NOVA IMS Information Management School

MEGI

Mestrado em Estatística e Gestão de Informação

Master Program in Statistics and Information Management

A comparison of the performance of Green bond funds, bond Mutual funds and bond ETFs

Ana Rita Rosado Rodrigues

Dissertation presented as partial requirement for obtaining the Master's degree in Statistics and Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

A COMPARISON OF THE PERFORMANCE OF GREEN BOND FUNDS, BOND MUTUAL FUNDS AND BOND ETFS

by

Ana Rita Rosado Rodrigues

Dissertation presented as partial requirement for obtaining the Master's degree in Statistics and Information Management, with a specialization in Risk Analysis and Management.

Advisor: Prof. Luís Alberto Ferreira de Oliveira, PhD

February 2023

ACKNOWLEDGEMENTS

Reaching the end of this dissertation was not easy, so I would like to share my gratitude.

To my family, for their unconditional support, concern and for giving me the opportunity to study and deliver this final work.

To my boyfriend, for all his advice, encouragement, patience and motivation. My greatest support throughout this process.

To my team at CGD, especially my director Isabel and my colleagues Pedro and Henrique, for understanding my position in the last month and for providing me all the support I needed.

To Professor Jorge Bravo, without whose help I would not have found the most appropriate topic and advisor for this dissertation.

Finally, to Professor Luís Oliveira, for accepting the challenge of being my advisor, for all the availability he has shown and all the wisdom he has shared.

Thank you all.

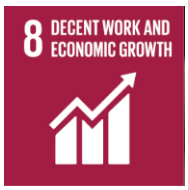
ABSTRACT

This dissertation measures and compares the performance of green bond funds, bond mutual funds and bond ETFs, all domiciled in the European region, from January 2005 to December 2019. This period has been divided into three subperiods in order to analyse the performance before, during and after the financial crisis of 2007-2008. The sample consists of monthly data for a total of 3,484 funds and their performance was assessed by using traditional risk-adjusted measures, namely Sharpe ratio, Treynor ratio, and Jensen's Alpha. The main findings show that, on average, bond mutual funds outperformed bond ETFs and green bond funds in all the studied subperiods. Furthermore, when analysing each fund category separately, all fund categories have performed best during the crisis period, which can be considered a fly-to-safety event, where the prices of safer assets tend to rise. Regarding the performance of green bond funds, although they have outperformed their peers at some points in time, there is no clear evidence to support this. However, some investors may prefer to invest in this type of asset due to the green bond funds' environmental contribution.

KEYWORDS

Risk-adjusted performance; Green bond fund; Bond mutual fund; Bond ETF; ESG; Financial crisis

Sustainable Development Goals (SGD):



INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ARCH	Autoregressive Conditional Heteroscedasticity
CAPM	Capital Asset Pricing Model
EGARCH	Exponential Generalized Autoregressive Conditional Heteroscedastic
ESG	Environmental, Social and Governance
ETF/ETFs	Exchange-Traded Fund / Exchange-Traded Funds
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
LAP	Loss-Averse Performance
NAV	Net Asset Value
OLS	Ordinary Least Squares
SRI	Socially Responsible Investment

1. INTRODUCTION

For some time now we are facing a climate change that results from two centuries of accumulated unsustainable development. Global emissions of greenhouse gases are rising. According to World Meteorological Organization (2021), concentrations of greenhouse gases in the atmosphere today are 149 per cent higher than pre-industrial levels and annual rates of increase have never been so high.

Over the past decade, people have shown signs of awareness in several areas of the financial system through capital mobilization to green activities. Green and sustainable finance can help improving environmental sustainability, reducing carbon emissions and developing a strong climate infrastructure (Agliardi & Chechulin, 2020). The world is increasingly concerned about this topic and the managers and investors themselves want to preserve the environment and the world that we live in. In order to do that, some companies are issuing dept in the form of green bonds.

One of the biggest drivers of investor awareness of risk management and sustainability was the financial crisis of 2007-2008. When the US housing market collapsed, it triggered a sub-prime crisis that led to the insolvency of some banks and other financial institutions. This quickly spread to the rest of the financial system and had a global impact, with banks and governments having to take emergency measures to stabilise the financial system and avoid a total collapse of the global economy (Samarbakhsh & Shah, 2021).

The crisis reinforced the need for a more responsible and sustainable approach, resulting in a greater focus on environmental, social and governance (ESG) issues. Nevertheless, the financial crisis of 2007-2008 can be seen as a turning point for the green bond market, as it gained greater visibility (Sampei, 2018).

As the green bond market has developed, so has the market for green bond funds. As an asset that can be invested in a diversified way, managed by professionals and with lower risk, it has gained popularity especially among private investors. However, it is useful to understand whether these green bond funds offer the same or better returns than their non-green peers.

There have been several academic studies on the performance of mutual funds and ETFs, although not so many on bond funds in particular, and even fewer that includes green bond funds. Given the ever-increasing environmental concerns around the world, as well as the important political decisions and economic changes of the last decade, green investing is a very interesting subject to study. As far as the author is aware, there is also a visible lack of studies on the European market, as many studies are carried out on the US market.

So, to address these gaps, the main purpose of this dissertation is to conduct a study on the comparison of the performance of funds that invest on green bonds, bond mutual funds and Exchange-Traded Funds (ETFs) that only invest on bonds, all European domiciled.

In order to achieve this goal, it is necessary to analyse the risk-adjusted returns of green bond funds, bond mutual funds and bond ETFs. In addition, the performance of these three categories of funds over the chosen period, which includes the financial crisis of 2007-2008, will also be analysed. To this end, the models used in this thesis include the Sharpe Ratio, the Treynor Ratio, as well as the Jensen's Alpha.

This dissertation is structured as follows. In chapter 2, a theoretical approach is made by presenting some concepts and models regarding sustainable investing and performance measurement, through a literature review. Further, in chapter 3, is presented the methodology and the data used to achieve the study objective. The results are presented and discussed in chapter 4 and the respective conclusions are stated in chapter 5, along with the limitations identified during the course of this study and the suggestions for future works.

2. LITERATURE REVIEW

2.1. SUSTAINABLE INVESTING

Over the centuries, bonds have become a stable source of income for institutions and individuals, such that, some authors believe that bonds are an important part of a balanced portfolio. A bond is a dept capital market instrument in which an investor loans money to an entity, in general a corporation or a government. The borrower uses the money to fund its operations and the creditor receives interest on the investment. (Deribew, 2017)

In practical terms, the issuer makes regular interest payments – cash flows – to the investor at a specified rate – coupon rate – that can be fixed or floating coupon, on the face amount (amount borrowed) until a specified date – the maturity date. Once the bond matures, the interest payments stop, and the borrower is required to repay the face amount of the principal to the investor. This type of bond is known as conventional or plain vanilla bond. (Choudhry, 2010) Due to these regular interest payments and the agreement of the rate and date, bonds are commonly known as fixed-income instruments.

For investors, bonds have become a stable source of income and a relatively low-risk form of investment. For corporations and governments, bonds are a financing source, because they permit these entities to obtain a large amount of capital without giving up equity ownership in their companies.

Being able to reduce risk and achieve stable investment performance is fundamental for any investor. However, this can be a challenge for non-professional investors, so instead of investing in individual bonds, they can purchase bond funds, which hold a diversified portfolio of bonds and offer the benefits of professional management. While bonds are considered a more conservative choice, according to Stankevičienė & Petronienė (2019), bond mutual funds and Exchange-Traded Funds (ETFs) are becoming increasingly attractive to risk-averse investors.

A bond mutual fund is an indirect method of investing in bonds and is considered to be a pool of investments that invests in several bonds or fixed-income securities. This type of fund is actively managed by professionals, which provide a well-diversified portfolio, but can also lead to high management costs (Bodie et al., 2014). Open-end funds are the most common type of mutual fund.

An alternative to bond mutual funds are bond Exchange-Traded Funds (ETFs), which are funds that track the movements of a certain index. The benchmark value of the index (which oscillates depending on market fluctuations) is the aggregation of the individual values of the index bonds, thus replicating the risk and return characteristics of the index (Hull, 2018). ETFs can be actively or passively managed by professionals.

According to Bodie et al. (2014), an ETF can be a better investment than a mutual fund. ETFs trade continuously, so compared to mutual funds, whose net asset value is quoted and therefore calculated only once a day, investors can buy a share more easily (at any time of the day). Another advantage of ETFs in comparison to mutual funds is that when investors decide to redeem their shares, the mutual fund managers have to sell some securities in which the fund has been invested in order to provide the necessary cash to repay the investors, whereas in ETFs this is not necessary as another investor

holds their positions (Hull, 2018). In addition, ETFs tend to offer lower management fees and taxes for the investor.

The fund industry took a turn for the worse in 2007-2008, when the global financial crisis hit. The starting date of the crisis is not unanimous for all authors. However, the majority agree that the crisis began in July 2007, when the Federal Reserve System and the Bank of England began injecting money into the economy. The crisis was caused by the bursting of the US housing bubble, driven by a combination of risky lending practices, a rapid increase in the number of subprime mortgages and a lack of regulatory supervision (Samarbakhsh & Shah, 2021). The value of mortgage-backed securities, which were widely held by investment funds, fell rapidly. This situation quickly evolved into a banking and credit crises, which affected the global economy and led to sovereign debt and economic crises. It was only after 2012 that the economy began to stabilise, after some European governments intervened and asked for bailouts.

During the crisis, many investment funds suffered huge losses and many investors saw the value of their portfolios decline massively. As a result, investors wanted to sell their positions and redeem their investments, which led to a sharp decline in the value of many investment funds. With this redemption, many investors opted for a less risky and more stable option, such as bond funds. (Filip et al., 2015).

Figure 2.1 - Examples of issues of ESG factors
Source: Vanguard, 2022

Sampei (2018) states that the 2007-2008 financial crisis disrupted the financial industry's strategy of focusing solely on competitive returns and brought new attention to the importance of considering Environmental, Social and Governance (ESG) factors in investment decisions. The integration of ESG considerations into investment processes has received increased attention, since many investors recognise that ESG factors can have a significant impact on a company's long-term financial performance. These ESG investments are also known as Socially Responsible Investment (SRI) or sustainable investment.

Environmental, Social and Governance principles are a set of standards for a company's operations that conscious investors use to screen possible investments. These three distinct factors cover a wide range of potential issues, as shown in Figure 2.1.

Sustainable Investment has become a major phenomenon in the financial world and is growing in popularity. Nizam et al. (2019) states that if financial institutions increased their access to ESG or environmental finance, their performance would improve. Dixon-Fowler et al. (2012) found empirical evidence of efficiency gains from environmental performance, and that it strongly influences market measures of financial performance.

The environmental aspect of ESG focuses on a company's impact on the environment and its efforts to reduce its carbon footprint. According to Inderst & Stewart (2018), this criteria typically includes climate change, carbon emissions, pollution, resource efficiency and biodiversity. Therefore, environmental criteria allow investors to make informed decisions and support corporations working towards a more sustainable future.

The best-known environmental investment is green bonds. By definition, green bonds are fixed-income securities with the same characteristics as other bonds, however they are bonds whose purpose is to raise capital to finance projects with specific climate or environmental benefits (Inderst & Stewart, 2018). As noted above for conventional bond funds, it is easier for retail investors to invest in green bond funds rather than individual green bonds, which may not be financially rewarding or easily accessible.

There have been several studies in this area. Chang et al. (2012) concluded that green mutual (equity) funds underperform when compared to other mutual funds in the same fund category. Paradinovic (2017) compared the performance of Islamic, SRI and green mutual funds and the findings obtained show that green mutual equity funds outperform both SRI and Islamic mutual equity funds, on average. Németh-Durkó & Hegedűs (2021) stated that the green bond portfolio underperformed the benchmark indices.

2.2. Performance Measurement

Typically, investors have two conflicting objectives. They expect to earn the highest possible return on their investments and, at the same time, they aim to minimize the risk, which is just another way to say that they want the lowest possible chance of losing money. There are several ways to measure investment performance. Vaidya (2017) defines risk-adjusted return as:

"(…) a technique to measure and analyse the returns on an investment for which the financial, market, credit and operational risks are analysed and adjusted so that an individual can make a decision on whether the investment is worth it with all the risks it poses to the capital invested."

