US Bond Mutual Funds Skill, Scale and Value Added

Auteur : Georgitsopoulos, Ophélie
Promoteur(s) : Hambuckers, Julien
Faculté : HEC-Ecole de gestion de l'Université de Liège
Diplôme : Master en ingénieur de gestion, à finalité spécialisée en Financial Engineering
Année académique : 2021-2022
URI/URL : http://hdl.handle.net/2268.2/15577

HEC LIÈGE
Management School - Liège Université

US BOND MUTUAL FUNDS SKILL, SCALE AND VALUE ADDED

Jury:
Supervisor:
Julien Hambuckers
Readers:
Caterina Santi
Philippe Hübner

Master thesis by
Ophélie Georgitsopoulos
For a master's degree in Business Engineering with specialization in Financial Engineering
Academic year 2021/2022

Ecole de Gestion de l'Université de Liège

Acknowledgements

First of all, I would like to thank all those who directly or indirectly has contributed to the elaboration of this research thesis about US bond mutual funds' value added. This report is the result of long and hard work which would not have been possible without the support of many people.

I would like to express my gratitude to my supervisor, Dr. Prof. Hambuckers, Professor of Finance at HEC Liège Management School. He suggested the topic to me and guided me throughout this project as an ideal teacher and thesis supervisor.

I would also like to express my gratitude to my readers, Dr. Prof. Santi, Assistant Professor in International Finance at HEC Liège and Philippe Hübner, teaching assistant of Prof. Hambuckers at HEC Liège and doctoral student in Finance, for their interests in my research thesis.

Thank to HEC Liège Management School for providing me with the financial data which were essential to achieve this project.

Contents

List of abbreviations

1. AIC: Akaike Information Criterion
2. AMISE: Asymptotic Mean Integrated Squared Error
3. BIC: Bayesian Information Critetion
4. DEF factor: Default Spread factor
5. ECDF: Empirical Cumulative Distribution Function
6. EIV: Error in Variable
7. FDR: False Discovery Rate
8. HML factor: High Minus Low factor
9. ID: Identification Number
10. LW: Load Waived
11. MOM factor: Momentum factor
12. OLS: Ordinary Least Squares
13. R^2_{adj}: Adjusted R squared
14. RF: Risk Factor
15. SMB factor: Small Minus Big factor
16. TC: Total Cost
17. TERM factor: Term spread factor
18. TR: Total Revenue
19. US: United States
20. VaR: Value at Risk

1 Introduction

Mutual funds are experiencing substantial interest worldwide but especially in the United States where their importance does not seem to stop growing. Indeed, the global mutual fund industry knew an extensive growth in the 1990s, in part due to the globalization of Finance. In the United States,"the total net assets of mutual funds grew from USD 1.6 trillion in 1992 to 5.5 trillion in 1998, equivalent to an average annual rate of growth of 22.4%" (Fernando et al., 2003, p.2). Currently, according to the Mordor-Intelligence (2021) website, this amount keeps increasing and reached $21.3 trillion in total net assets in 2019, making the United States mutual fund industry remaining the largest in the world. In 2020, although equity funds alone accounted for 55.3% of the US mutual fund net assets, bond mutual funds were the second-largest category with 22% of the net assets (Mordor-Intelligence, 2021). Despite this importance, bond mutual funds are much less studied in the literature and more generally the bonds market receives little attention relatively to the stocks market. Yet in at least 1995, the bond market value was several times higher than the equity market (Elton et al., 1995).

In addition, bond pricing models have received little attention compared to equity pricing models. Still, it seems important to learn more about these models in order to know more about the performance of bond mutual funds. Indeed, mutual fund performance is still a frequent research topic but almost always related to the equity or hybrid mutual funds (eg. Barras et al., 2010, 2021; Sharpe, 1966).

As a result, this research aims to provide a different approach to analyze the performance of mutual funds such as those used in many previous studies and moreover to focus exclusively on bond mutual funds that are still important financial vehicles in the United States. To this end, we will follow the performance analysis of Barras et al. (2021) but on bond mutual funds. A relevant question in the literature is the presence or absence of positive excess return in the mutual fund industry and whether we eventually find positive excess return, then comes the question of presence and prevalence of luck for these performing funds (Barras et al., 2010). They conclude that a statistically indistinguishable from 0 number of funds were actually skilled (net of expenses). The difference in this approach is that we will break down the gross excess return, α, [1] into the skill and scale components of the fund to determine if the zero-alpha funds could not actually be skilled but too big to have a positive α. In other words, the difference in this work will be that we will not directly associate the fund's skill to its excess return but rather analyze this excess return in more depth to determine if the bond mutual funds' managers are skilled and if they create value. Barras et al. (2021) found that most equity mutual funds are actually skilled and can extract value from the market. Our results are consistent with their findings as we find that a large majority of bond mutual funds are truly skilled.

We are not unaware that mutual funds are experiencing diseconomies of scale, so mutual funds' managers could have skill and still get an average net excess return of 0 (Berk and Green, 2004). Following their work and Barras et al. (2021), we decided to model the fund gross return as $\alpha_{i,t} = a_i + b_i \times q_{i,t-1}$ where $q_{i,t-1}$ is the lagged fund's size. The skill component (a_i) is measured as the gross alpha on the first dollar invested in the fund. Finally, the b_i stands for the scalability which is the fund's sensitivity to diseconomies of scale. With this first part of the research, we can determine how many bond mutual funds, between 2004 and 2020, are actually skilled ($a_i > 0$). A particular advantage of this measuring is that we could allow a_i and b_i to be fund specific which reflects reality. To be able to determine the fund gross

[1] That is the excess return before that any expense is withdrawn.

return, we needed to construct a suitable asset pricing model for bond portfolios. Following Bai et al. (2019), we decided to take their newly introduced bond related risk factors that have been proven to be more efficient than the previous common risk factors in explaining the bond returns over the risk-free rate. However, unlike their papers, we do the study on all the available bonds on the market as we include the investments of bond mutual funds and not only corporate bonds. Consequently, we decided to add 4 more factors from Fama, French and Carhart (1997) in our asset pricing model as it was the best fit among the models we tested.

Then, based on this first part of the research, we analyze the value added of these funds. Still following Barras et al. (2021), we measure the value added as $va_i = E[\alpha_i \times q_{i,t-1}] = E[(a_i + b_i \times q_{i,t-1}) \times q_{i,t-1}]$. This measure of value added is quite simple because it has been defined as the product of the fund gross alpha and its size. This can be compared with the value added of a company calculated as the markup price of its products (fund gross excess return) and the sold quantities (fund size). The value added analysis developed by Barras et al. (2021) enables to have a cross-sectional distribution of the funds' value added. Our results showed that just like the equity mutual funds, the bond mutual funds can extract value from the market. Over their entire life from 2004 to 2020, these funds created about \$382K[2] of value and the range of value added is quite large as the standard deviation is almost 3 times the mean.

In addition, as the lagged fund's size varies over time depending on which period in their lifetime we are and that the previous value added measure takes only into account the entire period value added, we also decided to study the dynamics of the value added over time. To this end, we create subperiods and analyze in particular the last subperiod value added, $va_i(5)$.

Finally, a remarkable advantage of their method is that we could estimate the cross-sectional distributions of these different measures (a_i, b_i, va_i and $va_i(s)$) thanks to a nonparametric approach enabling to estimate the distributions more precisely than would have the parametric approach if the imposed density structure of the latter was incorrect. The motivations for following this procedure will be explained later.

[2] In terms of the base year that is 2000.

2 Literature Review

Many studies have already been performed on mutual fund performance and the results are often similar in the sense that mutual funds do not outperform the passive benchmark (Gruber, 1996; FRENCH, 2008; Fama and French, 2010). Although the measures of performance may differ, most results point in this way saying that the alpha is often negative or equal to 0. Sharpe (1966) studied performance of mutual funds with a measure of return and risk called the reward-to-variability ratio. They suggested that the lack of persistence in mutual funds performance could be partially due to the difference in management skill which goes in the sense of this thesis in the idea that we want the skill coefficient to be fund-specific as difference in management skill can occur across funds. However, they still concluded that actively managed mutual funds do not outperform the passively managed portfolios. Unfortunately, bond mutual funds do not seem to be exception to the rule although less in-depth studies have been conducted on this industry. Indeed, Elton et al. (1995) associated the performance of bond funds with their alpha (gross and net) and they found that the gross alpha (before expenses) was very small, so most of the net alphas of the funds are negative. As a result, they concluded that they found no evidence that managers were performing superior returns on the portfolio they manage. Thus, the skill of the managers is a frequent question in the literature since the general rule for mutual funds is well that they do not outperform the passive portfolios.

In that sense, Berk and Green (2004) wrote that there exists uncertainty that managers add value through their activities because performance is sometimes seen as luck. This latter statement would make no economic sense because then we would have no reason to reward it (fund managers offer really expensive services). In conclusion, Berk and Green (2004) affirmed that managers have differential skills to create abnormal return (alpha) but that this ability is undermined by scale. This statement goes in the sense of our thesis approach but has not been studied in the same way which makes our thesis research relevant.

However, there also exist some theories that do not go in the sense of our research. Indeed, ones still maintain that most active funds that can get gross α (before expenses) positive, have just been lucky (Fama and French, 2010; Barras et al., 2010). This supposes that managers' skills do not have anything to do with this positive excess return relative to common benchmarks. This supposition is driven by the performance measure that is often used (α) and from which we usually draw conclusions regarding the managerial skills whereas other measures could be better. Fama and French (2010) investigated the equilibrium accounting in the context of mutual fund performance, which supports that active investing must be a zero sum game[3]. The latter statement implies that if some active funds have true positive gross α, then other active funds must have negative gross α. This theory does not imply that every single fund exhibit an α equal to 0 but that some funds may actually exhibit high positive α due to the superior managers' skills on condition that other funds exhibit negative α possibly due to inferior managerial skills. According to Fama and French (2010), a lot of extreme events have happened by chance and not imperatively because managers are skilled. Indeed, their results on long-term performance showed that true net α (returns to investors) is negative for most of active funds. In other words, we can say that most of managers cannot beat the market by more than their cost which imply that the net return to investors is negative (after expenses and fees). On the other hand, their research examined a portfolio of active funds and studied the average α which does not imperatively mean either that any manager is skilled and that no value can be extracted from the market. Indeed, as it will be more explained in Section 3.1,

[3] In a game with 2 players, if one gains 10, the other loses 10. So the sum of all the gains subtracted by the sum of all the losses is always 0.

we decided to work on the gross returns as the returns to investors (net α) is not always the best choice to assess the real performance of mutual funds.

Regarding the article of Barras et al. (2010), they used an approach more similar to ours as they studied the performance of each fund separately and not via the mean of a portfolio of all funds. Indeed, they used a multiple-testing procedure and have built a measure (False Discovery Rate (FDR)) to identify the percentage of lucky funds among the funds that exhibit significant α different from 0. As our method is constructed on the same will to identify the performance at the fund-level, their findings are pretty relevant for us if we want to go further with respect to the analysis of the $\alpha_{i,t}$ by multiple-testing. Moreover, their results indicate that 76.6% of the fund managers have abilities even after accounting for the FDR but as soon as the expenses are taken into account the net α is still driven to 0. Although the net returns to investors is understandably often used as a measure for performance, it is not entirely dependent on managerial skills (Berk and van Binsbergen, 2015). So it can lead biased results if we only use this measure to identify the managers' abilities.

This is one of the reasons why this research thesis is relevant and interesting as we assess performance with another measure as many previous studies even though our measure is linked to the precedents (often gross and/or net α). As introduced earlier, our performance measure is the value added, va_i, that the funds create in the market which is different from numerous previous studies. Then, this work is also relevant for its intent to really determine if mutual fund managers are skilled (analyze of the skill coefficient, a_i) as it is also a frequent question in the literature. Finally, another major reason for showing interest in this research is that it is applied on bond mutual funds and not on equity funds which are the main industry studied in the literature whereas bond mutual funds also exhibit impressive records as introduced earlier.

3 Data

3.1 Bond Funds data

For this research, we needed the gross returns of US mutual funds that invest exclusively in bonds. We have taken the gross returns (return before any fees) as we want to correctly measure managerial skill. The choice of the gross excess returns instead of net excess returns is therefore conscious. Although the gross excess return is not yet a value measure, starting from it to compute our skill coefficient, a_i, is the best we can do to have a correct measure. Indeed, the net excess return is driven to 0 by the investor competition, so we could not take this as a managerial skill measure as it is not determined by the managers' skill (Berk and van Binsbergen, 2015). As introduced, our managerial skill measure, a_i, is the value extracted from the market on the first dollar invested in the fund, so it is our measure of value for the skill of managers.

One of the challenging parts in this research was to obtain the gross returns at the fund level. Indeed, Morningstar gives the monthly gross returns at the share class level for the entire population of bond mutual funds. Thereby, we took these data from July, 2004 to December, 2020 from the Morningstar database (open-end funds category). We obtained information on 17 098 funds' share classes, which corresponds to 5060 funds (retrieved by the Identification Number (ID) of the fund).

Each mutual fund offers different products, referred as share classes, based on the investment horizons of the investors. The most popular are the share classes A, B and C (Thune, 2021).

