio of securities or even other mutual funds, which motivates increasingly more investors to resort to this type of investment vehicle. However, there is a difference between traditional mutual funds and ETFs. Unlike some mutual funds (closed-end funds), ETFs can be sold as a common stock on a stock exchange market.

There are also other advantages regarding liquidity, other fees, expenses, and exposure to the daily fluctuation of stock prices and day-end prices adjustment (Net Asset Value). It is recognized that ETFs triggered systematic shifting in the investment vehicles used by investors, from traditional actively managed open-ended mutual funds, towards passive managed ETFs.

ETFs on the Portuguese PSI20 index were first launched in September 2010, by NYSE Euronext in cooperation with Commerzbank, conferring more investment alternatives and diversified financial instruments to the Portuguese stock market. The PPP ETF which tracks the performance of the PSI 20 index brought access to a novel class asset class on the Portuguese stock market, which combined both the simplicity of a daily stock trading mechanism and the virtues of highly diversified investments funds.

For the Portuguese stock market, integrated into the Euronext worldwide platform, this was an additional and innovative step towards competing on equal footing with other jurisdictions and provided more depth to the market.

Undoubtedly, PSI20 benefited from the experience from early entrants in the ETF market tapping the benefits of a tested, credible instrument and an appealing one for investors due to the above-expended motives: easiness of access, simplicity, wide diversification and low trading expenses.

ETFs are remarkably less expensive than traditional mutual funds and they feature a very efficient way to track/replicate a stock, a commodity, or even a bond index regardless where it is being traded. Seven years have passed since ETFs' inception in the Portuguese stock index, providing market data for a topical and timely analysis. Investors should be interested in a market-grounded analysis on whether ETFs are effectively earning substantially more (abnormal) return, "in excess" of the market return, given their level of risk. It can be argued that traditional finance tenets, namely portfolio theory, already answered the question whether investors are better off with a diversified, low-cost financial instrument or with a more traditional one, costlier and not so diversified. The issue, however, is that it seems that the excess or abnormal returns have not always been there. So, this is a question still requiring an empirical analysis. We aim at filling this gap in the literature, delving into the market data.

Our main purpose is to analyze the performance of Portuguese ETF securities through the lens of the expected return theory – the Capital Asset Pricing Model (henceforth denoted 'CAPM'). In other words, empirical tests will ascertain whether the return of ETFs can be even higher than the risk-adjusted return provided by CAPM (see Sharpe, 1964), in which the inherent risk is measured by the beta (β) coefficient or the sensitivity to market return variation.

Our work extends the analysis to another level: in case we find empirical evidence that ETFs outperform the market, then this validates the theoretical underpinnings of the CAPM model – riskier assets should have higher return than less risky assets – and it would suggest CAPM has a consistent application to the real stock world (efficient stock-pricing mechanism). If not, we have an opportunity to unveil the underpinnings (see Lintner 1965, and Sharpe 1964).

To provide an overview of the framework that we will use for PSI20, in the next sections we present a similar exercise for the U.S. mutual fund market, based on Jensen (1968)'s model. We posit that the U.S. mutual fund market has some resemblance with the dynamics of the Portuguese investment funds market. The remainder of this article is organized as follows. The next section addresses the two central factors of traditional Finance, risk and return. Section 3 presents the recent studies more related to our research. Section 4 caters to the research design. Results are presented in Section 5 and the following Section concludes.

2. Literature Review

2.1 Motives for ETFs

The systematic growth of investment funds and tradable "pools of money", that allow investors to trade individual shares in the exchange market, without incurring in high costs or expensive managerial fees imposed by actively managed funds, paved the way for more sophisticated and highly developed passive managed funds.

During the past decade, as investors started to realize the advantages of ETFs and their beneficial tradability features, the appetite for ETFs peaked, and both the volume of assets under management (AUM) and the number of ETF players reached record-breaking figures. Thereby, it was only a matter of time until ETFs took over the markets, increasingly conquering new investors, disseminating new products, and providing a diversified and well-structured range of products to investors, thus fulfilling on a consistent basis their appetite for highly profitable, diversifiable and global range financial products.

Twenty years ago, the assets under management of the main sponsors were almost negligible, whereas nowadays they exceed $3tn. In the last year, U.S. mutual funds experienced a massive outflow of $130.7bn according to the annual report provided by research firm Morningstar (see Authers & Newlands, 2016). There is no doubt that ETFs triggered systematic shifting in the investment vehicles used by investors, from traditional actively managed open-ended mutual funds, towards passive managed ETFs, as depicted in Figure 1.

<Figure 1. ETFs Take Over>

In turn, the number of ETF products traded worldwide, including equity, commodities, and fixed income ETFs, reached huge proportions. In 2000, there were approximately less than 300 ETF products available worldwide; while in 2016, the number of ETF products was surprisingly close to 5,000 (see, Figure 2).

<Figure 2. Number of ETFs Products Available Worldwide>

Based on the information above, analyzing the performance of ETFs is undoubtedly a topical issue as it can contribute to the financial literature, providing a better understanding of the forces at play in terms of ETFs' return, volatility, efficiency, and liquidity. ETFs brought about dramatic changes and have had a significant impact on stock trading and capital markets worldwide.

2.1.1 Creation and Redemption of ETFs

Another important feature of ETFs is their ability to be created or redeemed to meet and satisfy large-scale investors' needs. Through the creation mechanism, an Authorized Participant (AP) can acquire the securities that the ETF intends to hold. The acquired securities are delivered to the ETF provider, and in turn, the AP demands from the providers a block of equally valued ETF shares, called a creation unit. Thus, whenever an ETF provider wants to disseminate and create shares available for traders or launch new products to meet investors demand, he can resource to a market maker with strong buying power (AP). The AP will acquire the required securities in the exchange, in most cases in blocks of 50,000 ETF shares, which can be resold whenever it becomes profitable to do so.

The AP can remove ETF shares from the market by purchasing a significant amount of ETF shares to form a creation unit, and then deliver these shares to the ETF provider, receiving, in turn, the same value converted into underlying securities of the fund. This is called the redemption mechanism. This feature is very important because it can force the readjustment process of ETF shares price, whenever there is an imbalance regarding the market value of an ETF and the value of its underlying securities. For example, if many investors want to buy an ETF, triggering a significant shift in demand toward ETF, the ETF shares prices might rise at higher speed and scale as compared to its underlying securities. Thus, in this situation, the AP is expected to intervene forcing the overpriced ETF shares prices to come down to 'their fair value, by purchasing the underlying shares that compose the ETF, selling afterward the shares in the open market pushing the price down.

When the market value of an ETF is above their NAV or the value of each share of its underlying securities we can say that the ETF is trading at a premium. In turn, when ETF shares are trading below the price of their NAV, the ETF is trading at a discount. Thus, when the first situation occurs, an AP must trigger the redemption process, removing ETF shares from the market, preventing market asymmetries that might arise in case investors decide to close immediately their long positions on ETF holdings, dumping them into the market.

2.2 Past studies

The research on previous studies follows two mains strands: (i) performance and tracking efficiency models, and (ii) conventional asset pricing models.

2.2.1 Performance and Tracking Efficiency

A large part of the debate over traditional index funds and ETFs concerns their ability to beat the market, performing relatively better than the inherent benchmark, without compromising their major goal (tracking the index as closely as possible at lower cost).

However, this ability and operational efficiency do not apply to all ETFs. Gastineau (2004) documents that ETFs linked to popular U.S. stocks indexes, such as Russel 2000 and S&P 500 exhibited a weak performance as compared to their competitors. Gastineau (2004)'s results show that iShares Russel ETF underperformed the inherent benchmark by 53 basis points (bps) or 0,53%, in 2001, while the Vanguard Small Cap Index Fund beat the Russel 2000 benchmark by 76bps (0.76%) in the same period. To explain the weak performance of ETFs as compared to competing products, like mutual funds, Gastineau (2004) focused on the management policy undertaken by ETFs' portfolio managers, which he believed to be the source of a structural problem. Whenever there is a change in the composition of the inherent index, the portfolio manager will have to decide when and how to adjust the fund formation in order to reflect the changes occurred in the index.

This process of matching the weights of the securities held by the ETF accordingly to the size of the index is supported by the creation and redemption mechanisms. As stated by Gastineau (2014), the transaction costs incurred each time a market manager or Authorized Participant (AP) decides to create new shares or extinguish existing ones, cannot be ignored. In fact, the price charged to investors tends to reflect the administrative costs related to handling the creation or redemption basket, as the ETF manager attempts to replicate the returns and the risk of the underlying market index, as closely as possible.

Elton et al. (2005) examined the performance of SPDR (Standard & Poor's Depository Receipt), commonly known as a spider, a famous and popular ETF that tracks the S&P 500 index. They found that, on average, the NAV (Net Asset Value) return on the spider underperforms the S&P 500 return, by 28bps (0.28%) per year. Elton et al. (2005) posit that the difference in performance hails from the cost structure and the tracking efficiency of spiders.

The management fee, as well as other management expenses, constitute a clear disadvantage for ETF investors. The return of a spider is reduced in a significant way by the transaction costs incurred in replicating the index, adjusting the weights of the holdings every time there is a change in the composition of the index. Once again, the intrinsic structure of ETFs, involving the creation and redemption of shares seems to affect their ability to outperform the underlying benchmark.

The replication strategy should be such that, at any point in time the number of stocks held by the ETF and the stocks in the S&P 500 feature an exact match in terms of weights. This situation does not always occur, and when it does not, the efficiency of the ETF in replicating the index is clearly undermined. In the meantime, Elton et al. (2005) found no significant evidence that connects the tracking error to the inferior performance of the spiders relative to the benchmark.