There is a long list of academic studies that combine different indicators to assess fund performance. The most popular measures for the evaluation of the risk-return profile of the investment are: Sharpe ratio (Sharpe, 1966), Treynor ratio (Treynor, 1965) and Jensen's alpha (Jensen, 1968).

First introduced by economist William F. Sharpe (1966), the Sharpe ratio, formerly known as the "Reward-to-Variability ratio", is often used by professionals as a reference performance measure for comparing the potential returns of investment portfolios against their underlying risks. This ratio is a relative risk-adjusted performance measure that uses the expected return of the portfolio in excess of a risk-free rate as numerator and the standard deviation of the same portfolio as risk measure.

Because of its simplicity, this ratio is one of the most widely used performance measures to compare portfolios. However, it is important to note that the Sharpe ratio assumes that the average return on an investment is normally distributed. In other words, it assumes that most returns are symmetrically grouped around the mean, with fewer returns in the tails of the curve. This could be a limitation, as investment returns may not follow a normal distribution.

To deal with issues such as skewness, kurtosis and fat tails over time, several adjustments to the Sharpe ratio have been developed. The Adjusted Sharpe ratio, developed by Pézier and White (2006), considers skewness and kurtosis. Later, to account for the fact that the Sharpe ratio becomes negative when the risk-free rate is higher than the return on the portfolio – which can occur during large and prolonged market downturns – Israëlsen (2005) developed the Modified Sharpe Ratio. However, despite these limitations, the Sharpe ratio (1966) is still considered the reference measure.

Later, Sharpe (1994) proposed another measure of performance, called Information ratio, which assesses the quality of the portfolio manager's investment decisions. The main difference between this ratio and the Sharpe ratio is that it uses the returns of the benchmark instead of the risk-free rate, and the measure of risk is given by the standard deviation of the difference between portfolio and benchmark returns (tracking error).

Treynor (1965) proposed two "Reward-to-Volatility ratios", both based on the systematic risk of the portfolio, i.e., the beta, which is obtained from a single-index regression. The first ratio is the Treynor ratio, in which the manager's compensation is evaluated relative to the market risk exposure. The second ratio is the Appraisal ratio, which measures the excess performance achieved by the manager over the risk premium offered by the market.

Several authors have proposed alternative versions of the Treynor ratio. Srivastava & Essayyad (1994) replaced the traditional beta with a modified version defined as the ratio between partial moments, while Bacon (2008) proposed to multiply the systematic risk sensitivity of the investor's portfolio return by the total risk of the market portfolio.

Regarding the absolute risk-adjusted performance measures, the most popular one is the Jensen's alpha, introduced by Jensen (1968). This measure is based on the Capital Asset Pricing Model (CAPM), which is a single-factor model, and evaluates the capabilities of the portfolio manager. There are some variants of the Jensen's alpha, such as the Zero-beta CAPM suggested by Black (1972), which is also

based on a single-index model. This measure uses the expected return of a zero-beta portfolio instead of the risk-free rate as Jensen's alpha.

Another variant is the Net Selectivity index proposed by Fama (1972), which translates the excess return earned by the manager that could not have been earned by investing in the market portfolio. It compares the excess return earned by the manager with a specific risk and the excess return that could have been earned with the same amount of systematic risk. Later, Connor & Korajczyk (1986) presented a multi-factor model, which is a generalization of the CAPM model, that identifies the relevant risk factors by using traditional model specification techniques.

Other types of performance measures have been developed. These include those based on some general characteristics of the return distribution, for instance, the Bernardo & Ledoit (2000) Gain-Loss ratio and the Keating & Shadwick (2002) Omega measure, the Loss-Averse Performance (LAP) measures (Simple LAP and House-money LAP) developed by Gemmill et al. (2006), the Sortino et al. (1999) Upside-Potential ratio and, the Rachev ratio and Generalized Rachev ratio both developed by Biglova et al. (2004).

3. DATA AND METHODOLOGY

3.1. METHODOLOGY

The use of returns to measure performance has proven to be inadequate as it does not account for risk. Hence, risk-adjusted measures are used for the evaluation and comparison of the performance of green bond funds, bond mutual funds and bond ETFs.

Similar to Jaksic et al. (2015) who used Sharpe ratio, Treynor ratio and Jensen's alpha to examine mutual fund performance; to Paradinovic (2017) who used, among others, the same measures to study the performance of bond mutual funds and bond ETFs; and to Filip et al. (2015) who equally used the same three measures to evaluate the performance of Romanian bond funds, the assessment of the risk-adjusted performance in this dissertation is also made by the Sharpe ratio (1966), the Treynor ratio (1965) and the Jensen's alpha (1968). These measures are described in more detail in the following subsections.

In order to apply the referred methods, it is necessary to obtain the monthly logarithmic returns for each fund after extracting their Net Asset Values (NAVs). The NAV is a value per asset, i.e., the price for buying or selling one unit of a fund. (The details of how the data were obtained are described in the section 3.2.). These returns are calculated by the difference of the lognormal NAVs to better approximate the distribution of returns to a normal distribution (Bodie et al., 2014).

$$R_{i,t} = \ln\left(\frac{NAV_{i,t}}{NAV_{i,t-1}}\right) = \ln(NAV_{i,t}) - \ln(NAV_{i,t-1}) \quad (1)$$

Where:

$R_{i,t}$ – Return of the fund i at month t

$NAV_{i,t}$ – Net Asset Value of the fund i at month t

$NAV_{i,t-1}$ – Net Asset Value of the fund i at month $t-1$

Similarly, once the benchmark prices have been obtained, it is also necessary to calculate their monthly returns, using the same method:

$$R_{m,t} = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(P_t) - \ln(P_{t-1}) \quad (2)$$

Where:

$R_{m,t}$ – Return of the market at month t

P_t – Closing price of an index at month t

P_{t-1} – Closing price of an index at month $t-1$

Regarding the risk-free asset, its returns are also calculated using the monthly logarithmic method shown in equations (1) and (2). However, this first requires a transformation to obtain the price of a zero-coupon bond. This price is given by:

$$P_{ZC} = \frac{F}{1 + \frac{r}{12}} \tag{3}$$

Where:

P_{ZC} – Zero-coupon bond price at month t

F – Face value of the bond

r – Risk-free rate of the asset

Moreover, as the computed fund returns are expressed using continuous compounding, their volatility can be defined as the historical standard deviation of the monthly returns provided by fund. The volatility of the fund i is given by:

$$\sigma_i = \sqrt{\sigma_i^2} = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}\left(R_{i,t} - \bar{R_i}\right)^2} \tag{4}$$

Where:

σ_i^2 – Variance of the fund i expected return

$R_{i,t}$ – Return of the fund i at month t

$\bar{R_i}$ – Historical average of return, $\frac{1}{n}R_{i,t}$

n – Number of months

An alternative to this calculation would be to model volatilities using the autoregressive conditional heteroscedasticity (ARCH) model (Engle, 1982), the generalized autoregressive conditional heteroscedasticity (GARCH) model (Bollerslev, 1986) or the exponential generalized autoregressive conditional heteroscedastic (EGARCH) model (Nelson, 1991). These models attempt to capture variations in volatility, recognising that it is not constant through time. Nevertheless, in this dissertation, the volatility is estimated using the standard approach, as it is simpler and more straightforward.

The following subsections describe the risk-adjusted measures used to assess portfolio performance. The next referenced excess return of a fund over the risk-free rate for the month t is calculated using the last available risk-free rate in that month. Meaning that, for example, for the March 2006 excess return calculation, the last available rate for making a risk-free investment was February 2006.

3.1.1. Sharpe Ratio

The Sharpe ratio (Sharpe, 1966) measures the excess return of a fund relative to the risk-free rate for a given period, considering the volatility of the fund's returns. It is given by the following equation:

$$Sharpe\ Ratio = \frac{R_{i,t} - R_{f,t}}{\sigma_i} \quad (5)$$

Where:

$R_{i,t}$ – Return of the fund i at month t

$R_{f,t}$ – Return of the risk-free asset at the beginning of month t

σ_i – Standard deviation of the fund i returns

As already referred, the standard deviation is a measure of the volatility of the returns, and it is calculated by equation (4).

As far as the interpretation of the Sharpe ratio is concerned, it is quite straightforward. The higher the ratio, the better the performance and therefore the higher the return earned for each unit of risk taken (volatility).

3.1.2. Parameters Estimation: The Single-Index Model

Based on Markowitz (1952), Sharpe (1963) introduced the single-index model that describe the relationship between the returns of an individual asset and the overall market. This model is presented as:

$$\left(R_{i,t} - R_{f,t}\right) = \alpha_i + \beta_i \times (R_{m,t} - R_{f,t}) \quad (6)$$

Where:

$R_{i,t}$ – Return of the fund i at month t

$R_{f,t}$ – Return of the risk-free asset at the beginning of month t

$R_{m,t}$ – Return of the market at month t

α_i – Alpha of the fund i (Jensen's alpha)

β_i – Beta of the fund i

The variables α_i and β_i are estimated using an Ordinary Least Squares (OLS) regression[1].

Beta Estimation

Beta (β_i) is the systematic or market risk coefficient, i.e., it measures the volatility of returns relative to the overall market. Systematic risk represents the non-diversifiable risk or the risk that remains even after extensive diversification. This risk affects the market as a whole rather than a specific individual

[1] Ordinary Least Squares is a commonly used statistical method for estimating the parameters of a linear regression model. When certain assumptions are met, OLS provides the best linear unbiased estimate of the regression coefficients.

asset. (Bodie et al., 2014) As such, it is generally considered to be the level of risk that investors are compensated for taking on.

If the beta is negative, it means that the portfolio is negatively correlated with the market, i.e., as market risk increases, the portfolio's returns decrease, or vice-versa. A beta of 1 means that the volatility of the portfolio is perfectly correlated with the market, suggesting that the movement of the portfolio reflects the movement of the market. Therefore, a portfolio with a beta greater than 1 indicates a higher level of risk compared to the market average risk and a beta below 1 indicates that the portfolio is less risky than the market average risk.

Alpha Estimation

When investing in a fund, investors often consider how the manager has contributed to the fund's performance. Looking at alpha (α_i) is one way to measure whether a fund manager has added value beyond simply investing in the index.

3.1.3. Treynor Ratio

Another relative risk-adjusted measure is the Treynor ratio (Treynor, 1965). The purpose of this measure is to determine whether an investor has been adequately compensated for the risk taken by being above the market. It is calculated using the following equation:

$$Treynor\ Ratio = \frac{R_{i,t} - R_{f,t}}{\beta_i} \tag{7}$$

Where:

$R_{i,t}$ – Return of the fund i at month t

$R_{f,t}$ – Return of the risk-free asset at the beginning of month t

β_i – Beta of the fund i

Comparing equation (7) with equation (5), it is possible to see that while the Sharpe ratio uses standard deviation as a measure of volatility, the Treynor ratio uses the beta coefficient. The beta used to calculate equation (7) is obtained from the regression described in subsection 3.1.2.

In terms of interpreting the Treynor ratio, the higher the ratio, the better the performance and therefore the higher the return earned for each unit of systematic risk taken. A ratio less than zero shows that the investment has underperformed the market, while a ratio greater than zero shows outperformance.

3.1.4. Jensen's Alpha

The Jensen's alpha (Jensen, 1968) is an absolute measure defined as the expected return of the investor's portfolio relative to the market returns given its level of systematic risk. It is formulated as follows:

$$Jensen's\ alpha = R_{i,t} - \left[R_{f,t} + \beta_i \times (R_{m,t} - R_{f,t})\right] \quad (8)$$

Where:

$R_{i,t}$ – Return of the fund i at month t

$R_{f,t}$ – Return of the risk-free asset at the beginning of month t

$R_{m,t}$ – Return of the market at month t

β_i – Beta of the fund i

In terms of interpreting the Jensen's alpha, this measure assesses the contribution of the manager's decisions to the performance of the portfolio. The higher the alpha, the better the manager's performance. Moreover, if Jensen's alpha is negative means that the fund manager has underperformed the market, whereas if it is positive indicates that the fund manager has outperformed the market.