- A is the most attractive for long term investors with high initial investment as it has front-end charges but lower expense ratios if you let your money in the fund. Moreover, investors can benefit from "breakpoints discounts" as they invest a certain amount of money.

- B is attractive for investors with middle-term to long-term investment horizons but with lower capital amounts. This share class charges a contingent deferred sales charge which decreases as the holding period increases but the charges are the same for any amount of investment, investors cannot get reduction of the charges as they invest more money. Additionally, it has higher expense ratios than the other share classes.

- C is made for investors who do not want to let their money in the long-run, this share class charges yearly 'level-load' fee (around 1%) and this fee never goes away, so this is the most expensive share class if investors want to invest in the long-term. The main advantage of this share class is that investors can redeem their fund's shares in the short-term.

- The "Load Waived (LW)" category is for particular investors with particular conditions. It is also often seen and allows investors to be exempt of fees. Some bond mutual funds offer a LW option of their share classes (often on the share class A).

With these data, we could compute the fund level gross returns as Pollet and Wilson (2008) did it: We took the total fund size which is equal to the sum of the share class' net assets. Then, we computed the fund gross returns by making the share class' net assets weighted average of the share class level gross returns. Thanks to this, we can determine if the actual bond mutual funds managers (as of 2004) are skilled and can create value in the market at the fund-level and not at the share class level.

The initial computed data set contained all the funds of the period but we had to apply some selection rule in order to have enough observations to get some reliable estimators of a_i, b_i, va_i and $va_i(s)$. We chose to keep only those funds for which we had at least 60 monthly observations as it is a threshold commonly used in studies. Then, we had 2258 funds during the sample period on which we applied another selection rule discussed later in Section 4.3.1 to obtain 2257 funds in our final sample. Thereby, we are aware that a survivorship bias could arise first from the fact that we study a sampling period as of 2004 and then from the first selection rule imposed on the funds to be included in the final dataset. Thus, we maintain that this research will only contain conclusions regarding current and relatively new bond mutual funds as many funds (skilled and unskilled) could have existed before 2004 and have since disappeared. For the potential bias arising from the first selection rule (number of observations to 60), we know from Blake et al. (1993) that the survivorship bias in bond mutual funds is less important than for stock funds as it is less variable so fewer bond funds dissolve or merge. Moreover, since Barras et al. (2021) concluded that the survivorship bias coming from their selection rule (minimum number of monthly observations to 60) was minimal and that their threshold was optimal to mitigate both the survivorship bias and the reverse survivorship bias [4] while they studied stock mutual funds, we have good hope that the survivorship bias introduced by our first selection rule is minimal too.

As a measure of size, we took the aggregate share class net assets, which corresponds to the total assets invested in the fund and that is useful in gauging the fund size as defined by Morningstar. This definition is consistent with several studies using also the net assets as measure of the size of the fund (Chen et al., 2004; Pollet and Wilson, 2008). We decided to run a linear interpolation to compute the potential missing values of both the share class' net assets and the fund size from aggregate share class as this approach was also used by Barras et al. (2021). Once we had the share class' net assets and the fund size, we were able to calculate the proportion of funds coming from one share class or another. So we had all we needed to compute the funds gross returns as previously explained.

Finally, as our research covers several years and as we have measures in \$, we had to take into account the inflation because \$1 in 2004 is not equal to \$1 in 2020. We have therefore chosen a reference year, 2000 as it is the same as Barras et al. (2021), so we will be able to compare the value added measures. Thereby, we retrieved the yearly cumulative inflation rate with 2000 as base year from the website US Inflation Calculator (2022) and we adjust the size values of each year by the corresponding inflation rate. For example, the cumulative inflation rate between 2000 and 2016 is equal to 39.4%, so \$1 in 2000 is worth \$1.39 in 2016 (1×1.394). Thus, to get all the values in \$ back in 2000 (base year) we took the \$1 in 2016 divided by 1.394 to get the real worth of this Dollar in 2000 (\$0.72).

3.1.1 Descriptive Statistics

From the graph below (Figure 1), you can observe the distribution of the funds returns averages. It shows that on average, most US bond funds exhibit positive returns but below 1%. Moreover, the average of the standard deviations of the funds is equal to 1.6823% which is not very dispersed. You can also find the standard deviation distribution below (Figure 2) and you can indeed observe that the large majority of the funds exhibit a standard deviation between 0 and 2%. Knowing that this measures the average difference between the various returns of one fund and its average return, we can say that globally the returns of the funds do not extremely vary.

[4] Bias that arises if actually skilled funds disapppear after unexpectedly low returns (Barras et al., 2021).

Then, still based on the returns averages distribution (Figure 1), we can more or less think of a normal distribution. As normality is a common feature of distribution which is often checked when speaking about returns and as a lot of tests and results rely on normality assumption, we decided to run a Lilliefors test on the returns of each fund. We chose this test because we did not need to estimate the parameters of the null distribution in contrast to the Kolmogorov-Smirnov test with which we must estimate them beforehand. The results showed that out of the 2257 funds in our population, we rejected the null for 1741 funds which represent 77.14%. That means that most funds do not have normally distributed returns at the 5% level.

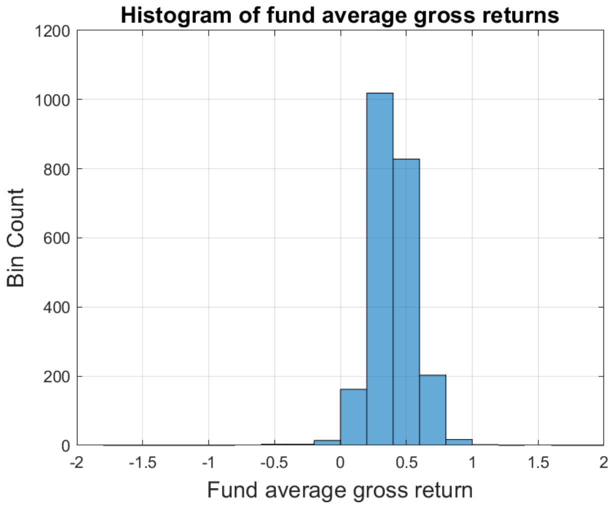

Figure 1: Histogram of the averages of bond funds gross returns

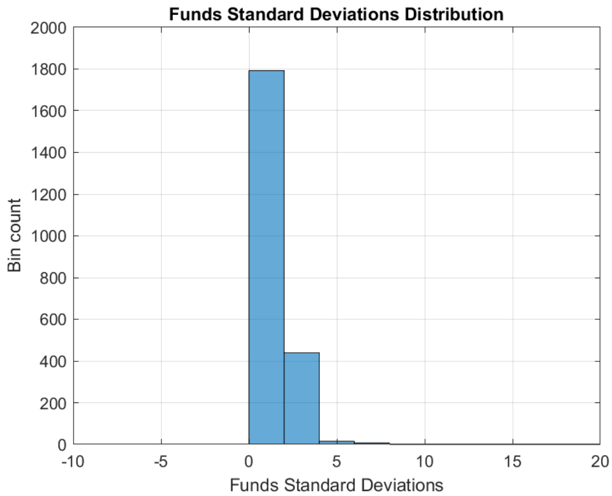

Figure 2: Histogram of the funds standard deviations

3.2 Risk factors data

To complete this research, we also needed bond related risk factors in order to extract the gross excess returns, α. Although their papers only focus on corporate bonds, we decided to use the 3 bond common risk factors introduced by Bai et al. (2019):

- **Downside risk**: measured with the 5% value at risk (VaR), so this is the potential decline in value of an asset over a certain period and given a certain probability. This risk factor accounts for the extraordinary events that could occur on the market as stock market crashes or bond market collapses. More and more attention is given to the risk management aptitude of financial and non-financial firms and simultaneously methods were developed to calculate the risk these firms face. The VaR is a primary tool to calculate this risk, so it seemed natural for us to use it. This factor is calculated by taking the value weighted average return difference between the highest VaR-portfolio and the lowest-VaR portfolio across the rating portfolios (Bai et al., 2019).

- **Credit quality**: measured via the credit ratings of the corporate bonds since these ratings capture information on bond default probability and the loss severity. These ratings furnished by rating agencies take into account information regarding both the bond issuer and the bonds themselves, what makes them good proxies to evaluate the credit quality of the bonds. This factor captures thus the difference of return you can get by choosing a lower rated bond rather than a better one as it is the value-weighted average return difference between the lowest rating portfolio and the highest rating-portfolio across the VaR-portfolios (Bai et al., 2019).

- **Bond illiquidity**: measured as $-Cov(\delta p_{itd}, \delta p_{itd+1})$ where $\delta p_{itd} = p_{itd} - p_{itd-1}$ is the log price change for bond i on day d of month t. This factor must extract the transitory component from bond price. It is calculated by taking the value-weighted average return difference between the highest-illiquidity and the lowest-illiquidity portfolios across the rating portfolios (Bao et al., 2011, cited by Bai et al. (2019)).

As you might have understood it, Bai et al. (2019) constructed their risk factors in a similar vein as Fama and French (1993), constructing bivariate portfolios by independently sorting their corporate bonds into quintiles based on the measures mentioned above. Their principal sorting variable was the credit quality as it is one of the most followed risk measure. Then, still following Bai et al. (2019), we also decided to add the excess bond market return as a risk factor. We took all these factors from Bali (2019). We were confident enough to choose the model of Bai et al. (2019) as our main model because according to their results, their newly proposed risk factors outperformed all other models considered in the literature in explaining the returns of the corporate bond industry. Indeed, the most effective bond pricing model in their research was the one built with these 4 risk factors mentioned above.

Although we decided to keep them as a potential model, we still wanted to try to complete this model by adding some other known risk factors because our dataset is made of bond mutual funds that do not necessarily invest exclusively in corporate bonds. Consequently, we also took the excess stock market return, the size factor (SMB), the book-to-market factor (HML) and the momentum factor (MOM) from the data library on the Kenneth R. French's website (French, 1993). Finally, we computed the default spread factor (DEF) and the term spread factor (TERM) following Bali et al. (2014) as the DEF factor is measured as the difference between yields on BAA-rated and AAA-rated corporate bonds and the TERM factor is measured as the difference between yields on ten-year and three-month Treasury securities. These yields data come from the St. Louis FRED website (Federal Reserve Bank of St. Louis, 1991).

4 Developments

4.1 Bond-oriented asset pricing model

The first step in our research was to find an adequate bond-oriented asset pricing model in order to perform a linear regression and be able to express the bond mutual fund's excess gross return relative to the risk-free rate[5] as a function of the fund's skill (a_i), the fund's monthly lagged size ($q_{i,t-1}$) and the chosen risk factors (f_t) (Equation 4.1).

$$Y_{i,t} = a_i + b_i \times q_{i,t-1} + \beta_i \times f_t + \epsilon_{i,t} \tag{4.1}$$

where $Y_{i,t}$ is the fund's gross return (before fees) over the risk-free rate, f_t is a column vector ($K_f \times 1$) for the common risk factors (RF) relative to excess returns and $\epsilon_{i,t}$ is the error term.

We had to find the factors, f_t, that we were going to use. Based on the results of Bai et al. (2019), we decided to start with a model of 4 common risk factors that they studied on corporate bonds and that we have introduced in the section 3.2. Indeed, according to their studies, the bond market, although the latter and the stock market are integrated, also has its own unique features such as the credit risk, due to the need for bond issuers to pay coupons; a particular sensitivity to downside risk as the upside payoffs on a bond are capped, so bondholders are more sensitive to downside risk than stockholders who can better benefit from great news in firm fundamentals; and a much less liquid market than the stock one. For these different characteristics, we decided to base ourselves on their asset pricing model.

However, in our research, we did not include only corporate bonds but all bonds in which mutual funds could have invested. This suggests to us that their asset pricing model could perhaps be completed by some additional factors as it concerns all the bond mutual fund investments. Consequently, we constructed 3 other linear regression models:

- Model 2: Composed of the 4 risk factors previously introduced along with the excess stock market return, the size (SMB), the book-to-market (HML) and the momentum factors (MOM). This model requires thus to estimate 9 coefficients.
- Model 3: Composed of the 4 risk factors previously introduced along with the default spread (DEF) and the term spread (TERM).
- Model 4: Composed of all the factors mentioned.

At this point, we had thus 4 different models for the linear regression below (Equation 4.2) from which we quickly want to compare the performance between each other at explaining the bond mutual funds returns before going on with our initial goal. In each model, the number of columns (K_f) for the vector f_t varies such that it is in order (from Model 1 to Model 4) equal to 4, 8, 6 and 10.

$$Y_{i,t} = \alpha_i + \beta_i \times f_t + \epsilon_{i,t} \tag{4.2}$$

[5] proxied by the 1-month Treasury Bill return from Ibbotson and Associates, Inc.

In order to select the best model among those introduced, we decided to look at the average adjusted R^2, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) since they allow to compare all the models with each other (nested or not) and they assign a penalty according to the number of parameters to be estimated. Below (Table 1) are the average values of these measures for each model.