Frino & Gallagher (2001) report the struggle faced by index-oriented fund managers on a daily basis, as they attempt to replicate the underlying index without incurring in substantial transaction costs. They highlight the fact that tracking error in performance or inaccurate replication strategy are unavoidable due to market frictions. Theoretically, the task of index managers to realign the weights of securities in the portfolio, trading index securities whenever there is a change in the composition of the underlying index, does not have to be a “difficult job”, as the market itself converges to a state of equilibrium .

Unfortunately, the real-world market conditions prevent, in many ways, transactions to take place smoothly. First, there is a cost involved, every time the index manager decides to buy or sell new shares. Second, transactions or trading schemes involving a considerable number of shares of stocks do not occur instantaneously, but rather depend on the investors’ willingness to perform quickly such trade and on the size of the stocks entering and exiting the index.

Therefore, market timing to incorporate a certain transaction is a really important issue for index managers. According to them, the magnitude and the variation of the tracking error arise directly from the volatility of the securities that constitute an index. Dividend distribution policies and the timing to effectively pay to investors can also cause tracking error.

Examining the 42 S&P 500 index mutual funds, Frino & Gallagher (2001) document the presence of tracking error in performance. Their results show that the cross-sectional average absolute tracking error[3] is equal to 5.9bp (0.059%) per month before expenses. The tracking error estimated as the volatility of the difference in returns is approximately 8.0bp (0.08%) per month or 27.6bp (0.276%) per annum. The correlation between dividend payments to the holders of S&P 500 securities and the tracking error is positive and statistically significant. There are signs of seasonality in the tracking error. The estimated tracking error is significantly higher in the months of January and May. By comparing active funds with passive S&P 500 index mutual funds, on average the latter outperform active funds, exhibiting a risk-adjusted excess return after expenses higher than that of large-cap active mutual funds.

Svetina & Wahal (2008) used a sample of 584 ETFs to demonstrate that on average ETFs underperform their underlying benchmark, due to transaction costs. Domestic equity ETFs representing approximately 60% of the sample underperform their benchmark by 0.26 percent per year, while international equity ETFs underperform by 0.19 percent per year. Contrary to Elton et al. (2005)’s study, Svetina & Wahal (2008) found that retail index-oriented mutual funds are outperformed by ETFs at about 0.30 percent per year, on average. Concerning the accuracy in tracking the underlying index, the results are consistent with the literature For the domestic equity ETFs and international equity ETFs, they estimate a tracking error of 0.47% and 1.13%, respectively.

[3] The absolute tracking error accounts for the difference in returns between the index portfolio and the underlying benchmark. Another common metric of tracking error also used in this paper is the standard deviation of the difference in returns between the index portfolio and the underlying index.

Shin & Soydemir (2010) found similar results, using three different methods to calculate the tracking error of 26 ETFs. They found evidence of persistence in tracking errors, which explains why the examined ETFs underperform their benchmark. This study highlights the inability of the passive investment strategy to beat the market.

Kostovetsky (2003) provides a consistent comparative analysis between ETFs and conventional index funds, by developing a simple one-period model which proved to be very useful in examining substantial differences between the two investment vehicles.

The competitive cost advantage of ETFs over indices relates to their ability to disseminate and eliminate shares in the market through a subtle and highly effective mechanism (creation and redemption), which allows large investors to purchase considerable amounts of ETF shares without any exposure to direct liquidity costs in the secondary market, offering only, in turn, a portfolio of securities matching the target index in weights and value.

Apart from the fund transaction costs and exploitation of arbitrage opportunities, which are quite irrelevant for ETFs due to the above mentioned motives, Kostovetsky (2003) points out three other relevant costs that can probably influence investors' choice between ETF or index funds. Rebalancing costs due to changes in the index or corporate activity (i.e. a merger and acquisition operation, forcing the merged or acquired company to leave the index) are relevant and affect both, ETFs and index funds, in the same proportion. Shareholder transaction costs are a major issue for ETF investors who must pay a commission to the broker in order to purchase shares in the secondary market (with exception of large investors). Contrarily, most index funds charge no commissions on transactions (i.e. no-load index funds). Finally, the tax efficiency is the third factor that affects ETFs and index funds in different ways. While tax efficiency associated with capital gains favors ETFs, index fund investors are negatively affected by the tax burdens. The in-kind redemption process of ETFs allows for the washing out of the most appreciated stocks, thus capital gains distribution is a rare operation for most ETFs.

Poterba & Shoven (2002) compare the pre-tax return and after-tax return on the SPDR trust (the largest ETF that tracks the S&P500), with the returns on the largest equity index fund, the Vanguard Index 500 fund. The results suggest a superior pre-tax and after-tax performance of the Vanguard Index 500 as compared to the SPDR trust. The superior performance of the index fund might be related with profitable trading strategies adopted by fund managers (i.e. purchasing shares of companies that intend to make part of the S&P 500 at the announcement date).

With the same goal of comparing the performance and the tracking efficiency of an ETF with an index mutual fund, Bello (2012) investigated small caps ETFs and index mutual funds that track the same benchmark (Russel 2000 index). His results suggest that small-cap ETFs are less diversified than index mutual funds, because ETFs generally overinvest in the top-ten companies they hold, thus only a small percentage of their investments is allocated into "less-valuable" securities. As expected, ETFs have lower expense ratios and high portfolio turnover than index mutual funds.

To evaluate ETFs and index mutual funds' relative performance, Bello (2012) uses Jensen's alpha and Sharpe information ratio. Consistent with previous research, he found that, despite the superior performance of index mutual funds relative to small caps ETFs, both underperformed their benchmark. Concerning the tracking efficiency, Bello (2012) found that ETFs exhibit a larger tracking error than index mutual funds.

Despite the similarities and discrepancies between these two instruments (conventional index mutual funds and ETFs), their coexistence is inevitable. They both provide to investors a diversified portfolio and alternative investment opportunities, competing against each other in a very sustainable way, exploiting systematically effective means to reduce managerial and transaction costs. Agapova (2011)'s study shows that conventional funds and ETFs are substitutes, but not perfect substitutes for one another. Which means that despite the innovative marketability and tradability features brought about by ETFs, conventional funds are still not easily replaceable. The competition and coexistence between these two products are sustainable and useful for investors, who benefit from increasing competition pushing prices down.

Cremers, Ferreira, Matos, & Starks (2016) hypothesize that the competitive pressure brought about by index funds not only forces the active funds to be more cost attractive but also to be more risk efficient, generating positive alpha. They found that on average active funds charge lower (higher) fees in markets with more explicit indexing funds (closet indexing[4]). Therefore, the entry of new index funds conceived explicitly to track the performance of a benchmark index, does benefit investors, as the fees charged by active funds tend to decrease. Consistent with the underlying hypothesis of the paper, Cremers, Ferreira, Matos, & Starks (2016) found that the average alpha generated by skilled active funds' managers is higher in countries in which low-cost passive managed funds are more popular, and lower in markets dominated by closet indexing.

Pricing deviations and situations of imbalance between ETFs' price and the market value of their underlying securities are other topical issue. Thirumalai (2003) finds evidence of price deviation from NAV, by analyzing the pricing efficiency of passive and active ETFs. DeFusco, Ivanov, & Karels (2011) focused primarily on the mismatch between the price of ETFs and the price of the market index being tracked. They used the Engle & Sarkar (2006)'s approach to calculate the price deviation as the difference between the price of the market index and the price of ETF. They found that the price deviation of the three most liquid ETFs (Spider, Diamonds, and Cubes) are predictable and stationary. The results suggest that, on average, the spider ETF is priced at 29 cents above S&P 500's price, Diamond's price is 8 cents higher than the price of Dow Jones Industrial Average (DJIA), and the Cubes ETF is priced at 25 cents below its underlying benchmark. Once again dividend accumulation and distribution policy, as well as the intrinsic nature of ETFs and their price formation mechanisms, are proclaimed to constitute the main source of the problem, causing pricing deviation and inefficient managerial performance.

Ivanov (2013) extends the previous work of DeFusco, Ivanov & Karels (2011), using high-frequency data. Ivanov (2013) found empirical evidence of high-frequency pricing deviations between ETFs and their underlying market indices. Price deviations of DIA, SPY, and QQQ amount to 0.0429, -0.0743 and 0.4298 respectively. These findings cast doubts on the efficiency of highly sophisticated monitoring computers and algorithmic-based trading programs that are responsible for clearing any signs of mispricing of ETFs. Pricing inefficiency of ETFs can benefit investors, who can make a riskless profit using an arbitrage strategy (i.e. take a short (long) on the overpriced (underpriced) ETF and buy (sell) the cheapest (expensive) basket of securities). In the meantime, arbitrage opportunities and situations of discrepancies in prices are quickly overcome, as the market converges to equilibrium.

[4] Closet indexing is a terminology used to describe funds that falsely claim to actively purchase investments but use passive strategy instead. They normally charge high fees, as if they are active, and yield average returns similar to a benchmark index.

Hilliard (2014) found that the median long-term premium of U.S. equity ETFs tends to converge to zero, which confirms the high efficiency of ETFs' arbitrage pricing mechanism in the long run.

Thus, for some ETFs the creation and redemption process are being executed in an efficient way, meaning that profitable arbitrage opportunities (when the NAV of the ETF differs from shares prices) are very limited due to the effectiveness of the creation and redemption strategy. International ETFs face a more complex process. In some countries, there are barriers to transferring and transacting shares overseas, making the arbitrage mechanism for international ETFs costlier and riskier than for domestic ones.

Engle & Sarkar (2006) highlight the problem faced by international ETFs in preventing their shares to be traded above or below NAV. Their findings indicate persistent and frequent premiums or discounts related to international ETFs.