3.2. DATA

For the analysis of the performance differences between green bond funds, bond mutual funds and bond ETFs, data was obtained from the Bloomberg platform. A fund screening tool was used to generate a list of funds based on a set of search criteria. Therefore, some common filters were applied. Only funds with an active market status and a primary share class of 'Yes' were included, to avoid possible bias from the variety of investment classes that have the same attributes and are managed by the same fund managers. Moreover, as the main intention of this study is bond funds, only fixed income was selected in the fund asset class focus. By Bloomberg's definition, a fund classified as fixed income means that at least 80% is invested in fixed income securities.

Apart from using common criteria, several specific filters were applied to distinguish each category of fund. For mutual funds, the fund type selected was open-ended and for ETFs, as the name suggests, the type of fund was "ETF". For green funds, since they are both mutual funds and ETFs, the general attribute "Environmentally Friendly" was used in addition to the previous filters. In order to avoid overlapping data in the sample, this last attribute was excluded from the first two categories of funds.

Furthermore, for the purposes of this dissertation, were considered funds domiciled in the European region and with a time horizon from January 2005 to December 2019. This period was chosen in order to assess the potential impact of the 2007-2008 global financial crisis on the performance of the above-mentioned funds. Funds with no data or with less than six months of data available on Bloomberg were excluded from the sample. Funds that disappeared or were merged/integrated into other funds were also not considered. Thus, the sample is not completely free from survivorships bias because it only considers the funds with uninterrupted end-of-the-month NAV between January 2005 and December 2019 to derive monthly rates of return.

Given that the prices were quoted in different currencies (e.g., US Dollar, British Pound, etc.) and to avoid dealing with multi-currency data, the prices were all collected in Euro. The NAVs do not include management fees, therefore the analysis and the results will be presented from an investor's perspective.

The application of the before mentioned methodology also requires a benchmark and a risk-free rate. The Bloomberg Euro Aggregate Bond Total Return Index was chosen as the benchmark and "proxy" for the European fixed income market. This index measures the market for investment grade, fixed rate bonds denominated in Euro, including Treasuries, government, corporate and securitised issues, and is therefore aligned with the portfolios under analysis. Although specific indices exist for benchmarking green bonds, it was decided to use a single market benchmark to facilitate interpretation and comparison of the performance of the three portfolios. The Bloomberg Euro Aggregate Bond Total Return Index, launched in June 1998 and maintained by Bloomberg, is also widely used by investors and fund managers as a benchmark for the performance of euro-denominated fixed income securities. It provides a comprehensive view of the eurozone bond market and is an important tool for assessing the performance of investment portfolios.

For the risk-free rate, consideration was given to using one-month Germany Government Bond, however, the one-month Euribor was chosen as it is a generally accepted market reference and has a high liquidity provided by the eurozone interbank market. Euribor rates are based on the interest rates at which European credit institutions borrow funds from each other. This market rate referential

reflects highest daily trading volume of business and the market intervenients denote a best quality credit rating, high ethical standards, and an excellent reputation. The level of Euribor rates is primarily determined by the law of supply and demand, but there are also external factors that can influence this level, such as inflation and economic growth (European Money Markets Institute, 2023). Both the benchmark index and the risk-free rate have been extracted from the Bloomberg platform.

Overall, the final sample comprises 3,100 bond mutual funds, 328 bond ETFs and 56 green bond funds, which translates in a total of 3,484 funds. Figure 3.1 shows the distribution of total number of bond funds in each year of the sample period (2005-2019).

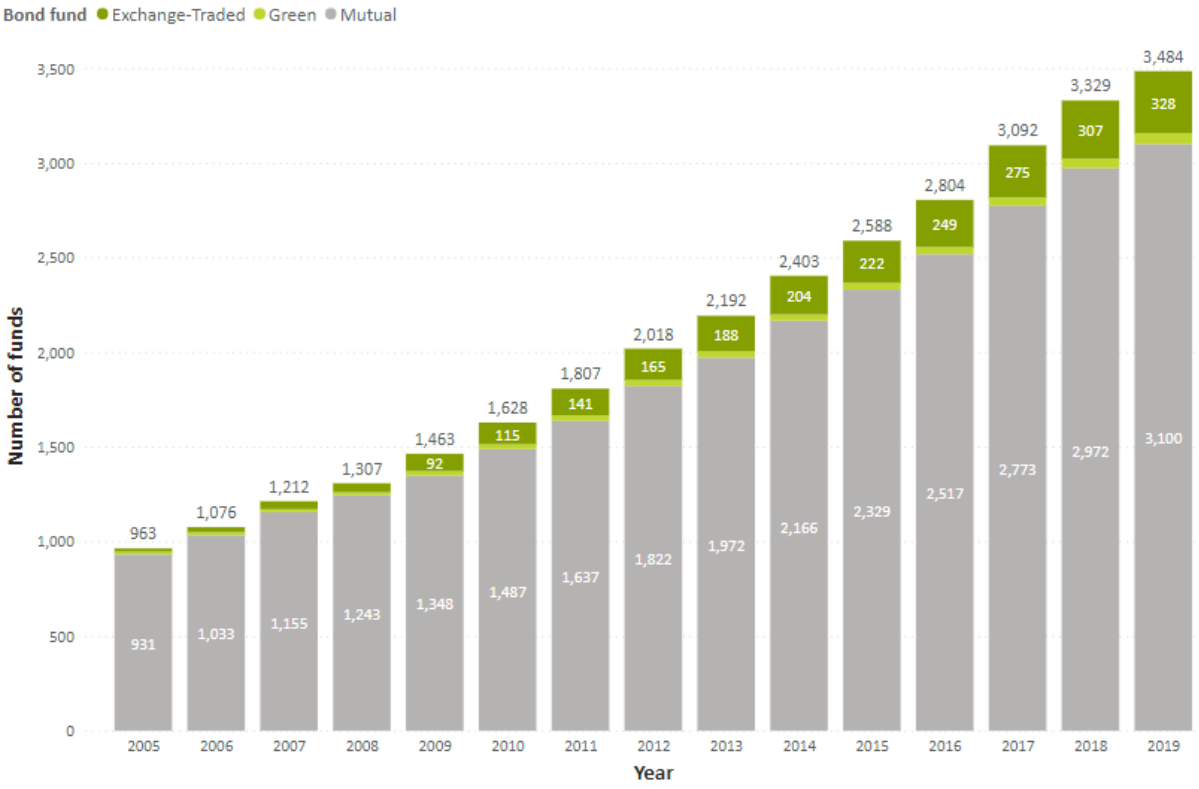


Figure 3.1 - Total number of funds by year

The overall increase in the total number of funds observed is remarkable, with an increase of around 262% since the first year. As there were 963 bond funds in 2005 and 3,484 bond funds in 2019. The largest increase was in bond ETFs, with an average annual growth of 131% over the period, followed by green bond funds, which have reported an average annual growth of 20% throughout the observed period. Finally, the smallest, but still notable, growth was seen in bond mutual funds, which increased by 17% per year, on average.

At the end of the observed period, bond mutual funds represent 89% of the whole sample, being the biggest group of funds considered. Bond ETFs account for 9.4% of the total funds considered, making them the second largest group, while green bond funds made up the remaining 1.6% of the sample, representing the smallest number of observed funds.

In addition, rather than just looking at the total number of observations, it might also be interesting to look at how many new bond funds were added to the sample each year. The Figure 3.2 displays the number of new bond mutual funds, bond ETFs and green bond funds launched each year.

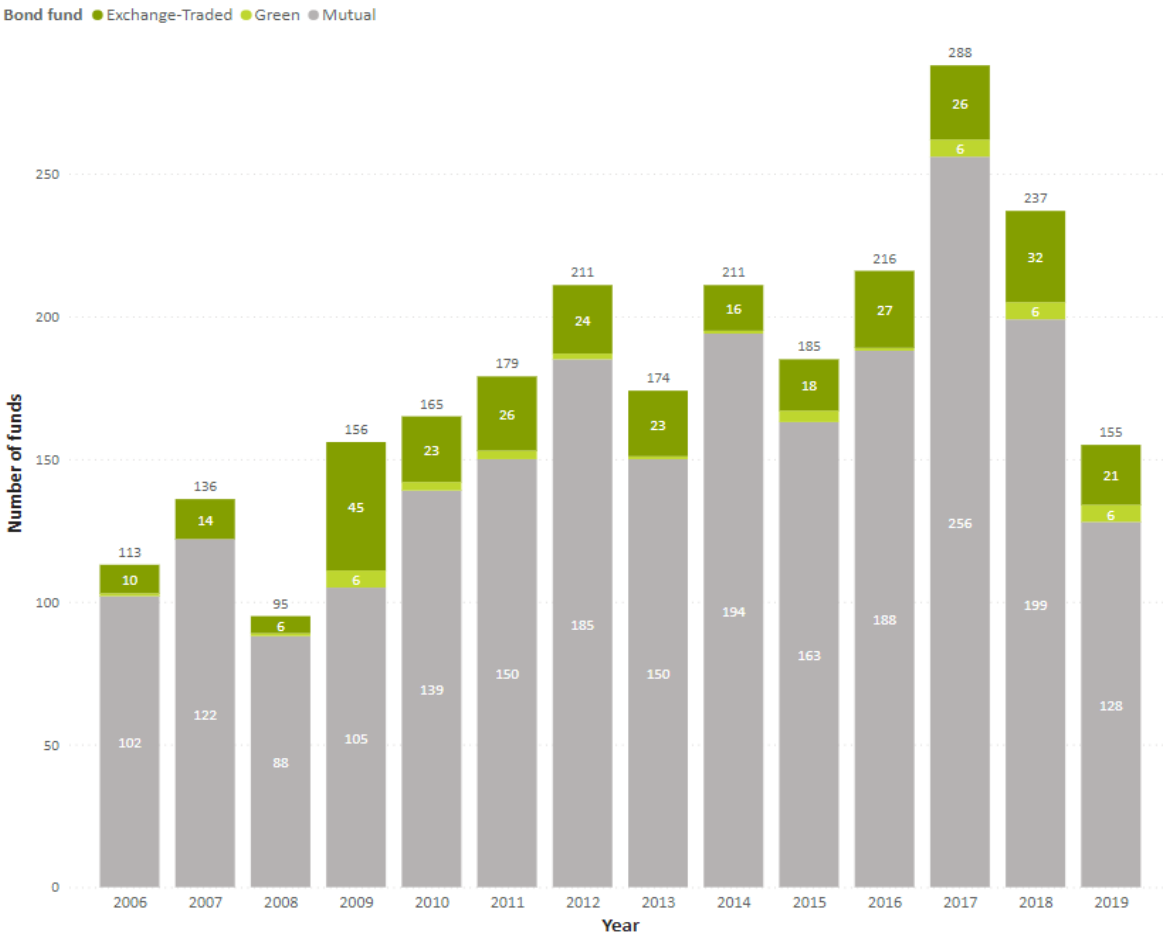


Figure 3.2 - Total number of new funds per year

Since 2005, there have been periods of ups and downs in the number of new bond funds launched each year. For instance, 2008 has the smallest increase in the number of bond fund inceptions during the sample period, accounting for only 95 new bond funds, while the period from 2016 to 2018 reported the largest increase in the number of bond funds incepted each year, from 216 to 288 new bond funds.

Regarding green bond funds, Figure 3.2 also shows that there were no new funds in 2007, so there was no growth that year. In 2009, there was a strong growth in this category of fund, with six new funds launched. In the following years, the growth was less evident, but since 2017 until 2019, the growth rate has been higher again, with six new funds created in each year.

In order to better analyse the differences in performance for each category of fund over the period, the sample was divided into three subperiods. After the international financial crisis in the United States and Great Britain in mid-2007, we observed the proliferation of its effects on other European markets. With the start of the sovereign debt crisis, financial markets experienced a period with a high level of uncertainty that, given the volatility levels, reflects distinct investment conditions than in previous periods.