	R^2_{Adj}	AIC	BIC
Model 1	47.31%	$-1.0048e+03$	-990.12
Model 2	53.64%	$-1.0233e+03$	-996.83
Model 3	49.29%	$-1.0085e+03$	-987.94
Model 4	55.41%	$-1.0265e+03$	-994.14

Table 1: Averages Adjusted R-square and Information Criteria

The 2 best values of these measures have been colored with the corresponding models. As you can see, the choice of the best model fit will be between model 2 and model 4 according to the AIC/BIC analyze. Although the BIC of model 2 and 4 are very close, the results of these criteria compete with each other. Therefore, we looked at the R^2_{adj} which only increases by less than 2% by adding 2 factors (model 4) which means that the variation of $Y_{i,t}$ that is explained by the model increases only by 2% as we add 2 more factors in the model, so they do not bring much extra explanation relatively to the penalty they cause.

Another useful tool to compare these 2 competing models is the empirical cumulative distribution function (ECDF) of their respective R^2_{adj}. Below (Figures 3 and 4), you can observe 2 graphs which represent for each model (2 and 4) the percentage of funds for which the R^2_{adj} was higher than the measure on the x-axis (as we took the measure $1 - F(x)$ on the y-axis because it was more legible). From these graphs, we know that a bit more than 34% of the funds exhibit a R^2_{adj} above 80% with the configuration of Model 2. To compare, with model 4, this percentage of funds increases to 37%, so these 2 ECDF are close to each other.

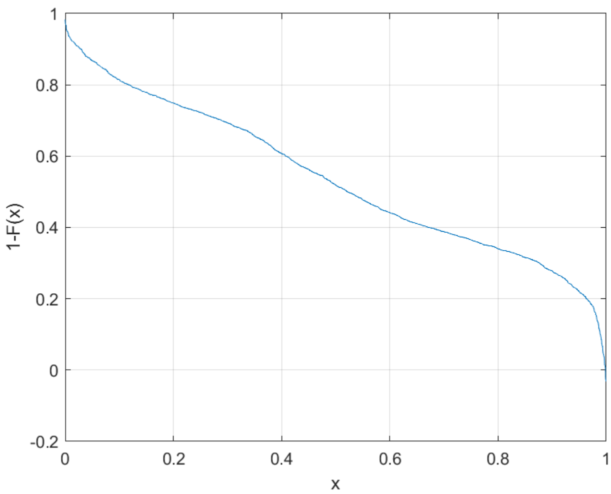

Figure 3: 1-F(x) for each value of R^2_{adj} from Model 2

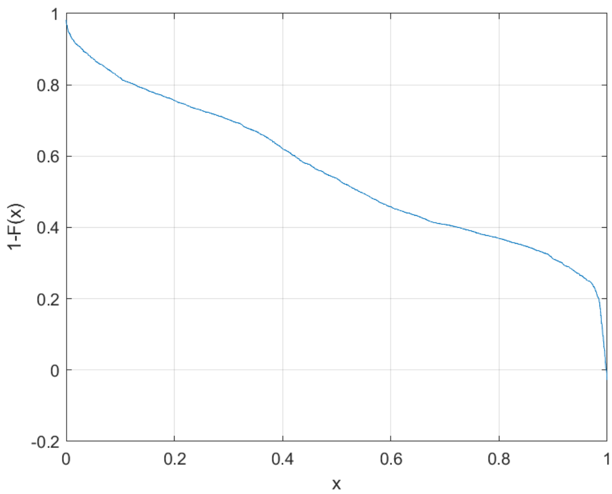

Figure 4: 1-F(x) for each value of R^2_{adj} from Model 4

As a result, following parsimony[6] and since the BIC is better for model 2, we decided to keep **model 2** as asset pricing model for our research. This choice comes from the will to keep a relatively parsimonious model while still fitting to the data as well as possible.

Now that the adequate risk factors are chosen, we can analyze in more depth how interact the bond mutual fund returns with these factors. In Appendix A.1, you can find the distributions of the intercept (α_i) and of the different coefficients for each risk factor included in the model.

The analysis of the intercept is not what will matter the most from this regression as our main purpose is to later break down this α_i for each fund into its skill and scale coefficients because our performance measure is not exactly this abnormal return. Nevertheless, it is still interesting to look a bit at its distribution to notice that the α_i are concentrated around 0 with positive and negative values although the majority of the funds seems to get an α_i that is between 0 and 0.002. Moreover, it seems that some funds are experiencing some very high abnormal returns compared to the average because the right tail goes far (positively skewed).

Regarding the analysis of the β_i, their distributions will show how our population of bond mutual funds interacts with the different risk factors in our model. Indeed, the β_i are the premia per unit of determinable risk (risks that are represented by the different risk factors) from which the mutual funds can benefit (or not) according to their investment. In other words, we also say that the β_i are the sensitivities of the fund i with respect to each variable. For example, the distribution of the β_i relative to the Bond Market factor (Appendix 12) shows that most of the coefficients are concentrated between 0 and 2.5 and that the distribution is positively skewed. This means that a large majority of the funds moves with the bond market, i.e. when the bond market is doing well, most bond mutual funds are doing well too as their β_i relative to this factor are mostly positive and even greater than 1.

We can observe other factors to which our bond funds population is on average positively sensitive (Table 2) like the Credit RF, Liquidity RF and Stock Market. From this, we know that most funds in our population invest in bonds from lower-rated companies and that can be quite illiquid as on average the β_i relative to these corresponding RF are positive. This means that on average our population benefits from the premia related to these determinable risks because they invest accordingly to them. On average, the Stock market β_i are also positive but the median is slightly negative which means that there is probably a little majority of funds that are negatively sensitive to this variable and the other part that is positively impacted. So we can conclude that the sensitivity to this RF in our population is rather mitigated and that still an important amount of bond mutual funds seem to move with the stock market (i.e. this could be explained by the presence of corporate bonds in their portfolio).

[6] We preferred a good model with less predictors to a slightly better model with more predictors. Indeed, we preferred to have 8 variables instead of 10.

	Mean	Median
Intercept	$9.54e-04$	$9.31e-04$
Bond Market Coefficient	0.57	0.52
Downside RF Coefficient	-0.045	-0.021
Credit RF Coefficient	0.086	0.015
Liquidity RF Coefficient	0.0025	$4.86e-04$
Stock Market Coefficient	0.022	-0.009
SMB Coefficient	-0.0082	-0.0134
HML Coefficient	-0.012	-0.022
MOM Coefficient	-0.019	-0.014

Table 2: Mean and Median of the Intercept and the different RF coefficients

On the other hand, our funds population is on average negatively sensitive to the Downside RF, SMB, HML and MOM which means that most funds do not invest accordingly to these risk and thus do not benefit from the corresponding premia. For example, from the results relative to the SMB factor, we could assume that on average the bond mutual funds invest in bonds issued by big companies. Moreover, concerning the downside risk, we can assume that on average our bond funds population is not exposed to this risk (as the average β_i is negative), so they invest more in bonds from instituions/companies with low VaR.

To conclude, if you look at the different distributions in Appendix A.1, you can observe that the whole distributions are concentrated around 0, meaning that although the average is sometimes negative (positive), some funds may (not) still benefit from premia for some RF while the average of funds does not (does). Furthermore, we can also observe that most of the distributions exhibit one tail that goes far (right or left) and thus are not very symmetrical except the distribution of the β_i relative to the momentum factor that is more symmetric. The asymmetry in the other distributions reflects that some funds appear to exhibit extreme coefficient values either positive or negative if they invest (maybe exclusively) in bonds that are risky with relative to the corresponding RF or not at all.

4.2 Skill, Scale and Value added measures

As Barras et al. (2021) followed Berk and Green (2004) to construct their linear model, we did the same. Indeed, they assumed that the costs incurred by actively managing is a convex function that depends on the size such that it can be expressed as $TC_{i,t} = b_i \times q_{i,t-1}^2$, then, we assumed that the total revenue of the fund over the benchmark was $TR_{i,t} = a_i \times q_{i,t-1}$. As a result, if we take the difference $TR_{i,t} - TC_{i,t}$ and after divide by $q_{i,t-1}$, we have the gross alpha expressed as $\alpha_{i,t} = a_i - b_i \times q_{i,t-1}$ such that it depends on the lagged fund size. Thereby, the skill coefficient a_i is measured as the performance of the fund when $q_{i,t-1} = 0$ which "captures the profitability of the fund's investment ideas" (Barras et al., 2021, p.9). In addition, the scale coefficient b_i reflects the fund's sensitivity to diseconomies of scale, so that is the fund's gross alpha decrease as the lagged size increases by one.

As it can be noted, the skill and scale coefficients are fund specific, so this approach does not use a panel specification which includes that all the funds should have the same a and b but treats them as random realizations from their cross-sectional distributions $\phi(a)$ and $\phi(b)$. This approach is consistent with the economic perspective that the diseconomies of scale should not be the same across all funds.

From this point, we could estimate the value added of the fund that is defined as the "the average product of the fund gross alpha and size" (Barras et al., 2021, p.10).

$$va_i = E[\alpha_{i,t} \times q_{i,t-1}] = a_i \times E[q_{i,t-1}] - b_i \times E[q_{i,t-1}^2] = a_i \times p \lim_{T \to \infty} \bar{q}_i - b_i \times p \lim_{T \to \infty} \bar{q}_{i,2} \quad (4.3)$$

where $\bar{q}_i = \dfrac{1}{T} \sum_{t=1}^{T} q_{i,t-1}$ and $\bar{q}_{i,2} = \dfrac{1}{T} \sum_{t=1}^{T} q_{i,t-1}^2$ denote the time-series averages of the fund size and its squared value and $p\lim$ denotes the limit in probability.[7] The convergence to an equilibrium value of $\bar{q}_i$ when the costs function is convex is rather logical as this supposes that the fund on the long run tries to minimize its cost by reaching and remaining at the critical value $\bar{q}_i$ that minimizes the costs function.

However, this measure (4.3) takes into account the value added that the fund creates throughout its entire lifetime over the studied period. As we know that the fund's size varies over time, this measure could not reflect precisely the potential fluctuations in the fund's value added. As a result, we split each fund's return history into S subperiods in order to examine the different value added measures on each subperiod which depend on the average size over the corresponding subperiod. Thereby, the value added of the subperiod s, $va_i(s)$, is computed like this:

$$va_i(s) = a_i \times \bar{q}_i(s) - b_i \times \bar{q}_{i,2}(s) \quad (4.4)$$

where $\bar{q}_i(s)$ and $\bar{q}_{i,2}(s)$ are the realized averages (respectively) of the fund size and its squared value over the subperiod s ($s = (1, ..., S)$). With this measure, we will be able to analyze deeper the dynamics of the size over our relatively recent studied period.

This approach to calculate the value added is also fund-specific as it is derived from a_i and b_i. Moreover, va_i and $va_i(s)$ are also treated as random realizations from their cross-sectional distributions $\phi(va)$ and $\phi(va(s))$ as they inherit the randomness of the 2 coefficients.

Before carrying on, some remarks need to be made about these different specifications. First, concerning the baseline specification, $\alpha_{i,t} = a_i - b_i \times q_{i,t-1}$, 2 comments must be made.

First of all, the estimation of a_i and b_i does not require that we specify the determinants of these coefficients across funds. In this way, we do not need to determine if a fund A has extremely talented managers compared to a fund B if the skill coefficients are different. In the same way for the scale coefficient, it could be different because fund A trades more efficiently compared to fund B. Instead, we can simply interpret a_i and b_i as functions of the characteristics of the funds like the managers and the fund's strategy (liquidity and turnover).

Second, this specification of the gross α could omit some time-varying effects that could lead the coefficients values. Indeed, the skill may also depend on time as the levels of industry competition vary just like the scale coefficient (b_i) may vary with time if the relationship between the gross α and the lagged size was actually non-linear. These problems may lead bias in the coefficients estimations and end with false conclusions regarding the presence of skilled funds in the population or the magnitude of the diseconomies of scale the funds face. To examine this issue, Barras et al. (2021) made an extensive analysis on the coefficients computed on short-time windows to evaluate the impact of time on them. As a quick conclusion to their analysis, they found that the empirical results are not driven by the omission of important variables in

[7] The limit in probability referred to the convergence in probability to a fund size of $\bar{q}_i$, as T is large.

their baseline specification. Since we have still done our analysis on mutual funds and that this specification is not directly linked to the kind of assets managed, we can hope that the same conclusions can be drawn for our case without exaggeration.

4.3 Method

Our goal is thus to estimate the density of the measure $m_i \in \{a_i, b_i, va_i, va_i(s)\}$ now we have chosen an asset pricing model. To achieve this work, we have to follow 3 steps:

- Estimate the different coefficients
- Apply the Kernel density estimation
- Adjust the density for the potential bias

To achieve these steps, we use the non-parametric approach from Barras et al. (2021). Indeed, this approach has several advantages that may apply in our case. First, the biggest advantage of this approach is that we do not have either to respect some assumptions about the data sample or to estimate the potential parameters that define the underlying distribution of the data. Thereby, we do not need to determine or to define the actual density of the data and by the way the number of parameters that underlie this distribution, so the great misspecification risk we avoid by this method is considerable. Moreover, it is an important advantage in our case where we do not have much guidance in the theory concerning the distribution of these measures (Barras et al., 2021). As a result, the distributions of the data will be given by the data sample itself and not guided by theoretical assumptions even though we will deduce the asymptotic properties of each estimated quantity and rely on econometric theory to make statistical inference.