Research on the performance of ETFs linked to European, Asian, Australian or other emerging markets is scarce. The absence of empirical research and the limited evidence on the performance of European and Asian ETFs are a clear limiting factor for our study. However, we should recognize that even today, innovative and dynamic products like ETFs are still new concepts for most investors.

Gallagher & Segara (2006) investigate the performance and trading characteristics of classical index-oriented ETFs in Australia. They provide a consistent analysis regarding the ability of ETFs to replicate persistently the Australian Stock Exchange, by comparing their tracking error volatility with other index funds operating in the market. Gallagher & Segara (2006) use two classical methods of calculating the tracking error, widely addressed in the literature (absolute difference in returns and standard deviation of the difference in returns between the ETF and the underlying index), and they found empirical evidence of significant tracking error across all ETFs and index-oriented funds tracking the Australian Stock Exchange index. In contrast with the previous studies, they found no evidence of superior or inferior performance of ETFs relative to their inherent benchmark.

Further results show that the wholesale index funds examined exhibit higher tracking error than ETFs, and the variation between the NAV and the quoted ETFs' prices are very small and do not occur frequently. Blitz & Huij (2012) investigate the overall performance of ETFs designed to track global emerging markets (henceforth denoted GEM) equity indexes (i.e. China, Brazil, India, South Africa and Russia). They argued that the tracking errors of those ETFs are much more likely to be higher than the tracking errors of the ETFs that mimic developing markets. The tendency of relative high tracking errors for GEM ETFs was explained based on the following arguments. First, the dispersion in stocks returns is larger for emerging markets; and second, the trading costs arising from low levels of stocks' liquidity is much higher in those emerging markets than in developed economies. The results confirm their hypothesis, as they found that GEM ETFs exhibit substantially higher levels of tracking errors than their developed markets counterparts. Consistent with the existing literature mainly focused in the investigation of ETFs linked to U.S. equity indexes, Blitz & Huij (2012) found that the GEM ETFs fall short of their benchmark indexes by around 85bp (0.85%), due to expected drag on returns caused by expenses and dividends taxation.

Regarding ETFs that provide passive exposure to European indices, the lack of empirical evidence is even more evident. Performance and pricing efficiency of European ETFs are poorly mentioned in the literature. Apart from Blitz, Huij & Swinkels (2012)'s research focused on the performance of European ETFs and index funds, not much has been said. Blitz, Huij & Swinkels (2012) posit that dividend withholding taxes as the main determinant of European ETFs' and index funds' inferior performance relative to their benchmarks. They concluded that dividend taxes are at least as important as fund expenses ratios in explaining European ETFs and index funds relative weak performance. They found that, on average, expense ratios drag down passive European funds' performance by 56bp (0.56%), while dividend taxes contribute to funds' performance at -48bp (-0.48%) per year.

Milonas & Rompotis (2006)'s study provides empirical evidence of the performance and marketability of 36 ETFs tracking the Swiss Stock Exchange index, an important player in the European ETF market. They applied a single index model to analyze Swiss listed ETFs ability to beat their inherent benchmark. The estimated coefficients are in line with previous studies - no evidence of superior performance was found, the mean alpha coefficient of the entire ETF sample was negative and statistically significant. The mean beta coefficient was significant and below one, meaning that the replication strategy adopted by those ETFs failed to fully mimic the return and volatility of their corresponding indexes, which can partially explain why they on average underperformed their benchmark by 7 basis points. By regressing ETFs average daily return on management fee ratio, using cross-sectional regression analysis, Milonas & Rompotis (2006) found that the performance of ETFs is significantly and negatively affected by management fees. The impact of management fees on ETFs tracking errors was found to be positive and statistically significant.

Yiannaki (2015) investigates the performance and tracking accuracy of ETFs domiciled in Ireland and Luxembourg. Consistent with previous studies, he found no evidence of abnormal returns or superior performance. The estimated Jensen's alpha parameters were insignificant and close to zero, as well as the tracking errors. They concluded that, despite their successful start, Ireland and Luxembourg listed ETFs underperform their benchmark.

Before we address the next topic of our literature review we would like to close this discussion over ETFs performance and tracking efficiency, summarizing the findings and their implications in our research. As spotlighted in literature, the ETFs' ability to replicate perfectly the performance of their inherent benchmark, matching its return and volatility, is not always accomplished; thus in line with the existing literature, we can say that tracking error is unavoidable.

The effectiveness of the creation and redemption process confers ETFs a trivial competitive advantage over conventional index funds, allowing them to be less expensive and more tax efficient.

These mechanisms also prevent ETFs shares to be traded above or below their underlying NAV, thus profitable arbitrage opportunities are very limited. However, findings concerning ETFs' pricing efficiency are ambiguous. There is, as of yet, no consensus.

2.2.2 The Risk-Return Model

Even before starting to think about investment pay off, investors should be aware of their money allocation. So, they start off with a simple question – "How much should I own in stocks (risky assets) and how much should I invest in a risk-free asset or other types of safe money securities?" By answering this question, the investor automatically derives a simple and very useful finding in portfolio theory, the investment opportunity set or all the efficient investments choices available to an investor at a given period.

Sharpe (1964) shows that the investment plan chosen by an investor is efficient, if and only if, it can provide the highest expected return for a given level of risk and the lowest risk for the same level of expected return. To maximize their utility as a function of risk and expected return, investors can choose among all possible portfolios that lie on the investment opportunity curve.

This simple illustration of how investors can allocate their capital is a clear demonstration of a simple asset allocation choice, which can help us understand the dynamic of Index models and guide us through the derivation process of the Capital Market Line, according to the CAPM (see Lintner, 1965). The equilibrium conditions of CAPM are restricted to a set of assumptions concerning investors' preferences and external market factors. CAPM assumes that all investors are rational, as they systematically apply the mean-variance theory developed in Markowitz (1952)'s seminal research[5], for optimization and utility maximization purposes.

Moreover, investors are considered to have identical investment horizons and share the same economic insight or homogeneous view of the market. So, they perform securities scanning using the same expected returns and covariance matrix of securities returns to create the optimal risky portfolio on the efficient frontier.

Investment decisions and portfolio construction strategies are defined according to the mean-variance theory.

Therefore, the risk-averse investor will rationally invest in the efficient portfolio in order to obtain the lowest volatility for any given level of expected return. Income taxes and transaction costs are assumed to be zero (frictionless market) and there is no restriction on credit - short selling is allowed.

Finally, single and isolated investors' trade decisions do not affect the market's overall equilibrium and security prices. Thus, investors are assumed to be price-takers and the market itself embeds a set of conditions underpinning what is denoted as perfect competition.

With no restrictions on borrowing or lending, investors can increase their investment opportunities by investing some money in Treasury bills (borrowing or lending at the risk-free rate) and the remaining money in stocks.

Black (1972) shows that the efficiency of the CAPM can be substantially improved under circumstances where investors can neither have access to the riskless asset nor borrowing or lending at the risk-free rate.

The model obtained by dropping one of the most important assumptions of the CAPM is consistent with the equilibrium models developed by Black, Jensen, & Scholes (1972) and contradicts Sharpe (1964) and Lintner (1965). The former contended that any point along the capital market line is desirable and attainable by investors if they behave rationally, adopting a diversified portfolio strategy.

The expected return rate increases proportionally to the level of risk incurred by investors – the price of risk -, yet investors can adjust their preferences combining the optimal risky portfolio with the riskless asset – the price of time -, thus avoiding the diversifiable part of the total risk. The power that this extensive diversification strategy has to eliminate part of the nonsystematic risk is empirically demonstrated by Lintner (1965) and later by Statman (1987). The former quantified the effects of portfolio diversification,

[5] Harry Markowitz is the mind behind one of the most important pillars of modern portfolio management, the Portfolio Selection Model. He provides consistent analyses towards optimal combinations of risky assets and risk-free assets, identifying the efficient set of portfolios or the efficient frontier of risky assets that minimizes the variance for any target expected return.

using data on NYSE (New York Stock Exchange) stocks. He found that, on average, the risk of equal weighted portfolios of randomly selected stocks decreases significantly with diversification[6]. Lintner (1965)'s seminal paper left some relevant recommendations to individual investors and mutual funds managers concerning expected returns and risk. Lintner testified that the best diversification strategy is the one that provides the highest ratio of the expected excess return to standard deviation. Thus, the optimal portfolio resulting from the best combinations of expected return and risk will provide the highest compensation or additional return per unit of extra risk incurred. The obtained results revealed that 57 of the 70 funds analyzed had a higher ratio of expected excess return to risk than the S&P 500, even though the majority of funds average lower returns.

The benefits of diversification were verified through the analysis of the residual risks, measured as the standard deviation of the estimate, obtained by regressing funds returns on the return of the market index (S&P 500 index). The results suggest that, apart from the market risk, the residual risk still accounts for a significant part of the total risk, even in situations of well-diversified and professionally managed portfolios[7]. Individual investors are highly recommended to pursue a prudent and effective diversification strategy to reduce the risk associated with the variability of the expected returns (which are significant, especially in case of common stocks), and to improve the expected return-risk relationship. They should also notice that the lower the correlation between the assets, the greater the gain in efficiency.

Consistent with the classical CAPM tenets, Sharpe (1964) demonstrated that the required rate of return of an asset converges to an equilibrium point that is linearly represented as a function of the asset's return rate responsiveness to changes in economic activity (systematic risk or market beta), plus the pure interest rate. Hence assets that are not affected by changes in the economic activity will provide the return of the risk-free asset and those which are sensitive to changes in the economic activity yield on average higher expected return rate. These findings lead to an important premise of the traditional capital market theory by showing that the aggressive securities or those which are more sensitive to market shocks are commonly priced above those whose response is less significant to changes in the economy (defensive securities). As a result, investors demand a higher expected return for investing in aggressive securities than they would require for investing in defensive securities. Which is to say that under the CAPM equilibrium conditions, investors can obtain the highest expected return regardless the level of risk, combining the least risky investment (i.e. U.S. Treasury bills), which has a beta of zero, and the riskier investment, the market portfolio of common stocks (beta is equal to one).