Therefore, the first subperiod corresponds to the pre-crisis period, from January 2005 to July 2007; the second subperiod corresponds to the US credit crisis and the subsequent European sovereign debt

crisis, from August 2007 to December 2012; and the third subperiod, corresponds to the post-crisis period, after December 2012 until December 2019.

Using equations (1) and (2) from section 3.1., the monthly returns have been calculated, respectively, for each fund and for the benchmark index. Table 3.1 gives a descriptive summary for each category of fund in the three mentioned subperiods. For the purpose of the calculations, in each subsample, funds with less than six months of data were not included.

Table 3.1 - Descriptive statistics of monthly returns of bond mutual funds, bond ETFs, green bond funds and benchmark index

		Pre	During	Post
Average Returns (%)	**Mutual**	0.039	0.315	0.157
	ETF	-0.184	0.350	0.188
	Green	-0.017	0.342	0.015
	Benchmark	0.013	0.551	0.353
1st Quartile (%)	**Mutual**	-0.200	0.113	0.008
	ETF	-0.333	0.156	0.035
	Green	-0.071	0.187	-0.320
	Benchmark	-0.835	-1.370	-0.733
Median (%)	**Mutual**	0.031	0.261	0.134
	ETF	-0.180	0.363	0.183
	Green	0.019	0.302	0.001
	Benchmark	0.295	0.186	0.161
3rd Quartile (%)	**Mutual**	0.189	0.510	0.281
	ETF	0.033	0.526	0.291
	Green	0.072	0.506	0.162
	Benchmark	0.886	1.819	1.304
Minimum (%)	**Mutual**	-2.085	-10.755	-8.007
	ETF	-0.726	-2.228	-1.448
	Green	-0.309	0.054	-0.394
	Benchmark	-2.486	-4.834	-3.185
Maximum (%)	**Mutual**	2.981	3.387	9.472
	ETF	0.125	1.910	1.312
	Green	0.221	0.913	0.931
	Benchmark	3.102	6.475	6.823
Standard Deviation (%)	**Mutual**	1.354	2.373	1.875
	ETF	1.280	2.013	1.556
	Green	1.583	2.202	1.545
	Benchmark	1.298	2.507	1.646

In this table "Pre" represents the pre-crisis period (Jan2005-Jul2007), "During" represents the crisis period (Aug2007-Dec2012) and "Post" represents the post-crisis period (Jan2013-Dec2019).

Considering Table 3.1, in the pre-crisis period bond mutual funds had the highest average monthly return (0.039%), in the crisis and post-crisis periods the highest average monthly returns were given by bond ETFs (0.350% and 0.188% respectively). Although these were the highest monthly average

returns, only in the pre-crisis period did bond mutual funds outperform the benchmark (Bloomberg Euro Aggregate Bond Total Return Index), while in the other two subperiods none of the bond funds outperformed the benchmark return.

It can also be seen that all categories of funds had the lowest average monthly returns from January 2005 to July 2007 and the highest average monthly returns from August 2007 to December 2012. This may be due to an event known as "fly-to-safety". This is a financial term used to describe a phenomenon in which investors move money out of volatile assets and into safer investments during periods of economic uncertainty or market turbulence. This flight to safety is often seen as a sign of market distress and can be triggered by various events such as political instability, natural disasters, or economic downturns. During such events, investors may sell their equities, which are typically seen as riskier assets, and buy safer investments such as government bonds or gold. This flight to safety can cause the prices of these safer assets to rise, while the prices of riskier assets may fall. This is aligned with the literature review conducted (Filip et al., 2015).

In volatility terms and except for the mutual bond funds and green bond funds in the pre-crisis period, and the mutual bond funds in the post-crisis period, all three categories of bond funds report small standard deviation than the benchmark. Bond ETFs had the lowest standard deviation in the pre-crisis and crisis periods (1.280% and 2.013%, respectively), while in the post-crisis period it was green bond funds that had a lower standard deviation (1.545%), making them the least risky funds in each subperiod. In contrast, in the pre-crisis period, green bond funds are the riskiest relative to others (1.583%), and in the crisis and post-crisis periods, bond mutual funds are the riskiest, with a standard deviation of 2.373% and 1.875% respectively.

4. RESULTS AND DISCUSSION

This chapter presents the results of the referred risk-adjusted measures. The comparative analysis of the performance is presented in section 4.1 using the Sharpe ratio; in section 4.2 using the Treynor ratio; and in section 4.3 using the Jensen's Alpha. The aim is to see where can be observed the best performance in terms of these three indicators.

Some notes to keep in mind for the next sections:

- Note that from here on, the word "bond" can be omitted, i.e., "bond mutual funds" can simply be called "mutual funds", "bond ETFs" can be called "ETFs" and "green bond funds" can be called "green funds";
- All calculations assume a 5% statistical significance level.

After obtaining the results of these performance measures, statistical tests were carried out. The F-test is used to compare two variances (Davies et al., 1968) and, according to that, the T-tests are used to compare the means. If the result of the F-test was that the variances were equal, then the two-sample T-test was used (Student, 1908), otherwise if the variances were unequal, then the Welch two-sample T-test was used (Welch, 1947). The results of all tests are displayed in appendices A, B and C.

4.1. SHARPE RATIO

Table 4.1 shows the Sharpe ratio results for bond mutual funds, bond ETFs and green bond funds in each subperiod, calculated using equation (5).

Table 4.1 - Sharpe ratio results

		Pre	During	Post	Total period
Average	**Mutual**	0.294	0.247	0.125	0.191
	ETF	-0.100	0.231	0.116	0.140
	Green	0.001	0.196	0.034	0.078
Positive (%)	**Mutual**	53.8	89.2	77.1	76.5
	ETF	27.6	91.2	81.7	81.6
	Green	56.2	100.0	50.0	66.0
Negative (%)	**Mutual**	46.2	10.8	22.9	23.5
	ETF	72.4	8.8	18.3	18.4
	Green	43.8	0.0	50.0	34.0
Total of funds	**Mutual**	1,043	1,729	3,100	3,100
	ETF	29	160	328	328
	Green	16	31	56	56

In this table "Pre" represents the pre-crisis period (Jan2005-Jul2007), "During" represents the crisis period (Aug2007-Dec2012), "Post" represents the post-crisis period (Jan2013-Dec2019) and "Total period" represents the entire period under study (Jan2005-Dec2012). "Average" represents the monthly average of the Sharpe ratio, "Positive (%)" and "Negative (%)" represent the proportion of funds with a positive or negative Sharpe ratio respectively and "Total of funds" represents the total number of funds considered.

The highest Sharpe ratio for every subperiod is achieved by the mutual funds. As explained in subsection 3.1.1., the higher the Sharpe ratio, the better the performance, so bond mutual funds are considered to be, on average, the best performers, which means that is the fund category that allows investors to earn more return for the same amount of risk taken. In contrast, over the entire sample period, green bond funds are the ones in which investors are less willing to invest, if only looking at returns' perspective.

The Table 4.2 shows the results of the two-sample T-test, comparing the average Sharpe ratios between fund category, for each subperiod.

Table 4.2 - Two-sample T-test for the average Sharpe ratios (by subperiod)

	Pre		**During**		**Post**	
	Mutual	**ETF**	**Mutual**	**ETF**	**Mutual**	**ETF**
ETF	Reject H_0	-	Not reject H_0	-	Not reject H_0	-
Green	Reject H_0	Not reject H_0	Not reject H_0	Not reject H_0	Reject H_0	Reject H_0

This table presents the two-sample T-test for the average Sharpe ratios, where H_0 is the null hypothesis, representing that the differences in the means are zero and the alternative hypothesis, H_1, represents that the differences in the means are different from zero. The decisions were made at the 5% significance level. "Pre" represents the pre-crisis period (Jan2005-Jul2007), "During" represents the crisis period (Aug2007-Dec2012) and "Post" represents the post-crisis period (Jan2013-Dec2019).

From Table 1B in appendix A it is possible to see that in pre-crisis period, the p-values of the T-test for mutual funds are too small relative to the considered significance level. Therefore, there is evidence to conclude that the average Sharpe ratio of mutual funds is statistically different from the average Sharpe ratio of either ETFs or green funds. In other words, there is evidence to conclude that bond mutual funds, on average, outperformed the other fund categories in the pre-crisis period. Regarding the crisis period, the p-values are too high given the 5% of significance level, so there is no evidence of any fund category outperforming the other two. In the post-crisis period, the only conclusion that can be drawn is that green bond funds are statistically different from their non-green peers. Since the monthly average Sharpe ratio of green funds is the lowest in this period, it can be concluded that this fund category has underperformed its non-green peers.

Despite this last conclusion, the slightly lower returns offered by green funds can be offset by the contribution they make to society in terms of environmental benefits. Investing in such assets means investing in companies that care about the environment rather than companies that may have a large carbon footprint.

Furthermore, Table 4.1 also shows that mutual funds reported the higher average Sharpe ratio from January 2005 to July 2007, while ETFs and green funds had the higher average Sharpe ratio from August 2007 to December 2012. Table 4.3 shows the results of the two-sample T-test comparing the average Sharpe ratios between subperiods, for each fund category.

Table 4.3 - Two-sample T-test for the average Sharpe ratios (by fund category)

	Mutual		ETF		Green	
	Pre	**During**	**Pre**	**During**	**Pre**	**During**
During	Not reject H_0	-	Reject H_0	-	Reject H_0	-
Post	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Not reject H_0	Reject H_0

This table presents the two-sample T-test for the average Sharpe ratios, where H_0 is the null hypothesis, representing that the differences in the means are zero and the alternative hypothesis, H_1, represents that the differences in the means are different from zero. The decisions were made at the 5% significance level. "Pre" represents the pre-crisis period (Jan2005-Jul2007), "During" represents the crisis period (Aug2007-Dec2012) and "Post" represents the post-crisis period (Jan2013-Dec2019).

The results in Table 4.3 indicate that in all subperiods, the average Sharpe ratios of bond ETFs are all statistically different at a 95% confidence level. Bond mutual funds perform worse on average after December 2012 compared to the other subperiods, since they give the lowest average Sharpe ratio in this period and the mean is statistically different from the other subperiods. In contrast, green bond funds outperform in the crisis period, in comparison with the other two subperiods.

4.2. TREYNOR RATIO

In order to obtain the results for the Treynor ratio, it is first necessary estimating and understanding the market risk coefficient. This coefficient has been calculated by the single factor model, using equation (6). Table 4.4 shows the estimated results of the OLS parameter Beta for bond mutual funds, bond ETFs and green bond funds in each subperiod.

Table 4.4 - Beta results

		Pre	During	Post
Average	**Mutual**	0.387	0.172	0.314
	ETF	0.826	0.392	0.583
	Green	0.218	0.192	0.332
Positive (%)	**Mutual**	81.6	59.9	75.7
	ETF	96.6	85.0	92.4
	Green	75.0	77.4	87.5
Negative (%)	**Mutual**	18.4	40.1	24.3
	ETF	3.4	15.0	7.6
	Green	25.0	22.6	12.5
Total of funds	**Mutual**	1,043	1,729	3,100
	ETF	29	160	328
	Green	16	31	56

In this table "Pre" represents the pre-crisis period (Jan2005-Jul2007), "During" represents the crisis period (Aug2007-Dec2012) and "Post" represents the post-crisis period (Jan2013-Dec2019). "Average" represents the monthly average of the OLS parameter Beta, "Positive (%)" and "Negative (%)" represent the proportion of funds with a positive or negative OLS parameter Beta respectively and "Total of funds" represents the total number of funds considered.

Table 4.4 exhibits that, on average, all betas are positive and inferior to 1, indicating that all fund categories are less risky than the market in every period. The highest average beta was given by ETFs

in the pre-crisis period (0.826) and the lowest in the crisis period, by mutual funds (0.172), making them the least sensitive to the market fluctuations. The vast majority of funds present a positive beta and, in percentage terms, negative betas are not very representative. However, for Treynor ratio calculation purposes the negative betas cannot be considered, in order to avoid bias in the results.

Table 4.5 shows the Treynor ratio results for bond mutual funds, bond ETFs and green bond funds in each subperiod, calculated using equation (7).