Then, the non-parametric approach is also simple and quick to implement, once you have chosen your non-parametric method, it provides an unified framework for estimating both density and its different characteristics and finally, it comes with a fully developed inferential theory.

Obviously, the non-parametric approach also has its disadvantages of being less powerful than the parametric approach if the assumptions of the corresponding parametric method are maintained. However, as we jointly study 4 measures, the task of correctly specifying a multivariate distribution whose marginals are potentially mixtures of distributions with different supports is really dissuasive. In addition, Cai et al. (2018) also uses a non-parametric approach to estimate the time-varying funds' alphas that they interpreted as skill indicator instead of the constant coefficient (the traditional α that is constant over time for one fund), so the method employed seems to be the best considering the measures we want to estimate.

We will explain deeper each step of the method in the following subsections. We insist that the whole method exposed here after is taken from Barras et al. (2021).

4.3.1 Estimation of the measures

Starting from the linear regression equation 4.1 which takes into account the decomposition of gross α and the common risk factors adapted to bond portfolios, we will estimate a_i and b_i. As a reminder, a_i and b_i are considered as random realizations from their cross-sectional distributions. Once we have these estimated measures for each fund, we can infer their density.

Although the non-parametric approach imposes few assumptions, there is still one assumption to be made here regarding the size of the fund over the long-term. Indeed, we have to assume that the size of the fund is asymptotically stationary. That means that as t grows large, the size of the fund will become independent of past size values. This assumption is accepted as it is consistent with any model that features diseconomies of scale at the fund or industry level (Chen et al., 2004). Moreover, this implies that in the early years (t is small), the size of the fund can change upwards as the investor's learning process is strong. Then, the size is stabilized around a certain mean (no more upwards trend) and with a certain fluctuation. The intuition behind this accepted assumption has been exposed earlier when speaking about the convex costs function.

Then, another important remark is the estimated coefficients that could be biased as the mutual fund size $q_{i,t}$ can be modelled with the regression below (Equation 4.5)[8]:

$$q_{i,t} = \theta_{q_i} + \rho_{q_i} \times q_{i,t-1} + e_{q_i,t} \tag{4.5}$$

As we know that $e_{q_i,t} = \epsilon_{q_i,t} + \beta_{q_i} \times x_t$ where $x_t = (1, f_t')'$, we know that the size at time t can be explained by a previous value of the time series (size at lag 1), the other independent variables f_t and what we called the innovation in size $\epsilon_{q_i,t}$. As a result, we can say that the lagged size is partially endogenous in time series. This violates the exogeneity assumption of the OLS regression model which requires that the error term is independent of the explanatory variables. In other words, it should not be possible to explain part of the error term through the explanatory variables. However, the driving forces of the reality cannot exactly be reproduced by theoretical models. Therefore, the error term appears to be a mix of the randomness of the process and a certain amount of omitted variables (MathWorks, 2022). When the omitted variable contain one important variable, it creates endogeneity.

To understand better where this violation comes from, let use this example from our case: Assume that the "true" model to explain our dependent variable $Y_{i,t}$ is

$$Y_{i,t} = a_i + b_i \times q_{i,t-1} + \beta_i \times f_t + \psi_i \times \epsilon_{q_i,t} + \nu_{i,t} \tag{4.6}$$

However, for some reasons, we did not include $\epsilon_{q_i,t}$ in the model, so it is estimated as previously in Equation 4.1:

$$Y_{i,t} = a_i + b_i \times q_{i,t-1} + \beta_i \times f_t + \epsilon_{i,t} \tag{4.7}$$

where $\epsilon_{i,t} = \psi_i \times \epsilon_{q_i,t} + \nu_{i,t}$ so that the innovation in size ($\epsilon_{q_i,t}$) has been absorbed in the bond returns innovation ($\epsilon_{i,t}$). If the correlation between $q_{i,t-1}$ and $\epsilon_{q_i,t}$ is not 0 (which is the case) and that this $\epsilon_{q_i,t}$ affects the bond returns $Y_{i,t}$ (meaning that $\psi_i \neq 0$), then one of our explaining variables $q_{i,t-1}$ is correlated with the error term $\epsilon_{i,t}$ and the exogeneity assumption is violated. As a result, the coefficients suffer from the small-sample bias as it is called (as it vanishes asymptotically) because the $\epsilon_{q_i,t}$ is positively correlated with the $\epsilon_{i,t}$.

To remove this bias, Barras et al. (2021) followed Amihud and Hurvich (2004) and Avramov et al. (2013) and we will follow their lead too. The idea is linked to our previous example as it is to include a proxy for the size innovation, $\epsilon^c_{q_i,t}$, in the explanatory variables in order to remove it from the mutual fund error term, $\epsilon_{i,t}$. We know that the innovation in size is positively correlated with the error term, i.e., $\epsilon_{i,t} = \psi_i \times \epsilon_{q_i,t} + \nu_{i,t}$ where $\psi_i > 0$. In particular,

[8] As t is not large, we are not talking about the asymptotic behaviour anymore.

$\epsilon_{q_i,t}$ denotes the size innovation coming from $\epsilon_{q_i,t} = e_{q_i,t} - \beta_{q_i} \times x_t$ where $x_t = (1, f_t')'$ and $e_{q_i,t}$ is the innovation of the size regression from Equation 4.5 : $q_{i,t} = \theta_{q_i} + \rho_{q_i} \times q_{i,t-1} + e_{q_i,t}$. If we do not adjust for this small-sample bias, the estimated coefficients, a_i and b_i, will be too high compared to their actual values since the correlation is positive.

Consequently, adding the regressor, $\epsilon_{q_i,t}$, removes the small-sample bias. To do it, we will replace the mutual fund error term, $\epsilon_{i,t}$ by $\psi_i \times \epsilon_{q_i,t} + \nu_{i,t}$ to have

$$Y_{i,t} = a_i + b_i \times q_{i,t-1} + \beta_i \times f_t + \psi_i \times \epsilon_{q_i,t} + \nu_{i,t} \tag{4.8}$$

and we will check that the exogeneity assumption is now verified, so that $E[\nu_i|X_i] = 0$ where $\nu_i = (\nu_{i,1}, ..., \nu_{i,T})'$ and X_i is the $T_i \times (K_f + 3)$ matrix of the available observations $x_{i,t} = (1, -q_{i,t-1}, f_t', \epsilon_{q_i,t})$ with K_f which is the number of factors. However, as we cannot observe the true innovation in size, $\epsilon_{q_i,t}$, we compute a proxy denoted as $\epsilon^c_{q_i,t}$ (Amihud and Hurvich, 2004; Avramov et al., 2013). Following their four-step procedure on each fund i individually ($i = 1, ..., n$), we first run the size regression to obtain the estimated coefficients, $\hat{\theta}_{q_i}$ and $\hat{\rho}_{q_i}$. Second, we compute the adjusted size innovation as

$$e^c_{q_i,t} = q_{i,t} - \hat{\theta}^c_{q_i} - \hat{\rho}^c_{q_i} \times q_{i,t-1} \tag{4.9}$$

where the second order coefficients corrected for the small sample bias are given by $\hat{\rho}^c_{q_i} = min(\hat{\rho}_{q_i} + (1 + 3\hat{\rho}_{q_i})/T_i + 3(1 + 3\hat{\rho}^2_{q_i})/T_i^2, 0.999)$ and $\theta^c_{q_i} = (1 - \hat{\rho}_{q_i})\bar{q}_i$. Third, we had to regress $e^c_{q_i,t}$ on the K_f factors to get $\epsilon^c_{q_i,t} = e^c_{q_i,t} - \hat{\beta}_{q_i} \times x_t$. Finally, we can insert $\epsilon^c_{q_i,t}$ in Equation 4.8 to obtain

$$Y_{i,t} = a_i + b_i \times q_{i,t-1} + \beta_i \times f_t + \psi_i \times \epsilon^c_{q_i,t} + \nu_{i,t} \tag{4.10}$$

Thereby, from this regression using least-square estimation method, we can get estimated coefficients for a_i and b_i that are corrected for the small sample bias, $\hat{\gamma}_i = \{\hat{a}_i, \hat{b}_i, \hat{\beta}_i, \hat{\psi}_i\}$, such that

$$\hat{\gamma}_i = \hat{Q}^{-1}_{x,i} \frac{1}{T_i} \sum_{t=1}^{T} I_{i,t} x_{i,t} Y_{i,t} \tag{4.11}$$

where $I_{i,t}$ is an indicator variable equal to 1 if the gross return of the fund i at time t is observable, T_i is the number of observations for fund i, $x_{i,t} = (1, -q_{i,t-1}, f_t', \epsilon^c_{q_i,t})'$ is the vector of explanatory variables and $\hat{Q}_{x,i} = \frac{1}{T_i} \sum_{t=1}^{T} I_{i,t} x_{i,t} x'_{i,t}$ is the estimated matrix of the second moments of $x_{i,t}$. The latter gives an overview of the linear dependence between the explanatory variables as it is a variance-covariance matrix of the explanatory variables.

For each fund individually, these coefficients were thus estimated in relation to the period during which the gross returns were observable. In order to have some reliable coefficients, we decided to apply 2 selection rules as Barras et al. (2021): the first one requires the fund to have at least 60 monthly observations as it is a recurrent threshold used in literature and the second rule aims at dealing with some potential collinearity problems.

Indeed, collinearity is a threat to correct estimations of coefficients measures as it greatly increases their variance, so it is important to check that any explanatory variable is not a linear combination of another variable. In order to check that, Barras et al. (2021) constructed a selection rule, following Gagliardini et al. (2016), based on the matrix $\hat{Q}_{x,i}$ such that $cond_i =$

$\sqrt{eig_{max}(\hat{Q}_{x,i})/eig_{min}(\hat{Q}_{x,i})}$ is the condition number of the matrix $\hat{Q}_{x,i}$ constructed by the ratio of the largest to smallest eigenvalues, eig_{max} and eig_{min}. This will need to be lower than the threshold, which is equal to 15, in order for the fund to be selected as it is the threshold advocated by Greene (2008) (as cited in Gagliardini et al., 2016)

4.3.2 Kernel Density: densities inference

In this section, we compute the density function, ϕ, thanks to the non-parametric method which is the Kernel estimation. Thereby, the estimated density, $\hat{\phi}$, for a given point m is computed like this:

$$\hat{\phi}(m) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{\hat{m}_i - m}{h}), \tag{4.12}$$

where n is the number of selected funds after the application of the 2 selection rules, K is a symmetric Kernel function and h is the vanishing bandwidth which determines how many observations around point m are used for estimation (Barras et al., 2021).

As pointed it out by Barras et al. (2021) and in the theory of the Kernel density estimation, the choice of the Kernel function is not crucial in the analysis, so they favored simplicity and chose the standard Gaussian Kernel $K(u) = \frac{1}{\sqrt{2\pi}} exp(-\frac{u^2}{2})$ and we decided to do the same.

With this method, the estimated density, $\hat{\phi}$, counts the percentage of observations which are close to m, which corresponds to $\frac{\hat{m}_i - m}{h}$ that will be small for each estimated $\hat{m}_i$ that is close to the true point m. The smaller $|u|$ is in $K(u)$, the higher the function K will be, so the weight of this estimation in the kind of average that we make to compute the estimated density $\hat{\phi}$ will be higher than the one for which $\hat{m}_i$ was far from m. The bandwidth h controls for the degree of smoothing.

The non-parametric approach hardly ever involves being indexed by a bandwidth or tuning parameter which controls the degree of complexity. The choice of this index is important for the implementation as it must imply a good bias-variance trade-off (Hansen, 2009). Indeed, if h is too small, we could overfit the data and increase too much the estimation variance. In contrast, if h is too big, we could smooth too much the density and create bias. As Hansen (2009) stated it, non-parametric methods must come with a data-dependent rules for determining the bandwidth but it is a difficult task and often the bandwidth is finally selected based on a statistical problem that is related to the length of the bandwidth.

We use the same approach as Barras et al. (2021) to determine the optimal bandwidth such that we take h^* that minimizes the Asymptotic Mean Integrated Squared Error (AMISE) of $\hat{\phi}(m)$. Thereby, we could control the trade-off between the bias and the variance exposed earlier.

Barras et al. (2021) formulated a proposition to examine the asymptotic properties of the estimated density, $\hat{\phi}$ as n and T grow large. They imposed that $n > T$ to derive these properties in order to capture the large cross-sectional dimension of the (bonds) mutual funds population.