As investors are risk-averse, their willingness to take some risk is compensated with the excess return they require to invest in the market, or the market risk premium, which is given by the expected market return minus the risk-free rate. The linear relationship between the expected return on investment and the market risk (beta) is commonly referred to as the Security Market Line, which is a straight line that starts at the point where beta is zero (investment in a risk-free asset). Then, the expected return on risky investments increases proportionally with the market risk – the risk that investors cannot avoid or mitigate through diversification.

[6] The estimated average standard deviation of returns of portfolios comprised of only one stock was about 49,2%, and it declines sharply to the minimum of 19.2%, as the number of stocks included in the portfolio increased.

[7] The residual risk was higher than the return of the risk-free asset.

In short, CAPM predicts that the rational investors aware of their aversion to risk, will hold the market portfolio or a basket of securities that mimic the market index, in order to obtain the highest level of diversification possible. Therefore the rate of return required by investors is the contribution of each security to the overall risk of the market portfolio. Analytically the expected risk premium on stocks can be expressed as follows:

$$E(risk\ premium\ on\ stock) = beta * E(risk\ Premium\ on\ market) \quad (1)$$

Thus, investors require a premium for holding a risky asset. This premium (or compensation) demanded by investors varies in direct proportion to beta, which measures the extent to which returns on the stock and the market move together.

Black et al. (1972)'s seminal research empirically tests the implications of CAPM on security pricing. They found disruptive results concerning the relationship between the expected risk premiums on individual assets and their systematic risk or beta. The linear relationship between an assets' return and their betas theoretically established by the CAPM does not always hold. Black et al. (1972) found that the expected excess return of an asset is not strictly proportional to its beta. They built a two-factor model relaxing one of the most important assumptions of conventional CAPM, which states that investors have unlimited access to riskless borrowing and lending. In equilibrium, the two-factor model predicts that the expected excess return on a security is then given by the expected excess return on a portfolio that is uncorrelated with the return on the market portfolio (zero-beta portfolio), plus the market risk premium subtracted by the zero-beta portfolio, all multiplied by the security's level of systematic risk. The estimated parameters were efficient and statistically significant, thus contradicting the traditional form of the CAPM. Further results indicated that the excess returns on high-beta securities are lower than the expected risk-adjusted return predicted by CAPM[8]. For a low-beta portfolio of stocks, the estimated excess return was higher than the risk-adjusted return.

Gibbons & Ferson (1985) hypothesize that conventional financial valuations models have failed to analyze the behavior of conditional expected returns over time and incorporate the changes in expectations into their models. The assumptions of constant risk premiums were relaxed and the refined model suggests that conventional financial valuations models can derive consistent and testable estimators of expected returns without observing the market portfolio or specify the state variables.

The immediate concern of our first model is to analyze the performance of ETFs that track the most popular European and American market indexes and determine whether or not they can beat the market. To analyze the ETFs' ability to beat the market or to generate returns that are higher than the risk-adjusted returns predicted by CAPM, we will take Jensen (1968)'s approach. Jensen (1968) was the first to investigate and systematically test the performance of mutual funds, methodically examining their ability to beat the market. The results show that the funds were generally unable to perform sufficiently better than the market or even beat the market". However, some funds exhibited stunning performance and surprisingly high returns given their level of risk. The best fund in terms of performance yields an alpha of 0.058, exceeding the author's expectations, and regarding their sensitiveness to market changes the estimated beta for the average fund was approximately 0.85, indicating that most funds were less risky than the market index.

---

[8] Under CAPM equilibrium, the expected excess return of an asset is given by the excess return on a market portfolio multiplied by the asset's expected return rate responsiveness to changes in the economy (systematic risk or beta): $E(\tilde{R}_j) = E(\tilde{R}_M) \times \beta_i$.

The empirical model underpinning Jensen (1968)'s study is easily derived and it can be constructed in a straightforward way. The parameters of the model can be estimated through OLS (Ordinary Least Squares) and its consistency and statistical significance can be tested using a simple "t-test". Algebraically, it can be defined as:

$$R_{it} - R_{f_t} = \alpha_i + \beta_i(R_{m_t} - R_{f_t}) + u_{i_t} \quad (2)$$

This equation is a clear proxy of CAPM, where the return on security i (R_{it}) discounted from the risk-free rate (R_{f_t}) or the expected excess return is given by the adjusted market risk premium ($R_{m_t} - R_{f_t}$) multiplied by its level of risk (β_i), plus the error term, u_{i_t}.

To determine whether funds can perform significantly better or worse than their benchmark we need to analyze the sign and the significance level of the alpha parameter. Intuitively, the parameter alpha famously known as Jensen's alpha, signals if the fund can earn significant abnormal returns, "in excess" of the market-required return, given the fund contribution to the overall risk. Securities that are able to provide an expected excess return rate higher (lower) than the risk-adjusted return given by CAPM exhibit a positive (negative) alpha.

The following expression clarifies the meaning of risk-adjusted return inherent to CAPM and it helps us understand why increasingly more investors are resourcing to financial instruments that provide "abnormal returns" or positive alpha.

$$R_{it} - R_{ft} = \alpha_i + \beta_i(R_{mt} - R_{ft}) + u_{it} \leftrightarrow \alpha_i = R_{it} - R_{ft} - [\beta_i(R_{mt} - R_{ft})] \quad (3)$$

$$\alpha_i = (R_{it} - R_{ft}) - [\beta_i(R_{mt} - R_{ft})] \quad (4)$$

If the alpha coefficient is positive, the expected excess return on fund i ($R_{it} - R_{ft}$) is automatically higher than the risk adjusted return $\beta_i(R_{mt} - R_{ft})$, which means that the reward to investors is even higher than the theoretical compensation required for bearing the risk. And it is a clear indication that the chosen fund outperforms the broad market index, and therefore investors would want to buy it, as suggested by Ferson & Lin (2014). Ferson & Lin (2014) contended that a more efficient measure of alpha based on investors preferences can provide reliable buy or sell indication. The results indicate that funds whose client-specific alpha is expected to be larger than the traditional alpha, tend to experience larger inflows. This paper contradicts the existing literature, as most of the studies on this field concluded that traditional alphas are not reliable proxies of the attractiveness of an investment. Therefore, instructions to sell or buy provided by the alpha parameter are neither precise nor conclusive.

Chen & J. Knez (1996) found that the client may be willing to short the asset even if it has a positive alpha. Contrarily, Glode (2011)'s paper suggests that clients may be willing to buy a fund even if it has a negative alpha.

3. Data, Methodology, and Econometric Setup

3.1 Data

We gather daily historical prices of the PPP ETF and the PSI 20 index from December 2012 to June 2017. And as a proxy for the risk-free rate we use the 10-year yield change rate of the German Government Bonds.

For this study, we use daily data – much like in previous studies (see Peltomäki (2017)) - as it is more adequate, seeing that the creation and redemption features of ETFs normally take place at the end of the day to meet the daily pricing pressure, making daily data particularly interesting for ETFs.

The data was obtained from the Bloomberg terminal and we set the Euro as the currency measure.

3.1.1 Summary Statistics

Table 1 reports the description of the variables, and **Table 2** presents the summary statistics.

<Table 1 Description of Variables>

<Table 2 Summary Statistics>

The return of the ETF that tracks PSI 20 exhibited some periods of high volatility, during 2012 and 2016. From 2016 onwards, it exhibits a steady pattern. Across the period of analysis, the average daily return of ETFs is close to 0.0122%, which is equivalent to an average of 4.561% ((1+0,012%) ^365-1) return per year. The corresponding benchmark exhibits an average daily return of 0.0008% or approximately 0.282% return per year. The return on the PPP ETF outperforms the return on the market index. It is also more volatile over time than the market index, see Figure 3, and Figure 4.

<Figure 3. ETF Daily Return>

<Figure 4. PSI20 Daily Return>

The average daily excess return on ETF is negative, but it exceeds the risk premium provided by the underlying market index. The excess return on the PPP ETF is more volatile than the corresponding benchmark index. In terms of correlation, as expected, the return on the PPP ETF is highly correlated with the return on PSI20. It exhibits a positive and statistically significant (at 1% level) correlation with the market return, as depicted in Table 3.

<Table 3. Correlation Matrix>

3.2 Methodology

3.2.1 Measuring the Abnormal Return (Jensen Alpha Approach)

To properly answer the question concerning the Portuguese ETF's ability to generate an abnormal return, we will adopt the Jensen's Alpha approach and inherent methodology, so greatly referenced in the financial academic-circle. The methodology and technical specifications behind Jensen (1968)'s seminal study are quite simple and straightforward.

The underlying theory and hypothesis of his paper were tested using a simple hypothesis test, commonly known as the t-test. To test whether the PPP ETF can beat the market, we employ the Single-Index Model discussed in the previous chapter, and therefore perform simple statistical inference to check the significance of the intercept parameter – alpha.

The underlying hypothesis is that PPP ETF can generate returns that exceed the risk-adjusted return predicted by an equilibrium model such as CAPM, thus the alpha parameter should be positive.

This means that the expected rate of return generated by the PPP ETF is even higher than the compensation initially required by investors for investing in that investment vehicle, given its level of risk.