Table 4.5 - Treynor ratio results

		Pre	During	Post	Total period
Average	**Mutual**	0.024	0.085	0.011	0.032
	ETF	-0.003	0.036	0.007	0.015
	Green	0.001	0.040	0.011	0.017
Positive (%)	**Mutual**	49.0	91.2	77.5	75.1
	ETF	28.6	92.6	83.8	83.1
	Green	50.0	100.0	51.0	64.7
Negative (%)	**Mutual**	51.0	8.8	22.5	24.9
	ETF	71.4	7.4	16.2	16.9
	Green	50.0	0.0	49.0	35.3
Total of funds	**Mutual**	851	1,035	2,347	2,347
	ETF	28	136	303	303
	Green	12	24	49	49

In this table "Pre" represents the pre-crisis period (Jan2005-Jul2007), "During" represents the crisis period (Aug2007-Dec2012), "Post" represents the post-crisis period (Jan2013-Dec2019) and "Total period" represents the entire period under study (Jan2005-Dec2012). "Average" represents the monthly average of the Treynor ratio, "Positive (%)" and "Negative (%)" represent the proportion of funds with a positive or negative Treynor ratio respectively and "Total of funds" represents the total number of funds considered.

Table 4.5 shows that, with the exception of ETFs in the pre-crisis period, all three fund categories outperformed the market on average, as they have average positive ratios. In the crisis period, only the sample of green funds' ratios are entirely positive, indicating that none of these funds underperformed the market. As mentioned above, the funds with a negative beta were excluded from the calculation of the Treynor ratio, which reflects an inferior total number of funds than that reported in the other measures.

The highest average Treynor ratio in all subperiods was achieved by bond mutual funds, and also by green funds in the post-crisis period, which is in line with the average ratios of the entire sample period, where the mutual funds had the best performance, followed by green funds and ETFs, i.e., bond mutual funds had the higher return per unit of systematic risk taken.

The Table 4.6 shows the results of the two-sample T-test, comparing the average Treynor ratios between fund category, for each subperiod.

Table 4.6 - Two-sample T-test for the average Treynor ratios (by subperiod)

	Pre		During		Post	
	Mutual	**ETF**	**Mutual**	**ETF**	**Mutual**	**ETF**
ETF	Reject H_0	-	Not reject H_0	-	Not reject H_0	-
Green	Reject H_0	Not reject H_0	Not reject H_0	Not reject H_0	Not reject H_0	Not reject H_0

This table presents the two-sample T-test for the average Treynor ratios, where H_0 is the null hypothesis, representing that the differences in the means are zero and the alternative hypothesis, H_1, represents that the differences in the means are different from zero. The decisions were made at the 5% significance level. "Pre" represents the pre-crisis period (Jan2005-Jul2007), "During" represents the crisis period (Aug2007-Dec2012) and "Post" represents the post-crisis period (Jan2013-Dec2019).

As explained in subsection 3.1.3., the higher the Treynor ratio, the better the performance. Considering Table 1B in appendix B, it can be seen that all the p-values of the T-tests are too high relative to the considered significance level, except for the p-values of the T-test for mutual funds in the pre-crisis period. Therefore, this is the only evidence that exist to support the outperformance of bond mutual funds compared with the other two fund categories.

Furthermore, Table 4.5 also shows that all funds' categories reported the higher average Treynor ratio from August 2007 to December 2012. Table 4.7 shows the results of the two-sample T-test comparing the average Treynor ratios between subperiods, for each fund category.

Table 4.7 - Two-sample T-test for the average Treynor ratios (by fund category)

	Mutual		ETF		Green	
	Pre	**During**	**Pre**	**During**	**Pre**	**During**
During	Reject H_0	-	Reject H_0	-	Reject H_0	-
Post	Not reject H_0	Reject H_0	Reject H_0	Reject H_0	Not reject H_0	Not reject H_0

This table presents the two-sample T-test for the average Treynor ratios, where H_0 is the null hypothesis, representing that the differences in the means are zero and the alternative hypothesis, H_1, represents that the differences in the means are different from zero. The decisions were made at the 5% significance level. "Pre" represents the pre-crisis period (Jan2005-Jul2007), "During" represents the crisis period (Aug2007-Dec2012) and "Post" represents the post-crisis period (Jan2013-Dec2019).

The results in Table 4.7 indicate that in all subperiods, the average Treynor ratios of bond ETFs are all statistically different at a 95% confidence level. Bond mutual funds perform better on average from August 2007 to December 2012 compared to the other subperiods, since they give the highest average Treynor ratio in this period and the mean is statistically different from the other subperiods. The green bond funds (in the same period) also outperformed the pre-crisis period.

4.3. JENSEN'S ALPHA

Table 4.8 shows the Jensen's alpha results for bond mutual funds, bond ETFs and green bond funds in each subperiod, calculated using equation (8).

Table 4.8 - Jensen's alpha results

		Pre	During	Post	Total period
Average	**Mutual**	0.00047	0.00239	0.00047	0.00103
	ETF	-0.00069	0.00166	-0.00021	0.00034
	Green	-0.00026	0.00240	-0.00116	0.00005
Positive (%)	**Mutual**	55.3	78.7	57.5	63.4
	ETF	51.7	73.1	44.2	53.6
	Green	50.0	83.9	32.1	50.5
Pos Significant (%)	**Mutual**	18.8	22.3	10.2	15.3
	ETF	3.4	16.9	5.5	8.9
	Green	0.0	9.7	0.0	2.9
Negative (%)	**Mutual**	44.7	21.3	42.5	36.6
	ETF	48.3	26.9	55.8	46.4
	Green	50.0	16.1	67.9	49.5
Neg Significant (%)	**Mutual**	2.8	3.3	4.2	3.7
	ETF	10.3	3.8	8.5	7.2
	Green	0.0	3.2	16.1	9.7
Total of funds	**Mutual**	1,043	1,729	3,100	3,100
	ETF	29	160	328	328
	Green	16	31	56	56

In this table "Pre" represents the pre-crisis period (Jan2005-Jul2007), "During" represents the crisis period (Aug2007-Dec2012), "Post" represents the post-crisis period (Jan2013-Dec2019) and "Total period" represents the entire period under study (Jan2005-Dec2012). "Average" represents the monthly average of the Jensen's alpha, "Positive (%)" and "Negative (%)" represent the proportion of funds with a positive or negative Jensen's alpha respectively, "Pos Significant (%)" and "Neg Significant (%)" represent the proportion of funds with a positive and significant or negative and significant Jensen's alpha respectively at the 5% significance level and "Total of funds" represents the total number of funds considered.

Table 4.8 exhibits that, mutual funds are the only fund category that did not have any negative average Jensen's alpha in any subperiod. They also had the higher proportion of positive and statistically significant alphas throughout all subperiods. In contrast, were the green funds that gave the lowest proportion of positive and statistically significant alphas.

Both the highest and the lowest alphas were recorded by green funds in the crisis period (0.0024) and in post-crisis period (-0.00116), respectively, representing the best and worst manager performance relative to the market. Although green funds have the highest number of positive alphas (83.9%), this only accounts for 9.7% in terms of significance.

Table 4.8 also demonstrates that the highest average Jensen's alpha in the pre- and post-crisis periods was achieved by mutual funds, and that it was the green funds category that reported the highest average Jensen's alpha in the crisis period. The Table 4.9 shows the results of the two-sample T-test, comparing the average Jensen's alphas between fund category, for each subperiod.

Table 4.9 - Two-sample T-test for the average Jensen's alphas (by subperiod)

	Pre		During		Post	
	Mutual	**ETF**	**Mutual**	**ETF**	**Mutual**	**ETF**
ETF	Reject H_0	-	Not reject H_0	-	Reject H_0	-
Green	Not reject H_0	Not reject H_0	Not reject H_0	Not reject H_0	Reject H_0	Reject H_0

This table presents the two-sample T-test for the average Jensen's alphas, where H_0 is the null hypothesis, representing that the differences in the means are zero and the alternative hypothesis, H_1, represents that the differences in the means are different from zero. The decisions were made at the 5% significance level. "Pre" represents the pre-crisis period (Jan2005-Jul2007), "During" represents the crisis period (Aug2007-Dec2012) and "Post" represents the post-crisis period (Jan2013-Dec2019).

From Table 1B in appendix C it is possible to see that in pre-crisis period, the p-value of the T-test is low relative to the considered significance level. Therefore, there is evidence to conclude that the average Jensen's alpha of mutual funds is higher than the average Jensen's alpha of ETFs. In other words, there is evidence to conclude that the managers of bond mutual funds performed, on average, better than those who managed the ETFs in the pre-crisis period. Regarding the crisis period, the p-values are too high given the 5% of significance level, so there is no evidence of any fund category managers outperforming the managers of the other two fund categories. In the post-crisis period, all fund categories are statistically different from each other, with mutual funds managers outperforming both ETFs and green funds managers, and ETFs management outperforming green funds management.

Furthermore, in Table 4.8 also shows that all categories of funds reported the higher Jensen's alpha from August 2007 to December 2012, i.e., during the crisis period. Table 4.10 shows the results of the two-sample T-test comparing the average Jensen's alphas between subperiods, for each fund category.

Table 4.10 - Two-sample T-test for the average Jensen's alphas (by fund category)

	Mutual		ETF		Green	
	Pre	**During**	**Pre**	**During**	**Pre**	**During**
During	Reject H_0	-	Reject H_0	-	Reject H_0	-
Post	Not reject H_0	Reject H_0	Not reject H_0	Reject H_0	Not reject H_0	Reject H_0

This table presents the two-sample T-test for the average Jensen's alphas, where H_0 is the null hypothesis, representing that the differences in the means are zero and the alternative hypothesis, H_1, represents that the differences in the means are different from zero. The decisions were made at the 5% significance level. "Pre" represents the pre-crisis period (Jan2005-Jul2007), "During" represents the crisis period (Aug2007-Dec2012) and "Post" represents the post-crisis period (Jan2013-Dec2019).

The results in Table 4.10 indicate that, at a 95% confidence level, all categories of bond funds have statistically different means between the crisis period and the other two subperiods and, given that Jensen's alphas are higher in the crisis period, there is evidence to suggest that bond funds were better managed in this period than in the other subperiods.

5. CONCLUSIONS AND FURTHER DEVELOPMENTS

There has been an increase in ESG investment in recent years, as both investors and financial institutions have considered the importance of sustainability in their investments. Nevertheless, ESG investing is still a work in progress. As stated above, the main driver for managers and investors to consider sustainable strategies is the expectation of higher long-term returns.

From the ESG principles, the environmental factor is the one that reflects the impact of a company in the environment and its efforts to reduce its carbon footprint. The green bond funds are inserted into this factor, and it is important to understand if this category of funds present a higher return than their conventional counterparties. Therefore, this dissertation provides a comparative performance analysis between green bond funds, bond mutual funds and bond ETFs.

To this end, monthly NAVs of 3,484 funds were extracted from January 2005 to December 2019. The data consists in 3,100 bond mutual funds, 328 bond ETFs and 56 green bond funds, all domiciled in Europe and previously converted for Euro currency. For analysis purposes, the data was divided in three distinct subperiods: the pre-crisis period (January 2005 - July 2007), crisis period (August 2007 - December 2012) and the post-crisis period (January 2013 - December 2019).

The study started by computing the monthly returns of all funds and conducting a preliminary statistic to assess, among others, the average returns as well as the level of risk of each, i.e., the standard deviation of each category of fund across the three subperiods. It was during the crisis that the higher volatility was observed, but also where the higher average returns were achieved, throughout all fund categories.

In order to assess the performance of each fund category, through all the mentioned subperiods, some risk-adjusted measures have been applied for each fund, such as the Sharpe ratio, the Treynor ratio and the Jensen's alpha, along with the F-tests to compare two variances and two-sample T-tests for the means.