Important remarks can be done and retained as content and consequences of their proposition (Barras et al., 2021). First, the estimated density, $\hat{\phi}(m)$, has different sources of bias that can be detected and for which we can adjust the estimated density to improve it. The first source of bias comes from the smoothing. Thereby, the first bias term adjusts for the bias coming from the difference between the actual density $\phi(m)$ and the smoothed density which can be written like this: $I_1 = \frac{1}{h}E[K(\frac{m_i - m}{h})] - \phi(m)$ where I_1 is representing the smoothing bias. Then, another bias term that can be deducted is the bias arising from the Error-In-Variable (EIV) bias, this bias can be written as $I_2 = \frac{1}{h}E[K(\frac{m_i + \eta_{i,T}/\sqrt{T} - m}{h})] - \frac{1}{h}E[K(\frac{m_i - m}{h})]$ where $\hat{\eta}_{i,T}/\sqrt{T}$ corresponds to the estimation error of $\hat{m}_i$ and I_2 is representing the EIV bias. This bias arises because of the noise we put in our estimation of $\phi(m)$ by using estimated measures $\hat{m}_i$ and not the true ones. Based on that, Barras et al. (2021) defined a bias term composed of 2 components, $bs_1(m)$ and $bs_2(m)$, in order to adjust the density. Thereby, following their approach, we can get a reliable estimated density if we have n and T that grow large. The bias term is the sum of its components, $bs(m) = bs_1(m) + bs_2(m)$ which accounts respectively for the smoothing and EIV biases.

However, a second important remark is that in the asymptotic case, the key driver of the bias is actually the EIV bias. Indeed, the latter depends on T the number of observations, so whatever the number of funds in the population, it will not disappear (we imposed that $T < n$). In contrast, the smoothing bias becomes negligible as we have n that is large, several thousands funds are enough to vanish this bias (Barras et al., 2021).

Third, it is important to know that "noisier estimation measures does not translate into a noisier estimation of the density $\phi(m)$" (Barras et al., 2021, p. 21). Thus, if EIV bias gets larger, it will not impact the variance of $\hat{\phi}(m)$. As a result, Barras et al. (2021) concluded that adjusting for the EIV bias enables to estimate $\phi(m)$ accurately in the asymptotic case.

Finally, a last remark regarding this proposition is that it has made it possible to obtain an expression for the optimal bandwidth h^*.

4.3.3 Adjusting for the biases

Barras et al. (2021) formulated a second proposition that enabled to construct the closed-forms of the 2 components of the bias and of the optimal bandwidth. To resume shortly, they applied a Gaussian reference model such that estimated measure m_i and the log of the asymptotic variance $s_i = log(S_i)$ follow a bivariate normal distribution: $m_i \sim \mathcal{N}(\mu_m, \sigma_m^2)$, $s_i \sim \mathcal{N}(\mu_s, \sigma_s^2)$ knowing that S_i is the asymptotic variance of the centered measure $\sqrt{T}(\hat{m}_i - m_i)$. This enabled them to compute good estimation of the bias that is closed to its true value.

As a result, by following their approach and computations, we could get the total bias (smoothing and EIV) in order to adjust the estimated density. This adjustment changes the shape of the estimated density $\hat{\phi}(m)$ in 2 ways. First, as the estimated measures, $\hat{m}_i$, tend to be larger because of the estimation noise, the tails of $\hat{\phi}(m)$ would be too big if we did not adjust for the bias. Thus, the bias-adjusted density will have lower tail probabilities. Then, this adjustment could be asymmetric as it could remove more mass from the left tail than from the right. This is explained because the correlation between the measure m_i and the *log* of the asymptotic variance s_i, $\rho = corr(m_i, s_i)$ could be positive. A positive ρ is found to be an empirical regularity for each measure $m_i \in \{a_i, b_i, va_i, va_i(s)\}$ (Barras et al., 2021). A fund with a positive measure m_i will exhibit a higher estimation variance S_i which implies that the estimated value $\hat{m}_i$ will be lower (left tail too thick).

Using the results of their second proposition, we computed the optimal bandwidth h^* and the bias-adjusted estimated density $\tilde{\phi}(m) = \hat{\phi}(m) - \hat{bs}_1^r(m) - \hat{bs}_2^r(m)$ where $\hat{bs}_1^r$ and $\hat{bs}_2^r$ are the approximations of the bias terms computed from the Gaussian reference model. With this bias-adjusted density, we could then compute the different characteristics of the distribution such as the moments, proportions and quantiles. The proportion of fund with a positive m_i will be given by the cdf estimate $\tilde{\pi}^+ = \int_0^\infty \tilde{\phi}(u)du$.

4.4 Empirical Results

4.4.1 Skill analysis

Thanks to the fund-level estimated values, $\hat{a}_i$, we could compute the bias-adjusted density $\tilde{\phi}(a)$ from which we can now draw some results in order to have more insight concerning the bond mutual funds' skill as of 2004 until now (skill measure of the relatively new bond mutual funds).

In the table below (Table 3), you can see the different characteristics of the skill distribution. It reveals that the monthly average of the skill coefficient is equal to 0.15% which corresponds to an annualized average of 1.85%. To analyze statistically this result, we ran a one-sided t-test[9] that says that we could reject the null at the 5% level meaning that the a_i come from a distribution with a mean that is significantly greater than 0 at this level. Moreover, the skill distribution is slightly positively skewed, which means that its right tail tends to be a bit larger than its left tail, so the probability of observing an extremely high skilled fund is higher than the probability of observing extremely low skilled funds. Some of our shape-related parameters are quite strongly consistent with the ones observed by Barras et al. (2021) in their research. Indeed, our distribution is positively skewed and exhibits a relatively high kurtosis as does their skill distribution except that our kurtosis is almost the double theirs, meaning that our skill distribution is a bit sharper than their skill distribution. You can find the bias-adjusted skill distribution in the Appendix 20.

Then, the most important observation we can make is that the bias-adjusted proportion of funds that are actually skilled ($a_i > 0$) is equal to 77.34% which is close to the rate of 76.6% of skilled managers from the results of Barras et al. (2010) that concerned US domestic-equity funds. Thus, the large majority of the bond mutual funds in the population is actually skilled which means that bond funds can find profitable investment ideas. These results also resonate with the findings of Barras et al. (2021) who found that about 87% of the equity mutual funds were skilled during the period from 1975 to 2019. Our research is different on 2 aspects: first, it concerns bond mutual funds and second, the targeted period is much more recent. As a result, before asserting that US bond mutual funds are less skilled than US equity mutual funds, we must be careful with the fact that the analyzed period is very different which makes the absolute comparison very difficult and irrelevant. Instead, we will rather draw the common conclusion that the majority of the US mutual funds seem actually skilled.

Finally, looking at the quantiles, we can discover that 5% of the funds in the population has a skill coefficient of less than −0.23% per month which corresponds to −2.75% per year. In contrast, for the highest skilled funds (5% of the population), they exhibit skill levels above

[9] By the Central Limit Theorem and our sample size of 2257 funds, we know we can use this test.

0.61% (7.58%) per month (per year) which is four times more than the average. We observe thus disparities across bond funds concerning their abilities to invest in profitable ideas.

	Skill coefficient
Mean (Monthly)	0.0015
Standard deviation (Monthly)	0.0039
Skewness	1.003
Kurtosis	38.89
Proportion Positive (\%)	77.34
Proportion Negative (\%)	22.66
(Monthly) Quantile 5%	−0.0023
(Monthly) Quantile 95%	0.0061

Table 3: Characteristics of the Skill distribution

4.4.2 Scale analysis

In the same way as the skill coefficient, we estimated the fund scale coefficient and then infer its bias-adjusted density with the non-parametric approach. The objective is to have more insights on the behaviour of the scale coefficient of our bond mutual funds population for the relatively recent period from July 2004 to December 2020.

Below, in the Table 4, you can observe the different characteristics of the scale distribution. We can see that on average, the bond mutual funds suffer from diseconomies of scale ($b_i > 0$ as we multiply the lagged size by -1) since the mean is positive and on average, the gross α decreases by 0.00051% (0.0061%) per month (per year) as the lagged size increases by $1,000,000[10]. Then, we can see that we have a slightly negative skewness but a very high positive kurtosis which means that the left tail is slightly longer than the right one but also that the concentration is highly around the mean as the bell is very sharp, so the tails are very small because of the high kurtosis. The distribution of the scale coefficient is thus not very dispersed. Indeed, the kurtosis of the scale distribution from Barras et al. (2021) is much smaller implying that our scale coefficients might be less dispersed and thus bond mutual funds might be more homogeneous with respect to their sensitivity to diseconomies of scale. However, as Barras et al. (2021) obtained these results based on standardized regression, the comparisons are very difficult as this could influence the distribution's characteristics. You can find the bias-adjusted scale distribution in the Appendix 21.

These results differ from the ones of Barras et al. (2021) in several other aspects but as already mentioned, the distributions characteristics are not really comparable as Barras et al. (2021) used standardized regression to interpret the scale coefficient and we did not in order to have the true impact of the size on the gross alpha. Moreover, since we do not compare the predictors coefficients with each other but only the intercept (a_i) and the scale coefficient (b_i), the standardization did not make sense to us. However, we can still make some comparisons as long as the scaling is not concerned, for example, the proportions of positive and negative should not differ too much between standardized and non-standardized coefficients. As a reminder, this also cannot be interpreted as a fair comparison between bond mutual funds and equity mutual funds as the periods analyzed are different, so any difference could be due either to the different types of assets studied (bond vs equity) or to the different periods.

[10] $1M in terms of the base 2000 as we control the sizes for inflation.

With this in mind, we can see that about 72% of the population experience real diseconomies of scale ($b_i > 0$). This turns out to be a bit less than the results of Barras et al. (2021) where this proportion reached 82.4% but we are still in agreement with many theories on mutual funds which affirm that the diseconomies of scale in the mutual fund industry is big. Indeed, it appears from our research that a large majority of funds are still experiencing diseconomies of scale but also that about 27% apparently experience economies of scale against 17.6% for Barras et al. (2021). Barras et al. (2021) classified the left funds (that had significant $\hat{b}_i < 0$ (at the 5% level)) as false discoveries as the true b_i was actually equal to 0. Indeed, they explained that the EIV bias adjustment does not enable to control perfectly for the estimation noise with a cluster of values for $\hat{b}_i$ around 0. In order to investigate in this direction, we choose to run a one-tailed t-test at the 5% level to identify the significantly negative b_i among the estimated coefficients. We found that only 184 b_i values were significantly negative at the 5% level which represent only 8.15% of the 2257 estimated b_i. Assuming that the Type I error is at 5% by the level of the test, we can suppose that among these 184 negative b_i, 9 are actually equal to 0. As a result, we still observe a very low proportion of funds that seem to exhibit economies of scale but this could be due to some estimation noise that could not be totally neutralized by the EIV bias as Barras et al. (2021) also supposed it.

Then, we could also compare the impact of a $\$100M$ increase in size on the gross alpha, for us this results in a decrease of 0.051% in the gross alpha per month which is equivalent to a decrease of 0.61% per year. In contrast, this same increase, lowered the gross alpha by 0.2% per year in the research of Barras et al. (2021). As a conclusion, we can say that either bond mutual funds are more sensitive to diseconomies of scale than equity mutual funds or that recently (during the period 2004-2020) mutual funds are more sensitive to diseconomies of scale than in the period from 1975 to 2019.

Finally, regarding the quantiles, we can observe that 5% of the funds suffer from levels above 0.00699% for their scale coefficients. That implies that some funds have large sensitivity to diseconomies of scale compared to the mean (around 14 times the average) although the dispersion in the distribution is rather limited. Indeed, the tails are small but go far. This confirms the choice of constructing a fund specific scale coefficient instead of imposing a constant b across all funds.

	Scale coefficient
Mean	$5.10e-06$
Standard deviation	$3.94e-05$
Skewness	-1.60
Kurtosis	67.20
Proportion Positive (%)	72.55
Proportion Negative (%)	27.45
Quantile 5%	$-2.20e-05$
Quantile 95%	$6.99e-05$

Table 4: Characteristics of the Scale distribution

4.4.3 Correlation between skill and scale coefficients

Following some insight from Barras et al. (2021), we decided to look at the correlation coefficient between the estimated skill and scale coefficients. As might be expected, they are positively correlated as the pairwise correlation between $\hat{a}_i$ and $\hat{b}_i$ is equal to 0.5732. This means that the bond mutual funds that exhibit the best skill (high a_i) also exhibit a quite high level of diseconomies of scale. In other words, the best skilled funds will have difficulties to grow in size as they want to maintain a positive and high gross α. As explained and illustrated by Barras et al. (2021), this correlation is partially linked to the fund's investment style. Indeed, as a_i and b_i depend both on the characteristics of the fund strategy, it is quite logical that they are correlated.

What we can learn from this correlation is that the most profitable funds may not be the ones which exhibit the highest skill as their sensitivity to diseconomies of scale would also be very high and thus would rapidly decrease the gross α. Alternatively, the most profitable funds may actually be the ones which exhibit a more balanced pair of coefficients as it will be discussed in the next section which introduces the value added measure.

4.4.4 Value added analysis

Finally, we can get the cross-sectional distribution of the value added computed based on a_i and b_i. This distribution concerns the value added that the bond mutual funds create during their entire lifetime from July 2004 to December 2020. Again, we put the different characteristics of the distribution $\bar{\phi}(va)$ in a table below (Table 5).