Our first model will take the form of:

$$r_{i_t} - r_{f_t} = \alpha_i + \beta_i(r_{m_t} - r_{f_t}) + u_{i_t} \quad (5)$$

Where: $r_{i_t} - r_{f_t}$ – is the return of security i (ETFs that track major European and American market indexes), at time t, minus the return of the risk-free asset (German Bunds 10-year yield change rate) – the expected excess return on ETFs ; $\beta_i(r_{m_t} - r_{f_t})$ – is the fraction of return due to movements in the overall market, it reveals the investors' "appetite" for the market risk or the market risk premium $(r_{m_t} - r_{f_t})$, given security's responsiveness to changes in the overall market index (β_i). The coefficient alpha (α_i), is the intercept of our equation, it measures the nonmarket premium or the abnormal return; and u_{i_t} – is the error term or the residual of our model, which is given by the difference between the actual ETFs excess return and its fitted value or the estimated values of ETFs excess return by our model.

The critical point here is not just to run the regression and obtain the parameters, but to verify if our model is capable of consistently explaining all the variation in the dependent variable or not. The first thing we can do to ensure that the estimated parameters or coefficients are reliable and significant – meaning that they represent the true relationship between the "regressed" and the predictor variable – is to perform statistical inference, aimed at testing to which extent we can reject our hypothesis of statistical significance of the parameter alpha.

To state matters plainly, the framework or the main scope of our hypothesis testing about the individual parameter alpha obtained from the sample of our bivariate linear regression model can be intuitively expressed as:

- $H_0: \alpha_i = 0$ – Null Hypothesis;
- $H_1: \alpha_i \neq 0$ – Alternative Hypothesis.

The null hypothesis being tested is that the true value of parameter alpha should be zero, against the alternative hypothesis – the true value of alpha is different than zero. The main interest here is to reject the null hypothesis in favor of the alternative hypothesis. Therefore a positive and significant alpha would suggest that European and U.S. located ETFs being examined are able to outperform the underlying market index, generating an expected excess return that is above the risk-adjusted return or the initial risk premium or compensation required by ETFs' investors.

The statistic to test the hypothesis stated above, is called the t-statistic or t-ratio of the parameter alpha, and it can be easily calculated as:

$$t_{\hat{\alpha}_i} = \frac{(\hat{\alpha}_i - \alpha_i)}{se(\hat{\alpha}_i)} \sim t_{n-k-1}. \qquad (6)$$

3.2.2 Performance Analysis

To analyze the performance of ETFs we will use some well-known performance indicators, such as the reward-to-volatility ratio, and tracking error. The reward-to-volatility takes many forms. Sharpe (1966) devised what we now call the Sharpe ratio, which is a consistent way of examining the performance or the desirability of an investment strategy by adjusting for its risk, or it can be defined as the extra return that investors demand, to bear a portfolio of risky assets – incremental return per incremental risk.

The analytical expression of the Sharpe ratio is given by:

$$Sharpe\ ratio = \frac{E(r_i) - R_f}{\sigma_i}. \qquad (7)$$

Where the excess return on ETFs $(E(r_i) - R_f)$ is adjusted to the level of risk that the portfolio represents to investor (σ_i).

Another common measure of the reward-to-volatility ratio is the Sortino ratio developed in Sortino & Price (1994)'s seminal research. It is a clear modification of the Sharpe ratio, but it uses a different measure of risk, as it takes as the denominator the downside deviation, which accounts for the "bad side of risk" or the standard deviation of the negative excess returns of the portfolio.

This improved version of risk-adjusted returns can mitigate the effect of high returns outliers, which is, in fact, a clear limitation of the Shape ratio since it does not distinguish between upside and downside volatility.

Treynor ratio is another version of the Sharpe ratio which strongly disagrees with the risk measure undertaken by Sharpe. Unlike Sharpe and Sortino ratio, Treynor ratio places major weight on the systematic risk of the portfolio, instead of using both (systematic and idiosyncratic risks).

$$\text{Sortino ratio} = \frac{E(r_i)-R_f}{DSD_i} \quad (8)$$

$$\text{Treynor ratio} = \frac{E(r_i)-R_f}{\beta_i}. \quad (9)$$

The denominators DSD_p and β_p account for the downside deviation of the ETF and the market Beta, respectively.

Tracking error basically indicates how closely a portfolio follows the inherent benchmark, so it can be either calculated as the absolute difference in returns of the index portfolio and the benchmark index or as the standard deviation of the difference between the portfolio and index returns (see Roll, 1992; Pope & Yadav, 1994, and Larsen & Resnick 1998).

Therefore, the two measures of tracking error can be expressed as follows:

$$TE_{1,i} = \frac{\sum_{i=1}^{n}(r_{it}-r_{bt})}{n} \quad (10)$$

$$TE_{2,i} = \sqrt{\frac{\sum_{i=1}^{n}(r_{it}-r_{bt})^2}{n-1}} \quad (11)$$

where:

r_{it} – is the return on ETF i in period t;

r_{bt} – is the return on the benchmark index, and n is the number of observations.

3.3 Econometric Setup

We perform several tests to address issues concerning model specification and frequent problems with long time series data, such as heteroscedasticity, instability in distributions, and serial correlation.

The underlying null hypothesis of White's test for heteroscedasticity was strongly rejected; therefore, we have clear evidence that the variance of the residuals varies systematically with known predictors, thus generating heteroscedastic estimators, (see Figure 5).

< Figure 5. Plot of Squared Residuals and Linear Prediction>

In the presence of heteroscedasticity, OLS estimators will produce biased coefficient estimates and misleading inferences. To avoid that, we will employ robust standard errors estimates, which are modified by conception to account for the heteroskedasticity.

The Fisher-type unit root test based on augmented Dickey-Fuller test reveals that the series is stationary, thus the null hypothesis that the series contain unit roots is strongly rejected.

The returns are calculated as the difference between the most recent closing price and the last closing price all divided by the last closing price. For estimation purposes, we use the continuous return also denoted the log returns[9], due to the benefits of normalization and consistency over long time series.

To improve the quality of data and avoid noisy observations, we discard the first month of data and extreme values are replaced at 1% and 99% percentiles.

4. Results

Table 4 shows that the tracking error is relatively low. The major goal of the PPP ETF or the fund's target, which is to track consistently the Portuguese market index is fulfilled.

The Sharpe and the Sortino ratio for both PPP ETF and PSI 20 are negative, meaning that the additional increment in the average return earned in excess of the risk-free rate per unit of risk taken is negative.

The expected return that investors appropriate by investing on the PPP ETF, or on the Portuguese stock market itself (PSI 20), tend to underperform the return of the riskless asset (German Bund 10-Year Yield).

<Table 4. Performance Analysis>

Also, the Sharpe ratio, as well as the Sortino and Treynor ratio, are higher for the PPP ETF. This results suggest that the PPP ETF outperforms the PSI 20 index. For each unit of risk taken, an investor would be more compensated for investing on the PPP ETF than on the PSI 20 index.

The following Figure 6 shows that the average return on the PPP ETF in excess of the market return tends to be positive over a significant period of time.

<Figure 6. Excess ETF Return Relative to the Benchmark>

Jensen's Alpha Results

Panel 1 of Table 5 shows that the estimated beta coefficient using robust regressions is statistically significant (at 1% level) and close to one, which confirms the sensitiveness of the PPP ETF to shocks in the Portuguese stock market (Panel 1). Index-tracking ETFs are designed to move alongside with their inherent benchmark indexes, therefore the beta coefficient is expected to be close to one.

The alpha coefficient is positive but statistically not significant (the p-value is very high). Thus, we do not have statistical evidence to affirm that 0.00009 (Panel 1) represents consistently the abnormal return of ETF, even though its positive.

Additionally, we perform a mean comparison test to evaluate the extent to which the returns on ETFs differ from the return provided by the benchmark indexes. We found no significant evidence of superior or inferior performance, the existent differences in returns are not statistically significant (t-stat are very low), see Panel 2 of Table 5.

<Table 5. Testing Abnormal Return>

[9] The continuous returns are calculated as the natural logarithm of the most recent closing price divided by the last closing price. Analytically it can be expressed as: $\ln(\frac{P_{i,t}}{P_{i,t-1}})$.

The positive alpha can be interpreted as a trace of ETF's outstanding performance potential, meaning that there is a possibility for investors to earn abnormal returns by investing in ETFs. The R-square measure is very high, meaning that the model explains almost all the variability in the dependent variable (see Figure 7). The good dimension of our sample - an average of 1,220 observations - may also explain the high level of the goodness of fit.

<Figure 7. Testing abnormal return>

The high quality of the linear fit is also present in Milonas & Rompotis (2006)'s research, and it highlights the fact that ETFs are doing a good job replicating the performance of the underlying market, meeting their tracking goals successfully.

5. Conclusions

The results and the statistical significance of the parameters suggest no clear indication for the prevailing theory that the PPP ETF can earn substantial abnormal returns or returns higher than the risk-adjusted return for the Portuguese market. However, the performance of ETFs combined with their lower cost structure and their trade characteristics, provide some guidance on the beneficial effects of investing in PSI20 ETFs, alongside the upside of a broader diversification.

The estimation results point out to a positive Alpha, suggesting the ability of ETFs to generate higher returns as compared to the market-required return given their level of risk. Nevertheless, ETFs incorporate changes occurred in the market, and therefore "bad times" can have an even higher impact on ETFs than they have on common stocks or mutual funds.

An avenue for future research might consider running multifactor models, such as the Fama-French 3-factor model, as they are more encompassing and incorporate specific characteristics of stocks composing the investment portfolio. This approach may lead to a better fit, minimizing the error of the CAPM model that we extensively used in this paper.

Most importantly, our analysis suggests that ETFs can enhance the competitiveness of the Portuguese Stock market, by stimulating and promoting the dissemination of high diversifiable financial instruments in the Portuguese market. ETFs contribute to additional liquidity and a wider range of investment solutions for investors in the Portuguese market who seek financial products that combine performance, lower cost structure, easiness in trading and low exposure to idiosyncratic risk.