The findings reveal that in the pre-crisis period, bond mutual funds had the highest average performances for all the calculated risk-adjusted measures, compared to the other two fund categories. The obtained T-tests showed that bond mutual funds average returns were statistically significant for both Sharpe and Treynor ratios at the 95% confidence level when compared with the other two fund categories. This is only the case for Jensen's alpha when compared to bond ETFs. In the crisis period, although bond mutual funds had the higher average Sharpe and Treynor ratios, and green bond funds had the higher average Jensen's alpha, there is no statistical evidence that these bond funds outperformed the other fund categories in this period. Finally, in the post-crisis period, there is evidence to conclude that green bond funds underperformed the other two categories, for both Sharpe ratio and Jensen's alpha. In addition, in terms of Jensen's alpha, bond mutual funds outperformed bond ETFs. About the Treynor ratio, nothing could be concluded, since the none of the T-tests reveal statistical differences between the means.

Analysing each fund category separately over the different subperiods, the following conclusions can be drawn. All fund categories have a consistent behaviour over time, evidencing its best average performance in the crisis period.

Overall, bond mutual funds are considered to be the best performers, followed by green bond funds and bond ETFs with a similar performance. As mentioned above, all fund categories had performed better in the crisis period, suggesting that investors seek for less risky investments in times of high volatility and, therefore, opt for safer investments. This can be considered a fly-to-safety event, where the prices of safer assets rise, and riskier assets' prices have a downward tendency.

Although green bond funds outperformed in some moments, the small number of funds in the sample may have led to the conclusions not being statistically significant. Additionally, investors seek a higher return with the lowest possible risk, and there is a wide range of investment options. Although green bond funds offer lower returns than other types of funds, some environmentally conscious investors may prefer to invest in these assets due to the contribution that these funds provide to the environment.

As mentioned in chapter 1, the main objective of this study was to understand whether green bond funds would outperform their peers, either bond mutual funds or bond ETFs. However, one of the main limitations was obtaining data. The best tools require an annual or monthly subscription and their interfaces are not user-friendly for obtaining information, as is the case with Bloomberg's terminal.

Additionally, by selecting only the subset of the population that "survived" or persisted over the entire sample period, one excludes the subset of funds that were discontinued due to poor performance, and this can cause a survivorship bias. Ignoring this poor performance of some funds can result into a misleading understanding of historical returns and an inaccurate representation of portfolio risk and return measures. To address this, the full subset of the population should be considered, i.e., including those that persisted over time, but also those that ceased to exist at some point during the sample period.

For the purposes of this study, only the net values were analysed, i.e., after deducting commissions. All the analysis was carried out from the investor's point of view, but it might be interesting to do the analysis from the manager's perspective, taking into account the gross values, i.e., before deducting commissions. Another interesting future analysis would be to look at a different geographical area where the funds are domiciled. This research focuses only on funds domiciled in Europe, but looking at funds domiciled in other regions, or even without any restriction on domicile, could lead to different results. And another possible approach in this study would be to model volatility using ARCH, GARCH or EGARCH methods instead of calculating historical standard deviation.

As a final suggestion, it would be pertinent to consider other time periods. This research has only focused on the 2007-2008 financial crisis, but after the Covid-19 epidemic crisis and the war in Ukraine, which caused an exceptional increase in the inflation rate, it will also be important to understand how this category of funds behaves nowadays.

BIBLIOGRAPHY

About Euribor® | The European Money Markets Institute (EMMI). (2023). emmi-Benchmarks.eu. https://www.emmi-benchmarks.eu/benchmarks/euribor/

Agliardi, E., & Chechulin, V. (2020). Green Bonds vs Regular Bonds: Debt Level and Corporate Performance. Journal of Corporate Finance Research / Корпоративные Финансы | ISSN: 2073-0438, 14(2), 83–99. https://doi.org/10.17323/j.jcfr.2073-0438.14.2.2020.83-99

Bacon, C. R. (2008). Practical portfolio performance measurement and attribution. Chichester Wiley.

Bernardo, A. E., & Ledoit, O. (2000). Gain, Loss, and Asset Pricing. Journal of Political Economy, 108(1), 144–172. https://doi.org/10.1086/262114

Biglova, A., Ortobelli, S., Rachev, S. T., & Stoyanov, S. (2004). Different Approaches to Risk Estimation in Portfolio Theory. The Journal of Portfolio Management, 31(1), 103–112. https://doi.org/10.3905/jpm.2004.443328

Black, F. (1972). Capital Market Equilibrium with Restricted Borrowing. The Journal of Business, 45(3), 444. https://doi.org/10.1086/295472

Bodie, Z., Kane, A., & Marcus, A. J. (2014). Essentials of investments (10th ed.). Mcgraw-Hill Education.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327.

Chang, C. E., Nelson, W. A., & Doug Witte, H. (2012). Do green mutual funds perform well? Management Research Review, 35(8), 693–708. https://doi.org/10.1108/01409171211247695

Choudhry, M. (2010). An introduction to bond markets. Wiley.

Connor, G., & Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory. Journal of Financial Economics, 15(3), 373–394. https://doi.org/10.1016/0304-405x(86)90027-9

Davies, O. L., Snedecor, G. W., & Cochran, W. G. (1968). Statistical Methods. 6th Edition. Applied Statistics, 17(3). https://doi.org/10.2307/2985653

Deribew, H.N. (2017). Risks and Returns of Green Bonds. Master's Thesis.

Dixon-Fowler, H., Slater, D., Johnson, J., Ellstrand, A., & Romi, A. (2013). Beyond "'Does it Pay to be Green?'" A Meta-Analysis of Moderators of the CEP-CFP Relationship. *Journal of business ethics*, *112*(2), 353-366.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50(4), 987–1007. https://doi.org/10.2307/1912773

ESG funds that reflect what matters most to you | Vanguard. (2022). Investor.vanguard.com. https://investor.vanguard.com/investment-products/esg

Fama, E. F. (1972). Components of Investment Performance. The Journal of Finance, 27(3), 551–567. https://doi.org/10.1111/j.1540-6261.1972.tb00984.x

Filip, A.-M., Pece, A., & Lăcătuş, V. (2015). Risk-adjusted Performance of Romanian Bond Funds during the Global Economic Crisis. Procedia Economics and Finance, 32, 1407–1413. https://doi.org/10.1016/s2212-5671(15)01517-8

Gemmill, G., Hwang, S., & Salmon, M. (2006). Performance measurement with loss aversion. Journal of Asset Management, 7(3-4), 190–207. https://doi.org/10.1057/palgrave.jam.2240213

Greenhouse Gas Bulletin: Another Year Another Record | World Meteorological Organization. (2021, October 25). https://public.wmo.int/en/media/press-release/greenhouse-gas-bulletin-another-year-another-record

Hull, J. (2018). Risk management and financial institutions (5th ed.). John Wiley & Sons, Inc.

Inderst, G., & Stewart, F. (2018). Incorporating Environmental, Social and Governance (ESG) Factors into Fixed Income Investment. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3175830

Israëlsen, C. L. (2005). A refinement to the Sharpe ratio and information ratio. Journal of Asset Management, 5(6), 423–427. https://doi.org/10.1057/palgrave.jam.2240158

Jaksic, M., Lekovic, M., & Milanovic, M. (2015). Measuring the performance of mutual funds: A case study. Industrija, 43(1), 37–51. https://doi.org/10.5937/industrija43-6677

Jensen, M. C. (1968). The Performance of Mutual Funds in the Period 1945-1964. Journal of Finance, 23(2). https://doi.org/10.2139/ssrn.244153

Keating, C., & Shadwick, W. F. (2002). A Universal Performance Measure. Journal of Performance Measurement, 6, 59–84.

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

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica, 59(2), 347. https://doi.org/10.2307/2938260

Németh-Durkó, E., & Hegedűs, A. (2021). Climate Change in the Capital Markets: A Study of Actively Managed Green Bond Funds. Financial and Economic Review, 20(4), 38–64. https://doi.org/10.33893/fer.20.4.3864

Nizam, E., Ng, A., Dewandaru, G., Nagayev, R., & Nkoba, M. A. (2019). The impact of social and environmental sustainability on financial performance: A global analysis of the banking sector. Journal of Multinational Financial Management, 49, 35–53. https://doi.org/10.1016/j.mulfin.2019.01.002

Paradinovic, I. (2017). Comparison of the performance of Islamic, Sri and green mutual funds. Run.unl.pt. https://run.unl.pt/handle/10362/25469

Pézier, J., & White, A. (2006). The Relative Merits of Investable Hedge Fund Indices and of Funds of Hedge Funds in Optimal Passive Portfolios. https://core.ac.uk/download/pdf/6560807.pdf

Samarbakhsh, L., & Shah, M. (2021). Fixed income mutual fund performance during and after a crisis: a Canadian case. Journal of Economics and Finance. https://doi.org/10.1007/s12197-021-09541-z

Sampei, H. (2018, September 12). ESG awareness is an enduring legacy of the global financial crisis. Fidelity International. https://www.fidelityinternational.com/editorial/blog/pesg-awareness-is-an-enduring-legacy-of-the-global-financial-crisisp-a5a9f2-en5/

Sharpe, W. F. (1963). A Simplified Model for Portfolio Analysis. Management Science, 9(2), 277–293. https://doi.org/10.1287/mnsc.9.2.277

Sharpe, W. F. (1966). Mutual Fund Performance. The Journal of Business, 39(1), 119–138. https://doi.org/10.1086/294846

Sharpe, W. F. (1994). The Sharpe Ratio. The Journal of Portfolio Management, 21(1), 49–58. https://doi.org/10.3905/jpm.1994.409501

Sortino, F., van der Meer, R. A. H., & Plantinga, A. (1999). The Dutch triangle - A framework to measure upside potential relative to downside risk. Journal of Portfolio Management, 26(1), 50–58.

Srivastava, S. C., & Essayyad, M. (1994). Investigating a new methodology for ranking international mutual funds. Journal of Economics and Finance, 18(3), 241–260. https://doi.org/10.1007/bf02920485

Stankevičienė, J., & Petronienė, I. (2019). Bond Mutual Funds vs. Bond Exchange Traded Funds: Evaluation of Risk Adjusted Performance. Administrative Sciences, 9(2), 31. https://doi.org/10.3390/admsci9020031

Student. (1908). The Probable Error of a Mean. Biometrika, 6(1), 1. https://doi.org/10.2307/2331554

Treynor, J. L. (1965). How to rate management of investment funds. Harvard business review, 43(1), 63-75.

Vaidya, D. (2017). Risk Adjusted Return | Top 6 Risk Ratios You must Know! Wallstreet Mojo. https://www.wallstreetmojo.com/risk-adjusted-returns/

Welch, B. L. (1947). The Generalization of `Student's' Problem when Several Different Population Variances are Involved. Biometrika, 34(1/2), 28. https://doi.org/10.2307/2332510

APPENDICES

A. Sharpe ratio statistical tests results

Table 1A. F-test to compare two variances for the average Sharpe ratios (by subperiod)

H_0	SD of x	SD of y	Df of x	Df of y	F-Statistic	P-value	Method	Decision
$\sigma_{MPr} = \sigma_{EPr}$	0.982	0.194	1,042	28	25.60	3.6E-15	F test	Reject H_0
$\sigma_{MPr} = \sigma_{GPr}$	0.982	0.136	1,042	15	52.01	6.5E-11	F test	Reject H_0
$\sigma_{EPr} = \sigma_{GPr}$	0.194	0.136	28	15	2.03	0.151	F test	Not reject H_0
$\sigma_{MD} = \sigma_{ED}$	0.479	0.238	1,728	159	4.05	0.0E+00	F test	Reject H_0
$\sigma_{MD} = \sigma_{GD}$	0.479	0.199	1,728	30	5.82	2.2E-07	F test	Reject H_0
$\sigma_{ED} = \sigma_{GD}$	0.238	0.199	159	30	1.44	0.242	F test	Not reject H_0
$\sigma_{MPt} = \sigma_{EPt}$	0.203	0.210	3,099	327	0.93	0.376	F test	Not reject H_0
$\sigma_{MPt} = \sigma_{GPt}$	0.203	0.208	3,099	55	0.95	0.750	F test	Not reject H_0
$\sigma_{EPt} = \sigma_{GPt}$	0.210	0.208	327	55	1.02	0.957	F test	Not reject H_0

This table presents the F-test to compare two variances for the average Sharpe ratios, where H_0 is the null hypothesis, representing that the differences in the variances are zero and the alternative hypothesis, H_1, represents that the differences in the variances are different from zero. The "decision" was made at the 5% significance level. "σ" represents the population variance. "M", "E" and "G" represents the bond mutual, bond ETF and green bond funds, respectively. "Pr" represents the pre-crisis period (Jan2005-Jul2007), "D" represents the crisis period (Aug2007-Dec2012) and "Pt" represents the post-crisis period (Jan2013-Dec2019). "SD of x" and "SD of y" represents the standard deviation of the variables and "Df of x" and "Df of y" represents the degrees of freedom of the variables.