	Value added
Mean (in Millions $)	0.3818
Standard deviation (in Millions $)	1.05
Skewness	1.93
Kurtosis	24.20
Proportion Positive (\%)	82.69
Proportion Negative (\%)	17.31
(In Millions $) Quantile 5%	−0.2425
(In Millions $) Quantile 95%	2.33

Table 5: **Characteristics of the Value added distribution.** The value added measures are expressed in $M in terms of the base year 2000.

As you can see, bond mutual funds create value in the marketplace as the average value created over the studied period is equal to almost $\$382K$ which is equivalent to a value added of about $\$24.6K$ per year from all funds in the population. Comparing to the results of Barras et al. (2021), bond mutual funds do not create much value in the market because equity mutual funds create on average $1.9M of value per year. Then, we can see that almost 83% of the bond mutual funds exhibit a positive value added, so a huge majority of the bond mutual funds from 2004 to 2020 create value. This is more than the findings of Barras et al. (2021) who found that 60% of the equity mutual funds had a positive value added over their studied period. However, as already explained, we cannot use these results to compare bond and equity mutual funds as the analyzed periods are different and also play a role in the results. Indeed, the value added depends on the average of the lagged size and the average of the squared lagged size over the period studied and since the lagged size varies over time, the periods analyzed can influence the results. Therefore, as our studied period is more recent, this higher proportion of funds that create value could be due to the learning process from which investors benefit during the period 1975-2004 and not necessarily because more bond mutual funds can extract value from the market. Precisely, as explained by Barras et al. (2021), the value added in the early stages of the mutual funds' life tends to be smaller as it takes time to investors to learn about skill and scalability and allocate the right amount of capital to each fund. Thereby, one hypothesis to explain this higher proportion could be that in our study we benefit from the investors' learning process without including the consequences of this period of process in our computations (or at least less consequences for funds that were created well before 2004), what could inflate the results compared to the ones of Barras et al. (2021).

Furthermore, we cannot miss the big gap that exists between the 5% and the 95% quantiles. Indeed, we still learn from Table 5 that 5% of the funds in the population destroy more than $\$242K$ of value unlike the most valuable funds (5%) which create more than $\$2M$ of value.

Then, from the plot below (Figure 5), you can see the effect the EIV bias adjustment had had on the value added density. As you can notice, the proportion of lower values would have been greater without the adjustment as the left tail of the unadjusted density is thicker, so the proportion of funds that destroy value would have wrongly been higher. Moreover, this adjustment also makes it possible to correctly detect the funds that create more value than the average as it would have been largely underestimated without it.

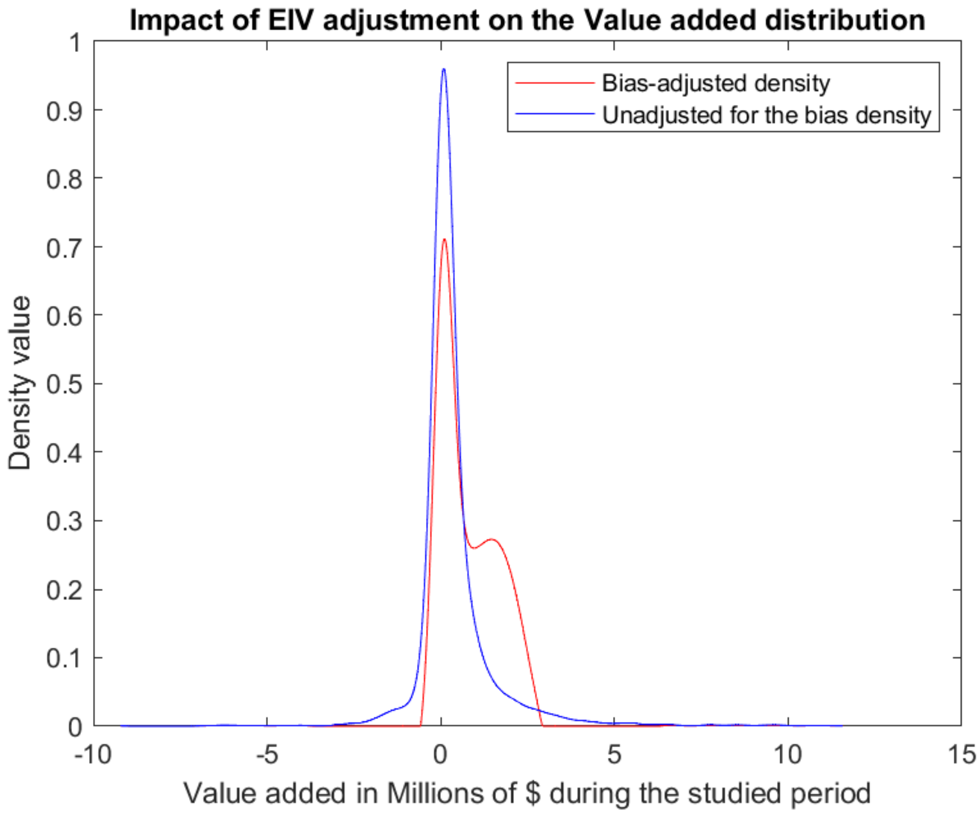

Figure 5: **Bias-adjusted and unadjusted densities of the Value added.** The value added measures are given in \$M in terms of the base year 2000.

Although there is a small proportion of funds that destroy value, we can still questioning what makes that these funds cannot create value. From our computations of the value added $va = a_i \times E[q_{i,t-1}] - b_i \times E[q_{i,t-1}^2]$, we know that the funds with a negative value added are either unskilled ($a_i < 0$) or are too big ($0 < a_i$ but $a_i \times E[q_{i,t-1}] < b_i \times E[q_{i,t-1}^2]$). By combining results from our skill analysis and value added, we discover that the funds population seems to exhibit more unskilled funds (22.66%) than there exist non-performing funds (17.31%). This supposes that some unskilled funds could still create value and go in the sense of some existing economies of scale. However, this may concern a very tiny proportion of funds that could arise because of some estimation noise that the EIV bias could reduce but not make it equal to 0, as already mentioned earlier (Barras et al., 2021). In view of the inconsistency of this result with the model, we will investigate deeper the fund size impact in the following sections in order to see if other results affirm or deny the hypothesis that funds could create value if they decreased their size. Indeed, several reasons could explain this abnormal result and we will confirm it to be an anomaly later.

4.4.5 Last Subperiod Value Added

As already mentioned, the previous section concerns the value added that the funds create over their entire lifetime from 2004 to 2020. However, as the lagged fund's size varies over time, it can be influenced by the start-up period of the fund and thus does not reflect well the value added of the funds when they get older. As a result, we decided to create 5 subperiods ($S = 5$) by splitting the total number of observations of each fund. Then, in order to have more insights concerning the dynamics of the lagged fund's size during our period of research, we computed $\Delta q_{i,s} = \bar{q}_{i,s} - \bar{q}_i$ where $\bar{q}_{i,s}$ is the average over subperiod s and $\bar{q}_i$ is the average over the full sample. We plot the median value of $\Delta q_{i,s}$ for each subperiod and as you can see in Figure 6 funds' sizes varied a lot during our period of analysis. As the median of the $\Delta q_{i,s}$ is negative, we know that the funds' sizes during this subperiod s are lower than their average. For the first subperiod ($s = 1$), we can see that the funds' sizes during this subperiod are well under their averages, the median gap is equal to $\$-29M$. On the other hand, we can see that for subperiods 3 and 4, the funds' sizes are above their averages and were particularly close to their average in subperiod 5. A maximum is reached in subperiod 4 before falling back towards the average. This graph supposes that bond mutual funds become bigger and bigger over their lifetime before reducing their sizes as they get older.

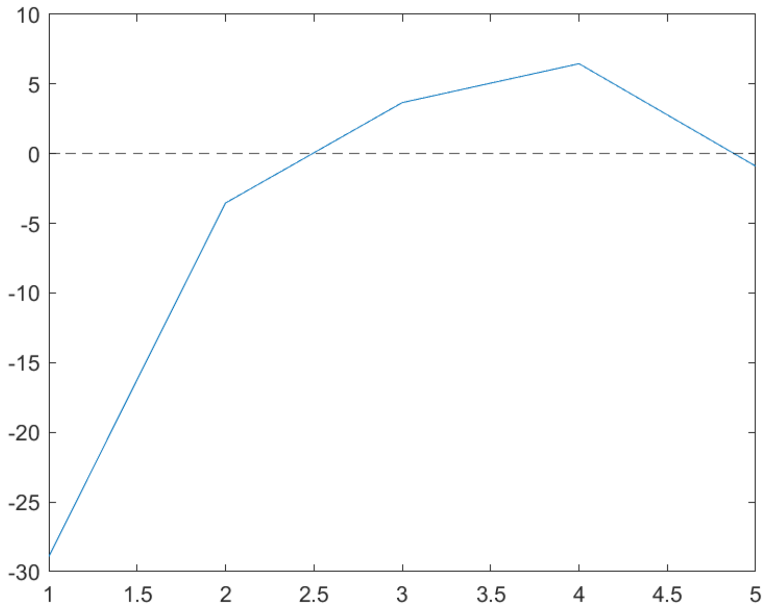

Figure 6: **Dynamics of fund size over time from subperiod 1 to 5.** The sizes are given in $M in terms os te base year 2000.

Motivated by the results of Barras et al. (2021), we decided to analyze the last subperiod ($s = 5$) in order to have more insights over the value added of the funds when they get older. To that end, we have computed the last subperiod value added density, $\phi(va(5))$ with the same non-parametric approach as previously. In the table below (Table 6), you can observe the characteristics of this distribution as for the precedents. Although the proportions of positive and negative value added did not change a lot, the average value added extracted from the market by all funds over the last subperiod is equal to $\$291K$ which is more or less equivalent to a value of $\$94K$ per year (vs $\$24.6K$ per year for the entire period mean).

	Subperiod Value Added
Mean (in Millions \$)	0.2916
Standard deviation	1.06
Skewness	1.07
Kurtosis	45.03
Proportion Positive (\%)	85.40
Proportion Negative (\%)	14.60
(In Millions \$) Quantile 5%	-0.1199
(In Millions \$) Quantile 95%	2.18

Table 6: **Characteristics of the last Subperiod Value Added distribution.** The value added measure are given in $M in terms of the base year 2000.

From these results, we can draw several conclusions: first unlike Barras et al. (2021), the proportion of funds in the population that create value over the entire analyzed period from 2004 to 2020 does not differ much from the proportion of funds that create value during their last subperiod (respectively 83% vs 85%). However, since the yearly average amount is largely higher, we can assume that these funds which create value, create on average more value during their last subperiod. So, the funds are getting more performing when getting older. This lets suppose that a learning process has still to be done but the funds might have benefited from the investors learning process about which Barras et al. (2021) talk. Consequently, the proportions of performing funds during the entire period or the last subperiod are the same, so we observe almost no fund that had negative value added during the entire period and then positive value added during the last subperiod, i.e. after reducing their sizes. This is another lead towards the previously drawn hypothesis that our higher proportion of funds that create value compared to the proportion obtained by Barras et al. (2021) is more due to the different analyzed periods than to the difference of performance between the 2 categories of funds. Another lead could also be that the increase in size that the funds have experienced was sufficient to decrease the average value added over the entire period but not enough to make it become negative (this would be more linked to the kind of assets managed by the funds as the bond funds sizes increase is more moderate). However, as mentioned before, our results show superior sensitivity to diseconomies of scale, so this hypothesis might be a bit less likely.

Then, a second conclusion that we can draw from this Table 6, is that this average value added created by the funds over their last subperiod is greater than the average value added created by the funds over the entire period (relatively to one year). This means that, while no additional funds create value, the funds create even more value added when they are getting older, we might therefore think that in the relatively recent period from July 2004 to December 2020, the funds benefit from the investors' learning process and the majority can create value but by getting older they can manage their size even better in order to create even more value added. To confirm this intuition, we can look at the shape-related characteristics of the distribution and see that the kurtosis is not far from twice the kurtosis of the entire period value added distribution. This means that this distribution is sharper and that there is more concentration around the average. You can observe the difference between these 2 distributions with Figure 7 below, you can see that the last subperiod value added distribution has a concentration at the mean much more important and also that the proportion of funds that create value added a bit larger than the average is higher. Furthermore, the tails of the last subperiod added distribution both go further than the tails of the entire period value added distribution. However the positive tail is even longer, meaning that some extreme values could positively influence the average.