References

Agapova, A. (2011). Conventional mutual index funds versus exchange-traded funds. *Journal of Financial Markets 14*(2), 323–343. https://doi.org/10.1016/j.finmar.2010.10.005

Authers, J. & Newlands, C. (2016). Exchange traded funds: taking over the markets. *Financial Times*.

Bello, Z. (2012). The investment performance and tracking errors of small-CAP ETFs. *Global Journal of Finance and Banking Issues 6*(6), 12–20.

Black, F. (1972). Capital market equilibrium with restricted borrowing. *Journal of Business 45*(3), 444–455.

Black, F., Jensen, M. & Scholes, M. (1972). *The Capital Asset Pricing Model: Some Empirical Tests*. *Studies in the Theory of Capital Markets*. https://doi.org/10.2139/ssrn.908569

Blitz, D. & Huij, J. (2012). Evaluating the performance of global emerging markets equity exchange-traded funds. *Emerging Markets Review 13*(2), 149–158. https://doi.org/10.1016/j.ememar.2012.01.004

Blitz, D., Huij, J. & Swinkels, L. (2012). The performance of European index funds and Exchange-Traded Funds. *European Financial Management 18*(4), 649–662. https://doi.org/10.1111/j.1468-036X.2010.00550.x

Chen, Z. & J. Knez, P. (1996). Portfolio performance measurement: Theory and applications. *Review of Financial Studies 9*(2), 511–555.

Cremers, M., Ferreira, M. A., Matos, P. & Starks, L. (2016). Indexing and active fund management: International evidence. *Journal of Financial Economics 120*(3), 539–560. https://doi.org/10.1016/j.jfineco.2016.02.008

DeFusco, R. A., Ivanov, S. I. & Karels, G. V. (2011). The exchange-traded funds' pricing deviation: Analysis and forecasts. *Journal of Economics and Finance 35*(2), 181–197. https://doi.org/10.1007/s12197-009-9090-6

Elton, E. J., Gruber, M. J., Comer, G. & Li, K. (2005). Spiders: Where are the Bugs? In *Exchange Traded Funds: Structure, Regulation, and Application of a New Fund Class* (pp. 37–59). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-27637-8_4

Engle, R. & Sarkar, D. (2006). Premiums-discounts and exchange traded Funds. *Journal of Derivatives XXXIII*(2), 81–87. https://doi.org/10.1007/s13398-014-0173-7.2

Ferson, W. & Lin, J. (2014). Alpha and performance measurement: The effects of investor disagreement and heterogeneity. *Journal of Finance 69*(4), 1565–1596. https://doi.org/10.1111/jofi.12165

Frino, A. & Gallagher, D. R. (2001). Tracking S&P 500 index funds. *Journal of Portfolio Management 28*(1), 44–55. https://doi.org/10.3905/jpm.2001.319822

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

Gastineau, G. L. (2004). The benchmark index ETF performance problem. *Journal of Portfolio Management 30*(2), 96–103. https://doi.org/10.3905/jpm.2004.319935

Gibbons, M. R. & Ferson, W. (1985). Testing asset pricing models with changing expectations and an unobservable market portfolio. *Journal of Financial Economics 14*(2), 217–236. https://doi.org/10.1016/0304-405X(85)90015-7

Glode, V. (2011). Why mutual funds "underperform." *Journal of Financial Economics 99*(3), 546–559. https://doi.org/10.1016/j.jfineco.2010.10.008

Hilliard, J. (2014). Premiums and discounts in ETFs: An analysis of the arbitrage mechanism in domestic and international funds. *Global Finance Journal 25*(2), 90–107. https://doi.org/10.1016/j.gfj.2014.06.001

Ivanov, S. I. (2013). High-frequency analysis of exchange-traded funds' pricing deviation. *Managerial Finance 39*(5). https://doi.org/10.1108/03074351311313834

Jensen, M. C. (1968). The performance of mutual funds in the period 1945-1964. *Journal of Finance 23*(2), 389–416. https://doi.org/10.1111/j.1540-6261.1968.tb00815.x

Kostovetsky, L. (2003). Index mutual funds and Exchange-Traded Funds. *Journal of Portfolio Management*, *29*(4) 80–92. https://doi.org/https://doi.org/10.3905/jpm.2003.319897

Larsen, A. G. & Resnick, G. B. (1998). Empirical Insights on indexing: How capitalization, stratification, and weighting can affect tracking error. *Journal of Portfolio Management 25*(1), 51.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *Journal of Finance 20*(4), 338–342. https://doi.org/https://doi.org/10.1111/j.1540-6261.1965.tb02930.x

Markowitz, H. (1952). Portfolio selection. *Journal of Finance* (2), 55–75. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

Milonas, N. T. & Rompotis, G. G. (2006). Investigating European ETFs: The case of the Swiss exchange-traded funds. *The Annual Conference of HFAA*, 1–28.

Peltomäki, J. (2017). Beta as a determinant of investor activity in sector exchange-traded funds. *Quarterly Review of Economics and Finance 65*, 137–145. https://doi.org/10.1016/j.qref.2016.06.006

Pope, P. F. & Yadav, P. K. (1994). Discovering errors in tracking error. *Journal of Portfolio Management*, *20*(2), 27 LP-32. Retrieved from http://jpm.iijournals.com/content/20/2/27.abstract

Poterba, J. M. & Shoven, J. B. (2002). Exchange-Traded Funds : A New investment option for taxable investors. *American Economic Review* 92(2), 422-427.

Roll, R. (1992). A mean/variance analysis of tracking error. *Journal of Portfolio Management 18*(4), 13–22. https://doi.org/10.3905/jpm.1992.701922

Sharpe, W. F. (1964). Capital asset prices: A theory of equilibrium under conditions of risk. *Journal of Finance 19(3), 425-442*. https://doi.org/10.2307/2977928

Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business 39(S1), 119-138*. https://doi.org/10.1086/294846

Shin, S. & Soydemir, G. (2010). Exchange-traded funds, persistence in tracking errors and information dissemination. *Journal of Multinational Financial Management 20*(4–5), 214–234. https://doi.org/10.1016/j.mulfin.2010.07.005

Sortino, F. A. & Price, L. N. (1994). Performance measurement in a downside risk framework. *Journal of Investing 3*(3), 59 LP-64. Retrieved from http://joi.iijournals.com/content/3/3/59.abstract

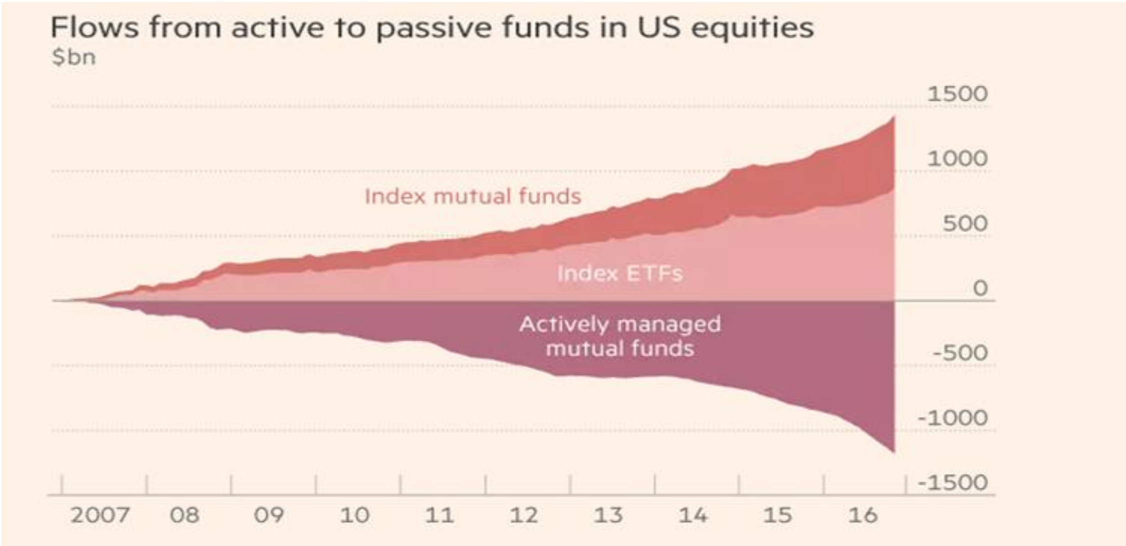
Statman, M. (1987). How many stocks make a diversified portfolio ? *Journal of Financial and Quantitative Analysis 22*(3), 353–363. Retrieved from http://www.jstor.org/stable/2330969

Svetina, M. (2010). Exchange Traded Funds: Performance and Competition. *Journal of Applied Finance 20*(November), 130–145. https://doi.org/10.2139/ssrn.1303643

Thirumalai, R. S. (2003). *Active vs. passive ETFs*. *Kelley School of Business Working Paper. Indiana University, Bloomington*.