Table 1B. Two-sample T-test for the average Sharpe ratios (by subperiod)

H_0	Mean of x	Mean of y	T-Statistic	P-value	Df	Method	Decision
$\mu_{MPr} = \mu_{EPr}$	0.294	-0.100	8.36	1.5E-12	80.9	T test (Unequal var)	Reject H_0
$\mu_{MPr} = \mu_{GPr}$	0.294	0.001	6.42	5.6E-08	48.0	T test (Unequal var)	Reject H_0
$\mu_{EPr} = \mu_{GPr}$	-0.100	0.001	-1.84	0.073	43.0	T test (Equal var)	Not reject H_0
$\mu_{MD} = \mu_{ED}$	0.247	0.231	0.71	0.478	296.6	T test (Unequal var)	Not reject H_0
$\mu_{MD} = \mu_{GD}$	0.247	0.196	1.35	0.184	36.6	T test (Unequal var)	Not reject H_0
$\mu_{ED} = \mu_{GD}$	0.231	0.196	0.77	0.443	189.0	T test (Equal var)	Not reject H_0
$\mu_{MPt} = \mu_{EPt}$	0.125	0.116	0.69	0.488	3,426.0	T test (Equal var)	Not reject H_0
$\mu_{MPt} = \mu_{GPt}$	0.125	0.034	3.29	0.001	3,154.0	T test (Equal var)	Reject H_0
$\mu_{EPt} = \mu_{GPt}$	0.116	0.034	2.70	0.007	382.0	T test (Equal var)	Reject H_0

This table presents the two-sample T-test for the average Sharpe ratios, where H_0 is the null hypothesis, representing that the differences in the means are zero and the alternative hypothesis, H_1, represents that the differences in the means are different from zero. The "decision" was made at the 5% significance level. "μ" represents the population mean. "M", "E" and "G" represents the bond mutual, bond ETF and green bond funds, respectively. "Pr" represents the pre-crisis period (Jan2005-Jul2007), "D" represents the crisis period (Aug2007-Dec2012) and "Pt" represents the post-crisis period (Jan2013-Dec2019). "Mean of x" and "Mean of y" represents the mean of the variables and "Df" represents the degrees of freedom.

Table 2A. F-test to compare two variances for the average Sharpe ratios (by fund category)

H_0	SD of x	SD of y	Df of x	Df of y	F-Statistic	P-value	Method	Decision
$\sigma_{MPr} = \sigma_{MD}$	0.982	0.479	1,042	1,728	4.20	0.0E+00	F test	Reject H_0
$\sigma_{MPr} = \sigma_{MPt}$	0.982	0.203	1,042	3,099	23.43	0.0E+00	F test	Reject H_0
$\sigma_{MD} = \sigma_{MPt}$	0.479	0.203	1,728	3,099	5.58	0.0E+00	F test	Reject H_0
$\sigma_{EPr} = \sigma_{ED}$	0.194	0.238	28	159	0.66	0.203	F test	Not reject H_0
$\sigma_{EPr} = \sigma_{EPt}$	0.194	0.210	28	327	0.85	0.633	F test	Not reject H_0
$\sigma_{ED} = \sigma_{EPt}$	0.238	0.210	159	327	1.28	0.061	F test	Not reject H_0
$\sigma_{GPr} = \sigma_{GD}$	0.136	0.199	15	30	0.47	0.124	F test	Not reject H_0
$\sigma_{GPr} = \sigma_{GPt}$	0.136	0.208	15	55	0.43	0.073	F test	Not reject H_0
$\sigma_{GD} = \sigma_{GPt}$	0.199	0.208	30	55	0.91	0.802	F test	Not reject H_0

This table presents the F-test to compare two variances for the average Sharpe ratios, where H_0 is the null hypothesis, representing that the differences in the variances are zero and the alternative hypothesis, H_1, represents that the differences in the variances are different from zero. The "decision" was made at the 5% significance level. "σ" represents the population variance. "M", "E" and "G" represents the bond mutual, bond ETF and green bond funds, respectively. "Pr" represents the pre-crisis period (Jan2005-Jul2007), "D" represents the crisis period (Aug2007-Dec2012) and "Pt" represents the post-crisis period (Jan2013-Dec2019). "SD of x" and "SD of y" represents the standard deviation of the variables and "Df of x" and "Df of y" represents the degrees of freedom of the variables.

Table 2B. Two-sample T-test for the average Sharpe ratios (by fund category)

H_0	Mean of x	Mean of y	T-Statistic	P-value	Df	Method	Decision
$\mu_{MPr} = \mu_{MD}$	0.294	0.247	1.44	0.149	1,346.1	T test (Unequal var)	Not reject H_0
$\mu_{MPr} = \mu_{MPt}$	0.294	0.125	5.53	4.1E-08	1,072.1	T test (Unequal var)	Reject H_0
$\mu_{MD} = \mu_{MPt}$	0.247	0.125	10.12	1.6E-23	2,079.2	T test (Unequal var)	Reject H_0
$\mu_{EPr} = \mu_{ED}$	-0.100	0.231	-7.07	2.9E-11	187.0	T test (Equal var)	Reject H_0
$\mu_{EPr} = \mu_{EPt}$	-0.100	0.116	-5.35	1.6E-07	355.0	T test (Equal var)	Reject H_0
$\mu_{ED} = \mu_{EPt}$	0.231	0.116	5.42	9.4E-08	486.0	T test (Equal var)	Reject H_0
$\mu_{GPr} = \mu_{GD}$	0.001	0.196	-3.52	0.001	45.0	T test (Equal var)	Reject H_0
$\mu_{GPr} = \mu_{GPt}$	0.001	0.034	-0.61	0.542	70.0	T test (Equal var)	Not reject H_0
$\mu_{GD} = \mu_{GPt}$	0.196	0.034	3.53	0.001	85.0	T test (Equal var)	Reject H_0

This table presents the two-sample T-test for the average Sharpe ratios, where H_0 is the null hypothesis, representing that the differences in the means are zero and the alternative hypothesis, H_1, represents that the differences in the means are different from zero. The "decision" was made at the 5% significance level. "μ" represents the population mean. "M", "E" and "G" represents the bond mutual, bond ETF and green bond funds, respectively. "Pr" represents the pre-crisis period (Jan2005-Jul2007), "D" represents the crisis period (Aug2007-Dec2012) and "Pt" represents the post-crisis period (Jan2013-Dec2019). "Mean of x" and "Mean of y" represents the mean of the variables and "Df" represents the degrees of freedom.

B. Treynor ratio statistical tests results

Table 1A. F-test to compare two variances for the average Treynor ratios (by subperiod)

H_0	SD of x	SD of y	Df of x	Df of y	F-Statistic	P-value	Method	Decision
$\sigma_{MPr} = \sigma_{EPr}$	0.175	0.010	850	27	297.36	0.0E+00	F test	Reject H_0
$\sigma_{MPr} = \sigma_{GPr}$	0.175	0.008	850	11	507.40	1.1E-13	F test	Reject H_0
$\sigma_{EPr} = \sigma_{GPr}$	0.010	0.008	27	11	1.71	0.353	F test	Not reject H_0
$\sigma_{MD} = \sigma_{ED}$	0.862	0.101	1,034	135	73.02	0.0E+00	F test	Reject H0H_0
$\sigma_{MD} = \sigma_{GD}$	0.862	0.083	1,034	23	107.44	0.0E+00	F test	Reject H_0
$\sigma_{ED} = \sigma_{GD}$	0.101	0.083	135	23	1.47	0.282	F test	Not reject H_0
$\sigma_{MPt} = \sigma_{EPt}$	0.268	0.058	2,346	302	21.45	0.0E+00	F test	Reject H_0
$\sigma_{MPt} = \sigma_{GPt}$	0.268	0.075	2,346	48	12.76	0.0E+00	F test	Reject H_0
$\sigma_{EPt} = \sigma_{GPt}$	0.058	0.075	302	48	0.59	0.010	F test	Reject H_0

This table presents the F-test to compare two variances for the average Treynor ratios, where H_0 is the null hypothesis, representing that the differences in the variances are zero and the alternative hypothesis, H_1, represents that the differences in the variances are different from zero. The "decision" was made at the 5% significance level. "σ" represents the population variance. "M", "E" and "G" represents the bond mutual, bond ETF and green bond funds, respectively. "Pr" represents the pre-crisis period (Jan2005-Jul2007), "D" represents the crisis period (Aug2007-Dec2012) and "Pt" represents the post-crisis period (Jan2013-Dec2019). "SD of x" and "SD of y" represents the standard deviation of the variables and "Df of x" and "Df of y" represents the degrees of freedom of the variables.

Table 1B. Two-sample T-test for the average Treynor ratios (by subperiod)

H_0	Mean of x	Mean of y	T-Statistic	P-value	Df	Method	Decision
$\mu_{MPr} = \mu_{EPr}$	0.024	-0.003	4.33	1.7E-05	777.1	T test (Unequal var)	Reject H_0
$\mu_{MPr} = \mu_{GPr}$	0.024	0.001	3.59	3.7E-04	440.0	T test (Unequal var)	Reject H_0
$\mu_{EPr} = \mu_{GPr}$	-0.003	0.001	-1.31	0.197	38.0	T test (Equal var)	Not reject H_0
$\mu_{MD} = \mu_{ED}$	0.085	0.036	1.74	0.083	1,163.9	T test (Unequal var)	Not reject H_0
$\mu_{MD} = \mu_{GD}$	0.085	0.040	1.43	0.153	246.3	T test (Unequal var)	Not reject H_0
$\mu_{ED} = \mu_{GD}$	0.036	0.040	-0.16	0.877	158.0	T test (Equal var)	Not reject H_0
$\mu_{MPt} = \mu_{EPt}$	0.011	0.007	0.52	0.600	2,159.1	T test (Unequal var)	Not reject H_0
$\mu_{MPt} = \mu_{GPt}$	0.011	0.011	0.00	0.998	76.9	T test (Unequal var)	Not reject H_0
$\mu_{EPt} = \mu_{GPt}$	0.007	0.011	-0.30	0.762	57.6	T test (Unequal var)	Not reject H_0

This table presents the two-sample T-test for the average Treynor ratios, where H_0 is the null hypothesis, representing that the differences in the means are zero and the alternative hypothesis, H_1, represents that the differences in the means are different from zero. The "decision" was made at the 5% significance level. "μ" represents the population mean. "M", "E" and "G" represents the bond mutual, bond ETF and green bond funds, respectively. "Pr" represents the pre-crisis period (Jan2005-Jul2007), "D" represents the crisis period (Aug2007-Dec2012) and "Pt" represents the post-crisis period (Jan2013-Dec2019). "Mean of x" and "Mean of y" represents the mean of the variables and "Df" represents the degrees of freedom.