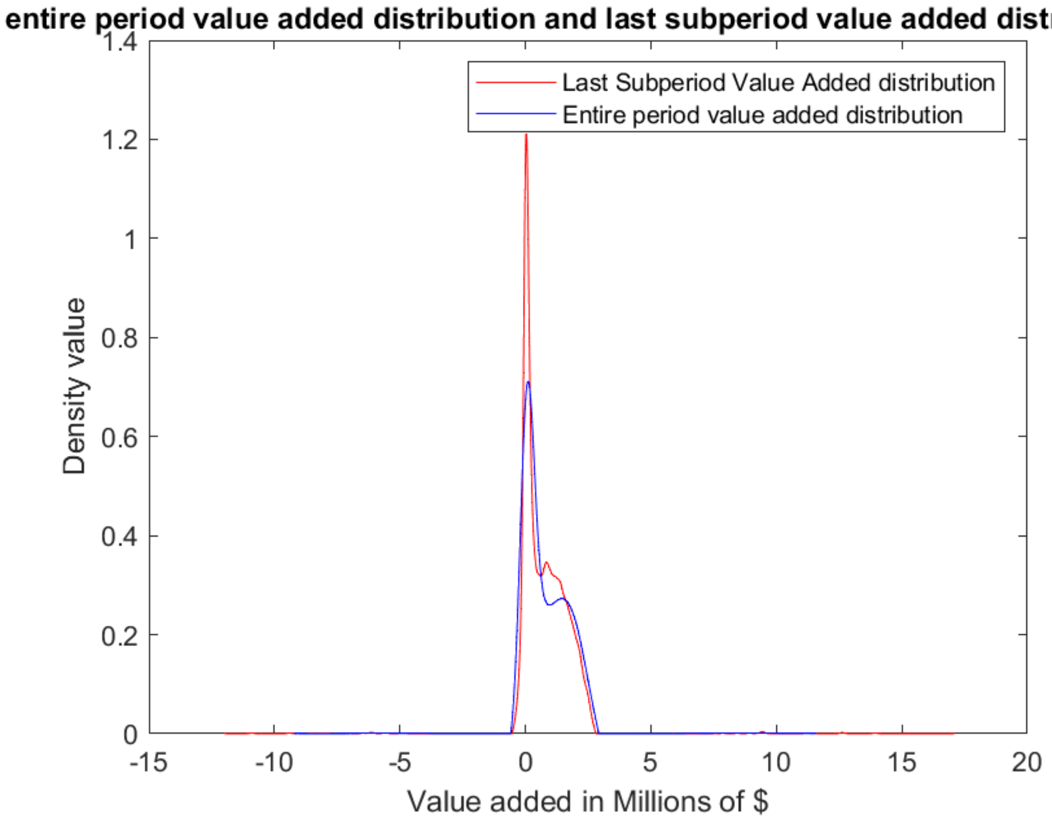

Figure 7: **Comparison between the last subperiod value added and the entire period value added distributions.** The value added measures are given in \$M in terms of the base year 2000.

4.4.6 Best combination of the coefficients

In this section, we decided to analyze deeper the value creation process in order to determine which funds can create the most value. Indeed, as the skill and scale coefficients are positively correlated, we know that the most skilled funds could not be the funds that create the most value as their good investments ideas are expected to be very difficult to scale up due to this correlation (high a_i means that b_i is high too). To examine this issue, we decided to compute the 10 deciles of each coefficient and then stretch each of them on a scale from 1=lowest to 10=highest. Next, we search after the funds having highly positive and significant estimated value added, $\hat{va}_i$. For that selection, we run a one-sided test with $H_0 : \hat{va}_i = 0$ and took the funds from the right tail that rejected the null with a 5%-significance level. In other words, we took the funds for which the t-stat was above the threshold (only the positive one), equal to 1.6513, in the case of a Student with 187 degrees of freedom[11]. We computed the $\hat{t}_i^{va}$ as $\hat{t}_i^{va} = \frac{\hat{va}_i - 0}{\sigma_i^{va}}$ where the standard deviation is computed with the estimated asymptotic variance $\hat{S}_i$ such that $\sigma_i^{va} = \sqrt{\frac{1}{T}\hat{S}_i}$. As a result, we ended up with 948 funds.

[11]Student distribution with a degree of freedom as high is very close to a normal distribution

Next, we found that the median skill and scale scores of these funds are respectively 5 and 3. Moreover, the proportion of these funds that actually exhibit the best score for skill (10) and scale (1) are respectively 8% and 14.6%, so the majority of the best performing funds have neither the best investment ideas nor the lowest scalability cost. Instead, the best performing funds actually seem to have a balance between skill and scalability. In other words, the best performing funds are those with less great investment ideas but that can scale up more easily as they have lower scalability cost.

These results are consistent with the ones of Barras et al. (2021) who found that the best performing funds had skill and scale scores of 7 and 4. We can therefore conclude that these results are valid for the mutual fund industry in any period of time.

4.5 Additional Results

4.5.1 Comparisons with the optimal value added

As we found earlier that eventually, all the funds that may be destroying value are unskilled, supposing that no funds got negative value added because they were too big. We wanted to deep a bit this hypothesis by trying to have insights regarding the size management of the bond mutual funds. To that end, we determine the optimal size, q_i^*, the fund should have to maximize its value added. In the same idea, this additional analysis could give us an overview of how far/close from the optimal value added funds are with their real sizes. To run this analysis, we base on several assumptions from the Berk and Green model and their equilibrium predictions (Barras et al., 2021). Thereby, the assumptions are first that the number of skilled funds is limited for supply; second that numerous investors compete for performance. By following these assumptions, the funds have power over the investors and can maximize their profit by charging investors fees $f_{e,i}$ equal to the gross alpha α_i (supply and demand law which says that as demand is high prices increase). As a result, the profit is equal to the value added since $\pi_i = f_{e,i} \times q_i = \alpha_i \times q_i = va_i$.

Then, from the linear specification $\alpha_i = a_i - b_i q_i$, and by taking the first order derivative $\dfrac{\partial va_i}{\partial q_i} = 0$, we know the expression of the optimal size $q_i^* = \dfrac{a_i}{2b_i}$ for the value added to be maximized (Barras et al., 2021). We know that the va_i function is concave as its second order derivative $\dfrac{\partial^2 va_i}{\partial^2 q_i} = -2b_i < 0$, so we also know that this maximization gives well a maximum value and not a minimum. By replacing q_i in the expression of the value added, we get:

$$va_i^* = a_i q_i^* - b_i (q_i^*)^2 = \frac{a_i^2}{4b_i}, \tag{4.13}$$

so we can compute the optimal value added and then compare it to the actual value added we computed earlier.

Of course, our comparison requires some prior precaution as checking that the va^* is positive. To achieve this, we decided to run an hypothesis test and to select only the funds which rejected the null. This test aims at finding the funds which exhibit highly significant and positive estimated values $\hat{va}_i^*$. As a result this is a one-tailed test for which the null is:

$$H_{i,0} : va_i^* = 0. \tag{4.14}$$

First thing to do to run this test is to construct the t-statistic in the same way we did before for the highly significant and positive actual value added. So $\hat{t}_i = va_i^*/\sqrt{\frac{1}{T}\hat{S}_i}$ is our t-statistic for the test with the asymptotic variance computed following the same approach as for the measures (following Barras et al. (2021)). Then, we could select all the funds that rejected the null, so the funds for which the $\hat{t}_i$ was above the threshold which is the 1, 5 or 10% t-distribution quantile (with 187 degrees of freedom). Indeed, we run this test using several significance levels such that the proportion of Type I errors[12] in the selected funds in each case is equal to the level of the test and we will be able to analyze how many false discoveries slipped in the selected funds when we increase the level of the test. In the table 4.13 here under, we report the results for the average optimal value added of the selected funds as well as the actual value added (entire period and last subperiod) that the very same funds actually exhibit.

	1% level		**5% level**		**10% level**	
	Mean	Ratio	Mean	Ratio	Mean	Ratio
Optimal value added (in Mio, per year)	0.2451		0.1955		0.1862	
Actual value added						
Entire period value added (in Mio, per year)	0.1	41.04%	0.0829	42.41%	0.0768	41.23%
Last subperiod value added (in Mio, per year)	0.1828	74.57%	0.1898	97%	0.1781	95.64%
Trimming the 10% smallest and highest values of a and b						
Optimal value added (in Mio, per year)	0.25		0.1990		0.1853	
Actual value added						
Entire period value added (in Mio, per year)	0.1029	41.11%	0.0845	42.47%	0.0784	42.33%
Subperiod value added (in Mio, per year)	0.1863	74.43%	0.1930	96.98%	0.1810	97.67%

Table 7: **Table of the averages of the optimal value added and actual value added of the selected funds before and after trimming the highest and lowest values for $\hat{a}_i$ and $\hat{b}_i$.** The value added measures are given in $M in terms of the base year 2000.

As it can be seen, funds capture between 41 and 43% of their optimal value added which is consistent with the fact that funds could maximize their value added if they were paying more attention to the management of their sizes. Additionally, we know that these results are not driven by high/low estimated a_i or b_i as after trimming 10% of the highest and lowest values for these coefficients, we still get very close results. Furthermore, as we want to have more insights regarding the actual size with respect to the optimal size of the funds, we decided to look at the difference between the optimal size and the average of the actual size for each fund selected at the 5% level. As we can observe it on the graph below (Figure 8), on average, some funds are indeed too big compared to their optimal size given by $q_i^* = \frac{a_i}{2b_i}$ as the difference $(q_i^* - \bar{q}_i)$ is negative but we learn that some funds are also too small on average at the equilibrium since this difference is also highly positive, so bond mutual funds should manage their sizes in 2 ways: (i) They have to grow until their optimal size, which is probably the most obvious goal of each fund, they intend all to grow but (ii) they have also to manage their growth in order not to be higher than their optimal size. Indeed, as they grow, the funds should try to have on average a size close to their optimal size and not higher as it will decrease their value added on the market. However, the majority of the funds still tend to be too big as 70.19% of the selected funds have an average actual size higher than their optimal size.

[12] Type I error is the fact to reject the null whereas it is actually true, so that is the funds that apparently reject the null whereas their $va_i^* = 0$

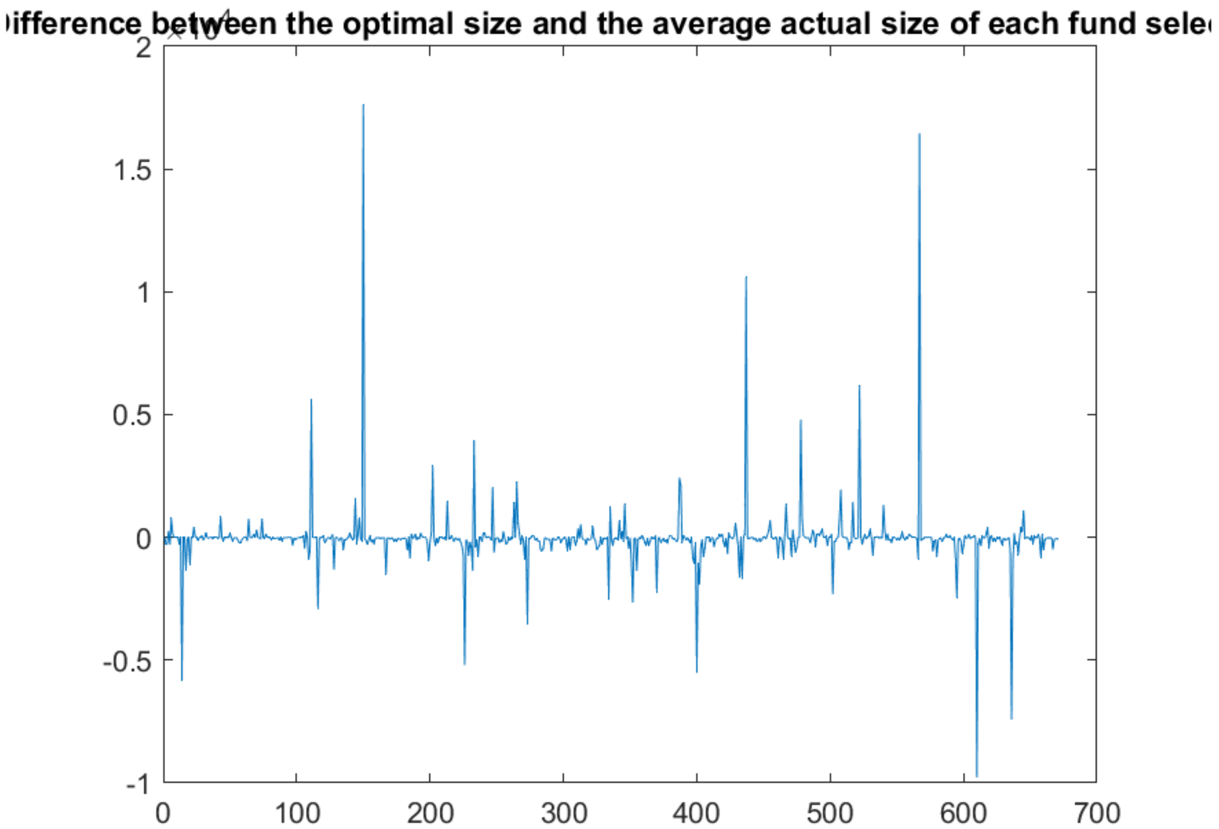

Figure 8: **Difference between q_i^* and $\bar{q}_i$ (in Mio) for the selected funds at the 5% level.** The sizes are given in \$M in terms of the base year 2000.