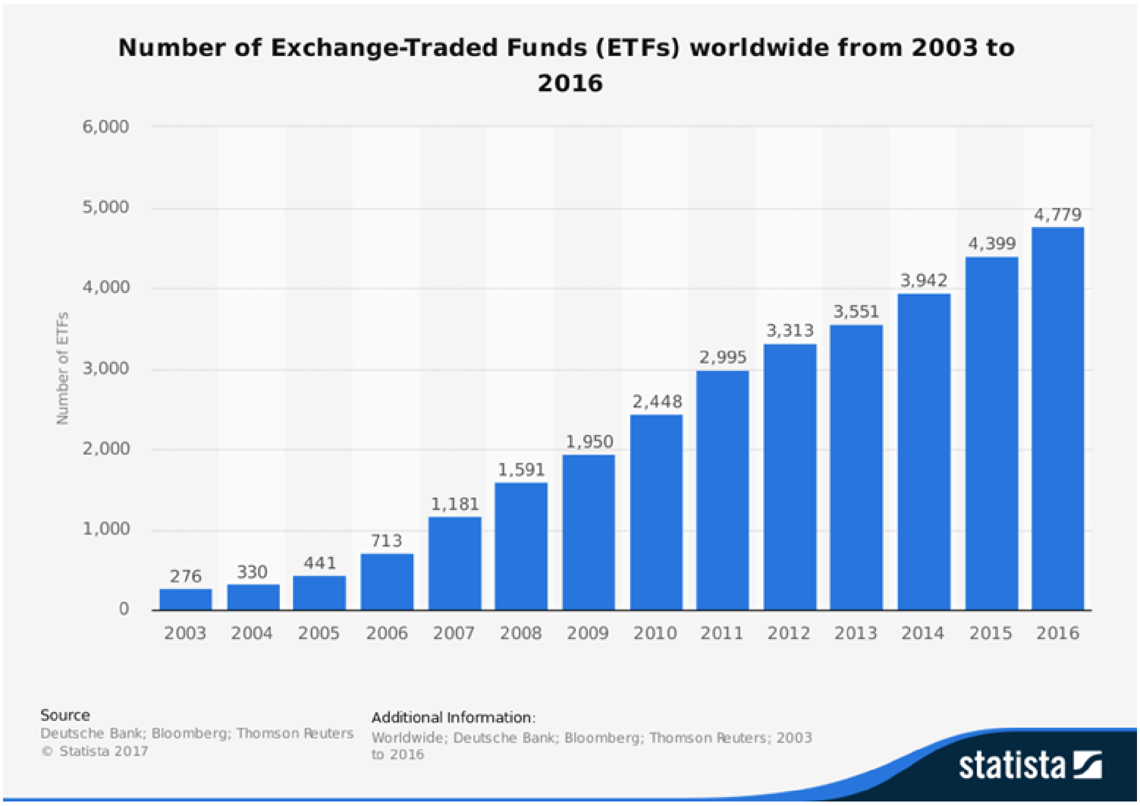
Yiannaki, S. M. (2015). ETFs performance europe - A good start or not? *Procedia Economics and Finance 30*(15), 955–966. https://doi.org/10.1016/S2212-5671(15)01346-5.

Figure 1. ETFs Take Over



Source: Financial Times, November 2017 (www.ft.com)

Figure 2. Evolution of ETF Products Worldwide



Source: Statista, 2018 (www.statista.com)