Table 2A. F-test to compare two variances for the average Treynor ratios (by fund category)

H_0	SD of x	SD of y	Df of x	Df of y	F-Statistic	P-value	Method	Decision
$\sigma_{MPr} = \sigma_{MD}$	0.175	0.862	850	1,034	0.04	0.0E+00	F test	Reject H_0
$\sigma_{MPr} = \sigma_{MPt}$	0.175	0.268	850	2,346	0.43	4.7E-44	F test	Reject H_0
$\sigma_{MD} = \sigma_{MPt}$	0.862	0.268	1,034	2,346	10.37	0.0E+00	F test	Reject H_0
$\sigma_{EPr} = \sigma_{ED}$	0.010	0.101	27	135	0.01	5.2E-22	F test	Reject H_0
$\sigma_{EPr} = \sigma_{EPt}$	0.010	0.058	27	302	0.03	7.2E-16	F test	Reject H_0
$\sigma_{ED} = \sigma_{EPt}$	0.101	0.058	135	302	3.05	1.3E-15	F test	Reject H_0
$\sigma_{GPr} = \sigma_{GD}$	0.008	0.083	11	23	0.01	9.4E-10	F test	Reject H_0
$\sigma_{GPr} = \sigma_{GPt}$	0.008	0.075	11	48	0.01	1.9E-09	F test	Reject H_0
$\sigma_{GD} = \sigma_{GPt}$	0.083	0.075	23	48	1.23	0.532	F test	Not reject H_0

This table presents the F-test to compare two variances for the average Treynor ratios, where H_0 is the null hypothesis, representing that the differences in the variances are zero and the alternative hypothesis, H_1, represents that the differences in the variances are different from zero. The "decision" was made at the 5% significance level. "σ" represents the population variance. "M", "E" and "G" represents the bond mutual, bond ETF and green bond funds, respectively. "Pr" represents the pre-crisis period (Jan2005-Jul2007), "D" represents the crisis period (Aug2007-Dec2012) and "Pt" represents the post-crisis period (Jan2013-Dec2019). "SD of x" and "SD of y" represents the standard deviation of the variables and "Df of x" and "Df of y" represents the degrees of freedom of the variables.

Table 2B. Two-sample T-test for the average Treynor ratios (by fund category)

H_0	Mean of x	Mean of y	T-Statistic	P-value	Df	Method	Decision
$\mu_{MPr} = \mu_{MD}$	0.024	0.085	-2.21	0.027	1,136.8	T test (Unequal var)	Reject H_0
$\mu_{MPr} = \mu_{MPt}$	0.024	0.011	1.71	0.087	2,302.2	T test (Unequal var)	Not reject H_0
$\mu_{MD} = \mu_{MPt}$	0.085	0.011	2.73	0.006	1,122.9	T test (Unequal var)	Reject H_0
$\mu_{EPr} = \mu_{ED}$	-0.003	0.036	-4.42	1.9E-05	146.8	T test (Unequal var)	Reject H_0
$\mu_{EPr} = \mu_{EPt}$	-0.003	0.007	-2.60	0.010	239.1	T test (Unequal var)	Reject H_0
$\mu_{ED} = \mu_{EPt}$	0.036	0.007	3.15	0.002	176.0	T test (Unequal var)	Reject H_0
$\mu_{GPr} = \mu_{GD}$	0.001	0.040	-2.23	0.035	23.8	T test (Unequal var)	Reject H_0
$\mu_{GPr} = \mu_{GPt}$	0.001	0.011	-0.83	0.412	51.9	T test (Unequal var)	Not reject H_0
$\mu_{GD} = \mu_{GPt}$	0.040	0.011	1.51	0.136	71.0	T test (Equal var)	Not reject H_0

This table presents the two-sample T-test for the average Treynor ratios, where H_0 is the null hypothesis, representing that the differences in the means are zero and the alternative hypothesis, H_1, represents that the differences in the means are different from zero. The "decision" was made at the 5% significance level. "μ" represents the population mean. "M", "E" and "G" represents the bond mutual, bond ETF and green bond funds, respectively. "Pr" represents the pre-crisis period (Jan2005-Jul2007), "D" represents the crisis period (Aug2007-Dec2012) and "Pt" represents the post-crisis period (Jan2013-Dec2019). "Mean of x" and "Mean of y" represents the mean of the variables and "Df" represents the degrees of freedom.

C. Jensen's alpha statistical tests results

Table 1A. F-test to compare two variances for the average Jensen's alphas (by subperiod)

H_0	SD of x	SD of y	Df of x	Df of y	F-Statistic	P-value	Method	Decision
$\sigma_{MPr} = \sigma_{EPr}$	0.0038	0.0025	1,042	28	2.29	0.009	F test	Reject H_0
$\sigma_{MPr} = \sigma_{GPr}$	0.0038	0.0016	1,042	15	5.86	3.1E-04	F test	Reject H_0
$\sigma_{EPr} = \sigma_{GPr}$	0.0025	0.0016	28	15	2.56	0.059	F test	Not reject H_0
$\sigma_{MD} = \sigma_{ED}$	0.0050	0.0045	1,728	159	1.20	0.139	F test	Not reject H_0
$\sigma_{MD} = \sigma_{GD}$	0.0050	0.0028	1,728	30	3.05	4.0E-04	F test	Reject H_0
$\sigma_{ED} = \sigma_{GD}$	0.0045	0.0028	159	30	2.54	0.004	F test	Reject H_0
$\sigma_{MPt} = \sigma_{EPt}$	0.0044	0.0024	3,099	327	3.37	0.0E+00	F test	Reject H_0
$\sigma_{MPt} = \sigma_{GPt}$	0.0044	0.0030	3,099	55	2.16	4.6E-04	F test	Reject H_0
$\sigma_{EPt} = \sigma_{GPt}$	0.0024	0.0030	327	55	0.64	0.021	F test	Reject H_0

This table presents the F-test to compare two variances for the average Jensen's alphas, where H_0 is the null hypothesis, representing that the differences in the variances are zero and the alternative hypothesis, H_1, represents that the differences in the variances are different from zero. The "decision" was made at the 5% significance level. "σ" represents the population variance. "M", "E" and "G" represents the bond mutual, bond ETF and green bond funds, respectively. "Pr" represents the pre-crisis period (Jan2005-Jul2007), "D" represents the crisis period (Aug2007-Dec2012) and "Pt" represents the post-crisis period (Jan2013-Dec2019). "SD of x" and "SD of y" represents the standard deviation of the variables and "Df of x" and "Df of y" represents the degrees of freedom of the variables.

Table 1B. Two-sample T-test for the average Jensen's alphas (by subperiod)

H_0	Mean of x	Mean of y	T-Statistic	P-value	Df	Method	Decision
$\mu_{MPr} = \mu_{EPr}$	0.0005	-0.0007	2.41	0.022	31.7	T test (Unequal var)	Reject H_0
$\mu_{MPr} = \mu_{GPr}$	0.0005	-0.0003	1.79	0.091	17.8	T test (Unequal var)	Not reject H_0
$\mu_{EPr} = \mu_{GPr}$	-0.0007	-0.0003	-0.62	0.539	43.0	T test (Equal var)	Not reject H_0
$\mu_{MD} = \mu_{ED}$	0.0024	0.0017	1.77	0.077	1,887.0	T test (Equal var)	Not reject H_0
$\mu_{MD} = \mu_{GD}$	0.0024	0.0024	-0.02	0.981	33.4	T test (Unequal var)	Not reject H_0
$\mu_{ED} = \mu_{GD}$	0.0017	0.0024	-1.18	0.244	63.9	T test (Unequal var)	Not reject H_0
$\mu_{MPt} = \mu_{EPt}$	0.0005	-0.0002	4.46	9.6E-06	593.7	T test (Unequal var)	Reject H_0
$\mu_{MPt} = \mu_{GPt}$	0.0005	-0.0012	4.02	1.7E-04	59.4	T test (Unequal var)	Reject H_0
$\mu_{EPt} = \mu_{GPt}$	-0.0002	-0.0012	2.26	0.027	67.6	T test (Unequal var)	Reject H_0

This table presents the two-sample T-test for the average Jensen's alphas, where H_0 is the null hypothesis, representing that the differences in the means are zero and the alternative hypothesis, H_1, represents that the differences in the means are different from zero. The "decision" was made at the 5% significance level. "μ" represents the population mean. "M", "E" and "G" represents the bond mutual, bond ETF and green bond funds, respectively. "Pr" represents the pre-crisis period (Jan2005-Jul2007), "D" represents the crisis period (Aug2007-Dec2012) and "Pt" represents the post-crisis period (Jan2013-Dec2019). "Mean of x" and "Mean of y" represents the mean of the variables and "Df" represents the degrees of freedom.

Table 2A. F-test to compare two variances for the average Jensen's alphas (by fund category)

H_0	SD of x	SD of y	Df of x	Df of y	F-Statistic	P-value	Method	Decision
$\sigma_{MPr} = \sigma_{MD}$	0.0038	0.0050	1,042	1,728	0.58	4.1E-22	F test	Reject H_0
$\sigma_{MPr} = \sigma_{MPt}$	0.0038	0.0044	1,042	3,099	0.75	2.7E-08	F test	Reject H_0
$\sigma_{MD} = \sigma_{MPt}$	0.0050	0.0044	1,728	3,099	1.30	3.7E-10	F test	Reject H_0
$\sigma_{EPr} = \sigma_{ED}$	0.0025	0.0045	28	159	0.30	4.4E-04	F test	Reject H_0
$\sigma_{EPr} = \sigma_{EPt}$	0.0025	0.0024	28	327	1.10	0.667	F test	Not reject H_0
$\sigma_{ED} = \sigma_{EPt}$	0.0045	0.0024	159	327	3.65	0.0E+00	F test	Reject H_0
$\sigma_{GPr} = \sigma_{GD}$	0.0016	0.0028	15	30	0.30	0.017	F test	Reject H_0
$\sigma_{GPr} = \sigma_{GPt}$	0.0016	0.0030	15	55	0.28	0.008	F test	Reject H_0
$\sigma_{GD} = \sigma_{GPt}$	0.0028	0.0030	30	55	0.92	0.829	F test	Not reject H_0

This table presents the F-test to compare two variances for the average Jensen's alphas, where H_0 is the null hypothesis, representing that the differences in the variances are zero and the alternative hypothesis, H_1, represents that the differences in the variances are different from zero. The "decision" was made at the 5% significance level. "σ" represents the population variance. "M", "E" and "G" represents the bond mutual, bond ETF and green bond funds, respectively. "Pr" represents the pre-crisis period (Jan2005-Jul2007), "D" represents the crisis period (Aug2007-Dec2012) and "Pt" represents the post-crisis period (Jan2013-Dec2019). "SD of x" and "SD of y" represents the standard deviation of the variables and "Df of x" and "Df of y" represents the degrees of freedom of the variables.

Table 2B. Two-sample T-test for the average Treynor ratios (by fund category)

H_0	Mean of x	Mean of y	T-Statistic	P-value	Df	Method	Decision
$\mu_{MPr} = \mu_{MD}$	0.0005	0.0024	-11.47	9.9E-30	2,629.0	T test (Unequal var)	Reject H_0
$\mu_{MPr} = \mu_{MPt}$	0.0005	0.0005	0.00	0.998	2,049.3	T test (Unequal var)	Not reject H_0
$\mu_{MD} = \mu_{MPt}$	0.0024	0.0005	13.41	6.0E-40	3,199.5	T test (Unequal var)	Reject H_0
$\mu_{EPr} = \mu_{ED}$	-0.0007	0.0017	-4.01	1.6E-04	67.5	T test (Unequal var)	Reject H_0
$\mu_{EPr} = \mu_{EPt}$	-0.0007	-0.0002	-1.02	0.309	355.0	T test (Equal var)	Not reject H_0
$\mu_{ED} = \mu_{EPt}$	0.0017	-0.0002	4.91	1.9E-06	202.5	T test (Unequal var)	Reject H_0
$\mu_{GPr} = \mu_{GD}$	-0.0003	0.0024	-4.13	1.6E-04	44.8	T test (Unequal var)	Reject H_0
$\mu_{GPr} = \mu_{GPt}$	-0.0003	-0.0012	1.62	0.113	48.0	T test (Unequal var)	Not reject H_0
$\mu_{GD} = \mu_{GPt}$	0.0024	-0.0012	5.43	5.2E-07	85.0	T test (Equal var)	Reject H_0

This table presents the two-sample T-test for the average Jensen's alphas, where H_0 is the null hypothesis, representing that the differences in the means are zero and the alternative hypothesis, H_1, represents that the differences in the means are different from zero. The "decision" was made at the 5% significance level. "μ" represents the population mean. "M", "E" and "G" represents the bond mutual, bond ETF and green bond funds, respectively. "Pr" represents the pre-crisis period (Jan2005-Jul2007), "D" represents the crisis period (Aug2007-Dec2012) and "Pt" represents the post-crisis period (Jan2013-Dec2019). "Mean of x" and "Mean of y" represents the mean of the variables and "Df" represents the degrees of freedom.