Figure 3. PPP ETF Daily Return

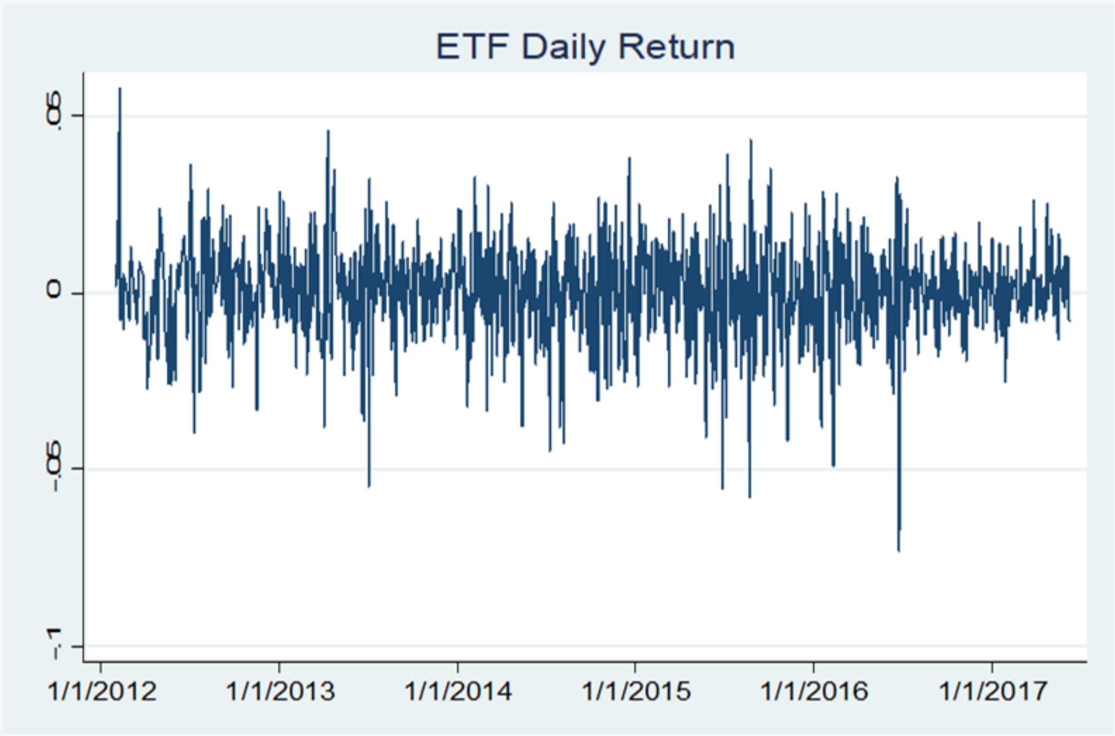


Figure 4. PSI 20 Daily Return

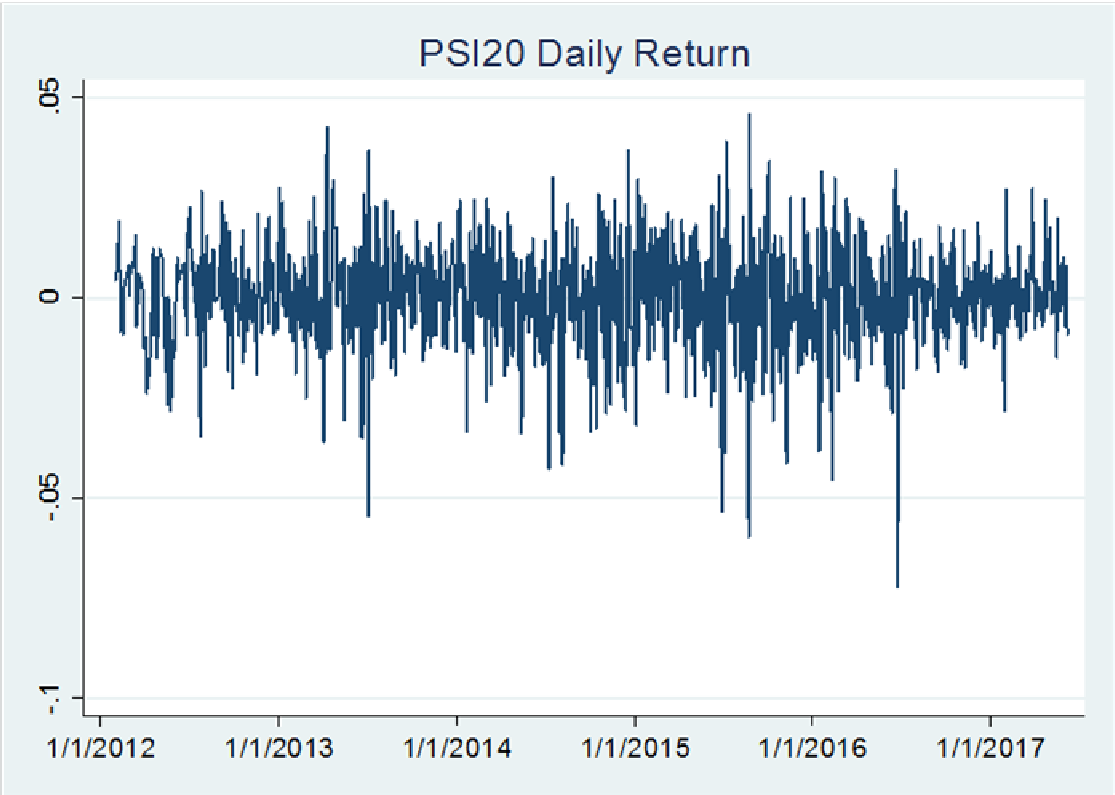


Figure 5. Plot of Squared Residuals and Linear Prediction

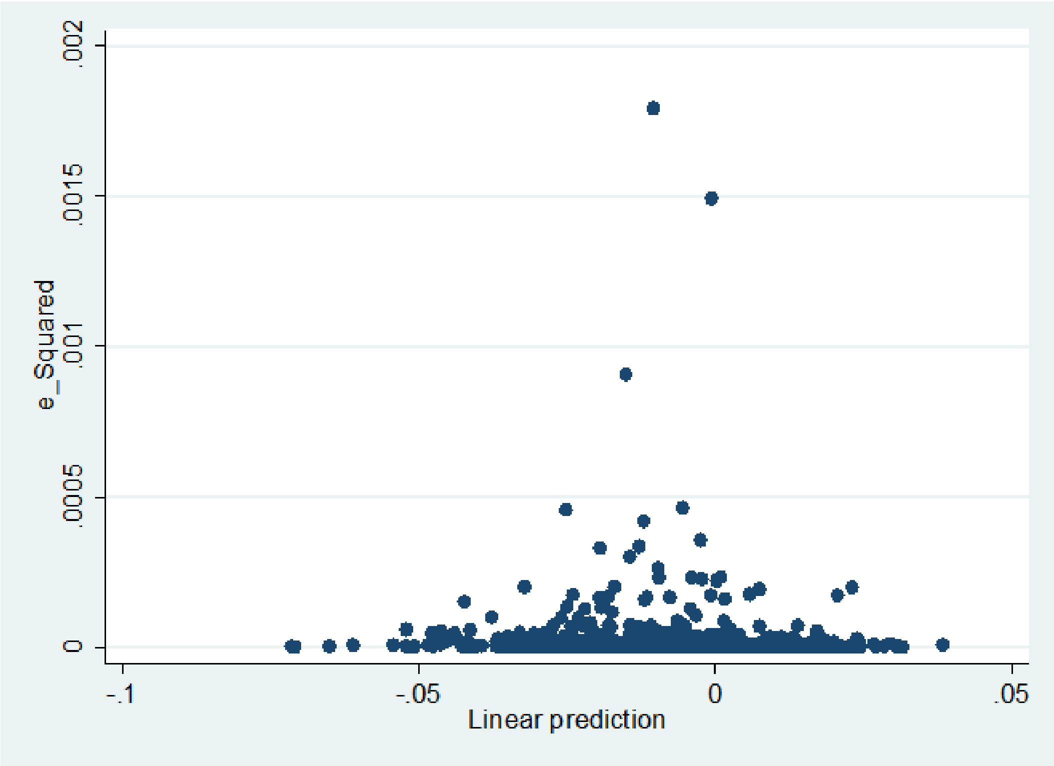


Figure 6. Excess ETF Return Relative to the Benchmark

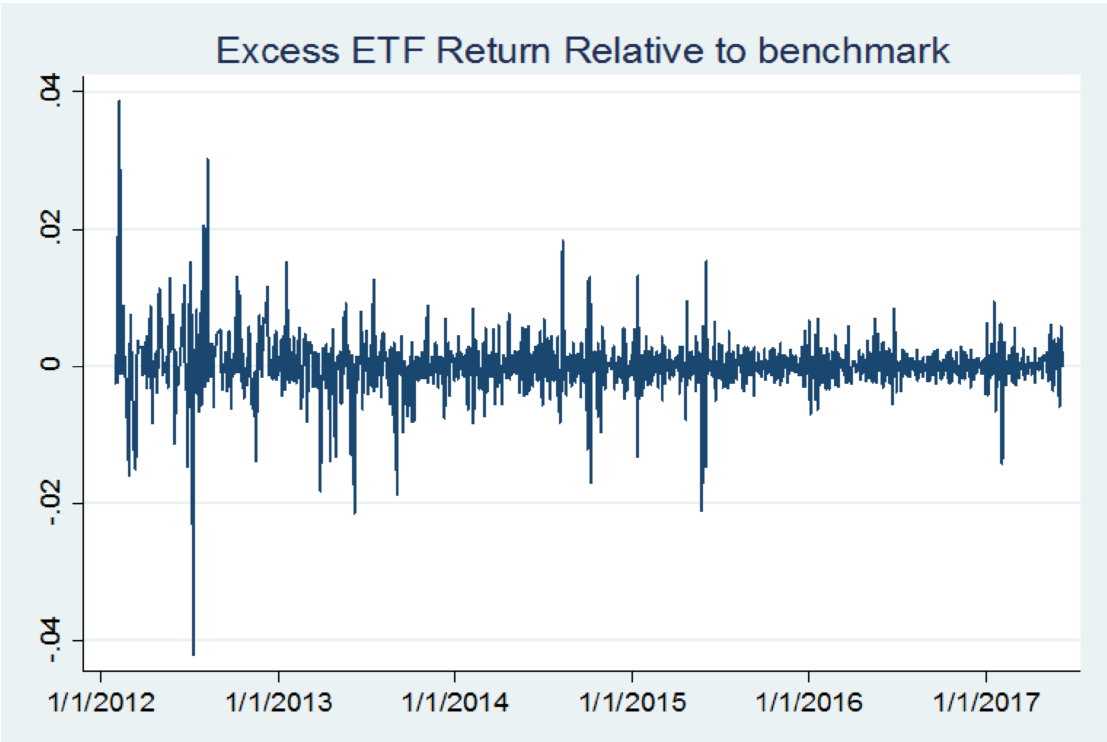


Figure 7. Testing Abnormal Return

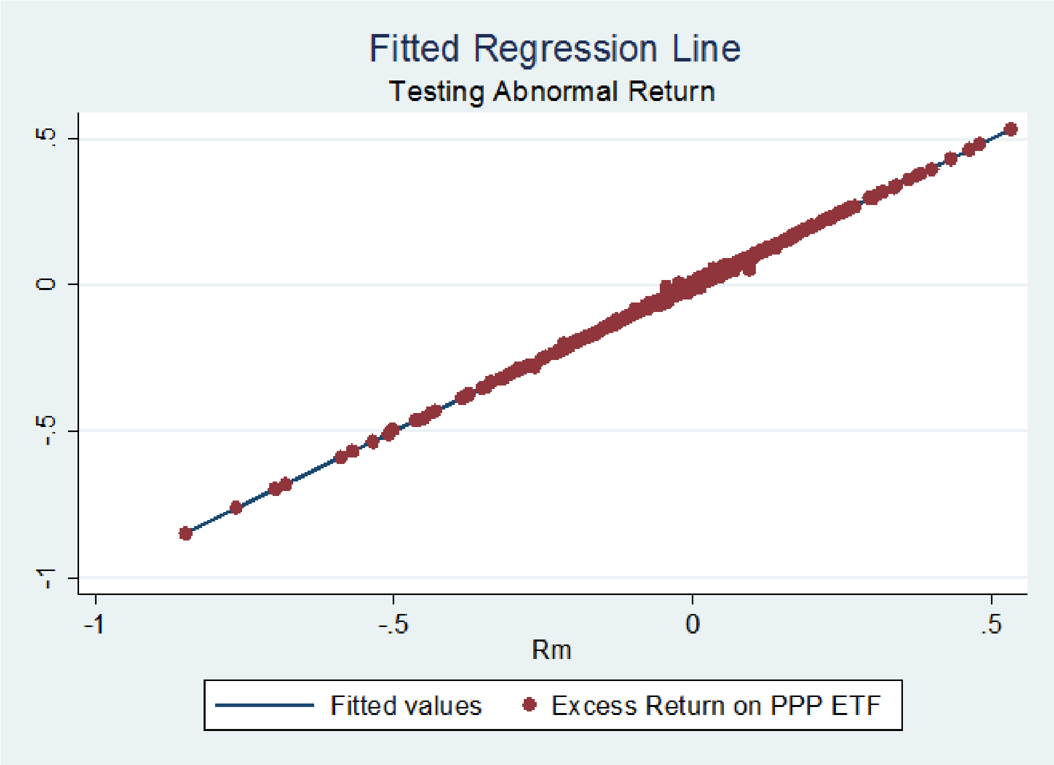


Table 1. Description of Variables

Variables	Description	Units	Source
ri	Return on ETF.	Percentage	Bloomberg
rm	Return on the benchmark.	Percentage	Bloomberg
Ri	Excess return on ETF.	Percentage	Bloomberg
Rm	Excess return on market indexes.	Percentage	Bloomberg
Rf	Risk-Free rate (German Bund 10-Year Yield).	Percentage	Bloomberg

Table 2. Descriptive Statistics

This table presents the summary statistics of the main variables used in our study aimed to test the ability of the PPP ETF to beat its underlying benchmark (PSI 20 index). ri stands for the daily return on the ETF; rm is the market return; Ri and Rm are the expected excess return earned by the ETF and by the inherent benchmark, respectively. Rf is the risk-free rate and N is the number of observations.

Variable	Mean	Std. Deviation	Min	Max	Skewness	Kurtosis	Median	N
ri	0.0122	1.3488	-7.3262	5.7820	-42.5047	485.5904	0.0839	1,233
rm	0.0008	1.2945	-7.2468	4.6042	-48.0658	494.8714	0.0344	1,233
Ri	-0.8716	1.4763	-7.2742	3.8040	-37.6218	384.0543	-0.7806	1,233
Rm	-0.8830	1.4287	-7.1948	3.8512	-32.8270	392.7456	-0.8149	1,233
Rf	0.8838	0.6275	-0.1840	2.0520	13.0460	161.0775	0.7730	1,233

Table 3. Correlation Matrix

This table presents the pairwise correlations among the variables included in our study. ri stands for the daily return on the ETF; rm is the market return; Ri and Rm are the expected excess return earned by the ETF and by the inherent benchmark, respectively. Rf is the risk-free rate and N is the number of observations

Variables	(1)	(2)	(3)	(4)	(5)
(1) ri	1.000				
(2) rm	0.954*	1.000			
(3) Ri	-0.013	-0.020	1.000		
(4) Rm	-0.020	-0.019	1.000*	1.000	
(5) Rf	0.088*	0.093*	-0.895*	-0.895*	1.000

* denotes significance at the 10% level.

Table 4. Performance Analysis

This table presents the computed indicators of reward-to-volatility for each ETF, using Sharpe, Treynor, and Sortino ratios. TE1 correspond to the absolute tracking error 1 and TE2 is the standardized version of tracking error.

ETF	Sharpe ratio (%)	Treynor ratio (%)	Sortino ratio (%)	TE1	TE2	N
PPP ETF	-3.767	-0.569	-3.590	0.00007	0.00388	1,233
PSI 20	-3.828	-0.578	-3.655	0.000	0.000	1,233

Table 5. Testing Abnormal Return

This table provides the regression results of the first model testing the abnormal return. We use robust OLS regression to test the ability of the PPP ETF to beat its underlying benchmark. Additionally, we provide results for the mean comparison test. The computed difference between ETFs and benchmark returns, as well as the estimated t-stat are presented on the right-side of the table. The dependent variables are the excess return on ETFs (Ri), and the independent variables are the excess return on the market (Rm).

	(1)	(2)
Model	Robust OLS	Mean Difference Test
Variables	Ri	Ri
Rm	1.000***	0.993***
	(0.000480)	(0.00738)
Constant	0.00009	0,0000553
	(0.000117)	(0.000127)
R-Squared	0.924	
Mean Difference in Returns		.0000875
t-stat		0.0143
N	1,220	1,220