Then, another previous hypothesis that we would like to clarify is the fact that some funds that destroy value ($\hat{va}_i < 0$) may be too big. Otherwise, the previous result supposed that the only reason why these funds destroy value is because they are actually unskilled. As a result, we will take the funds that destroy value among the selected funds in order to have insights even if this result will not be over the whole sample. As a reminder, the funds are selected as their estimated optimal value added $\hat{va}_i^*$ is significantly positive at the 5% level, so funds with actual negative value added among these funds is possible. Below on the graph (Figure 9), we can observe the actual value added in Mio[13] of these funds. We can see that several funds exhibit negative value added. In order to have insights regarding these non-performing funds and their position with respect to their optimal size, we construct a graph (Appendix 22) representing the difference $q_i^* - \bar{q}_i$ (as before) with the non-performing funds depicted as vertical dotted lines. Surprisingly, for most of the dotted lines ($va_i < 0$), the difference between the optimal size and the actual size of the selected fund is negative which cannot be a coincidence. This goes in the sense of the hypothesis of Barras et al. (2021) saying that maybe these funds could create value if they decreased their size. Thereby, the previous result implying that the reason why some funds destroy value is only because they are unskilled is probably driven by some estimation noise that either increases the proportion of unskilled funds in the population or decreases the proportion of funds that destroy value. For example, some funds may exhibit $a_i = 0$ that are classified as negative by estimation noise. Additionally, we also computed the difference $q_i^* - q_i$ for the other significance levels and you can find the median of these differences for each level in the table here under (Table 8). As you can see, the median at each level is negative, so lots of funds have well an actual size too big.

[13] In terms of the base year 2000.

	1% level	5% level	10% level
Median of the difference between actual size and optimal size (in Mio)	-22.24	-26.27	-24.32

Table 8: **Table of the medians of the difference between the selected fund's optimal size and its average actual size.** The sizes are given in \$M in terms of the base year 2000.

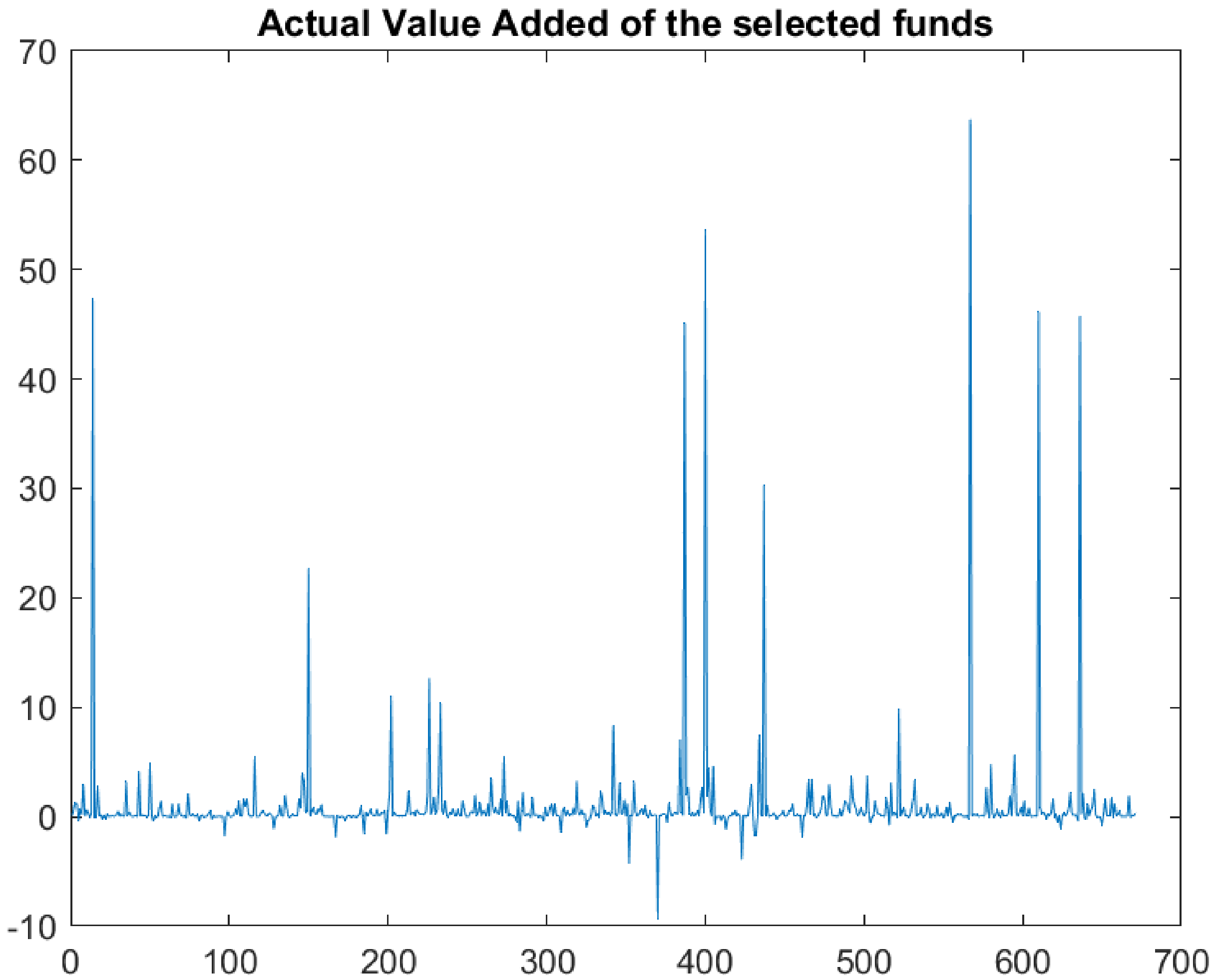

Figure 9: **Actual value added of the selected funds at the 5% level.** The value added measures are given in \$M in terms of the base year, 2000.

Finally, the last aspect we would like to analyze is the subperiod value added and the actual size dynamics with respect to the optimal value added and the optimal size. First, from the previous table 7, we can see that the last subperiod value added is much closer to the optimal value added, let us think that funds learn during their lifetime and might optimize their size as time goes.As a result, we computed the difference between the fund's optimal size and the average actual size for each subperiod $s = (1, ..., 5)$, $\tilde{q}_i(s)$. We computed these differences for the 5% significance level and then for each subperiod we took the median that you can see represented here under on the graph (Figure 10). As you can see, at the first subperiod, the actual size of the funds seems to be lower than the optimal and then it grows and in the last subperiod the actual size is on average higher than the optimal size whereas the value added in the last subperiod seems to be higher and thus closer to the optimal value added. A remark that we can make is that, it is not because the actual size of the fund is higher than the optimal size that the fund fails to optimize its value added as long as the funds in excess of the optimal size is invested passively. Indeed, this hypothesis is supported by the Figure in Appendix 22 which shows that although almost each dotted line ($va_i < 0$) seems to exhibit an actual average size greater than the optimal size, the inverse is not true. So each fund that has a negative difference between q^* and $\bar{q}_i$ has not obviously a negative va_i, i.e. if they invest the excess funds passively, they can still maximize their value added.

Furthermore, our results are consistent with the ones of Barras et al. (2021) as their last subperiod average actual sizes are also on average higher than the optimal size but still the last subperiod value added was closer to the optimal va_i^*. These observations drive to the outcome that funds might learn during their lifetime and get closer to their optimal value added if not at their optimal value added. Indeed, during the last subperiod, selected funds at the 5% and 10% levels seem to reach their optimal value added. One remark has to be made here regarding the potential false discoveries that were incorporated as we increased the level of the test (from 1% to 5%). When the level equals 1%, 389 funds are selected for having an optimal value added significantly positive, among them, only 4 funds can be considered as false discoveries (1%×389). Then, as we increase the level to 5%, 658 funds are selected (62% more) and among them 33 funds can be considered as false discoveries (8× more). As a result, the optimal value added average as the level increases can be underestimated because from the 240 additional funds that we discover for having a significantly positive optimal value added, 12% (29/240%) have actually an optimal value added equal to 0. This reduces the optimal value added average, which can explain the difference between the ratios from the 1% level and from the 5% level.

As a conclusion, we can say that since the last subperiod value added of the funds is closer to the optimal value added than the value added of the entire period, the funds probably learn and invest partially the excess of funds passively. So by getting older, bond mutual funds can become more performing and get closer to their optimal value added.

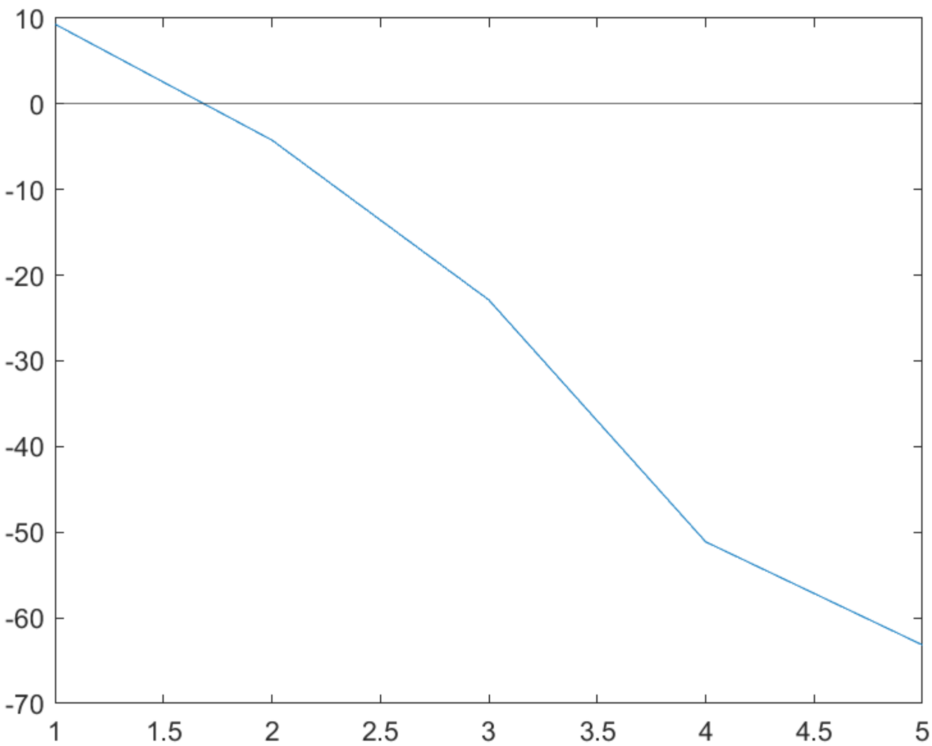

Figure 10: **Medians of the differences $q_i^* - \bar{q}_i(s)$ for the funds selected at the 5% level during each subperiod.** The sizes are given in \$M in terms of the base year 2000.

5 Conclusions

We have shown that during the period from July 2004 to December 2020, US bond mutual funds create value in the market although they are too large to achieve their optimal value added amount. Furthermore, our results support the popular statement that mutual funds are experiencing diseconomies of scale and more importantly we have provided a different measure to assess the managerial skills of bond mutual funds as it is expressed by one part of the gross α and not the value as a whole since we consider that the managerial skills can be good (in accordance with the funds managers' remuneration) but reduced due to the size of the funds. Based on our measurement, we can thus affirm that bond mutual funds are skilled because first, the distribution of the a_i has an average significantly above 0 at the 5% level and second based on the bias adjusted distribution of these a_i, we found that 88% of the funds exhibit a positive a_i.

This means that based on gross returns (before any fees and expense), funds appear to be skilled and create value. However, as it has also been demonstrated, the best performing funds are funds that have mitigated values for a_i and b_i because these coefficients are positively correlated. So unfortunately, the best investments ideas (highest a_i) do not offer the best performance (highest value added). These measures of performance can be interesting because they provide insights into managerial skills and the ability of the funds to manage their size and investments to create value. However, as previously mentioned, these measures might not be sufficient for investors who want to invest in bond mutual funds. Indeed, investors may be more interested in measures based on net returns as they are more related to the actual return they will get. It could be an idea to go further on this research thesis (doing this work on net returns).

Then, we could also open the discussion about the robustness of our results. Indeed, we should check the impact of the choice of the asset pricing model on our results. To get on overview of this impact we performed the procedure using the model 4 (introduced in Section 4.1). We found that our results may not be as robust as we might hope relatively to the asset pricing model choice. Indeed, although the proportions of performing funds ($va_i > 0$) still held for both the entire period and the last subperiod, the average value added drastically increased to \$1.9M for the whole period with the factors of model 4. Moreover, the averages of the skill and scale coefficients increased as well as the proportions of positive a_i and b_i (they both reached values close to 90%). So, even if the conclusions may not differ drastically, the results from each model still vary little. Indeed, we can confirm that the previous weird result implying that some economies of scale could occur, does not appear here since based on the model 4, we find that 85% of the funds that destroy value are unskilled[14], which support well the statement that a proportion of funds could create value if they decreased their size as we concluded. However, the distributions' shapes remained close to each other except for the means and standard deviation, so the estimation procedure does not give totally contrasted results from one model to another.

Finally, another point that could also be interesting to study based on this approach is the $\alpha_{i,t}$ distributions for each fund. Indeed, as we have α_i over time, we could have analyzed them and used the work of Barras et al. (2010) to obtain the actual percentage of funds that are performing based on the $\alpha_{i,t}$ distributions (another measure of performance that is the most used in literature) and see if the results are very different from the results obtained while

[14]We find that 13.61% of funds were unskilled and that 15.96% were non-performing, so 1-(13.61/15.96)% destroy value because they are too big.

analyzing the value added measure. Unfortunately, this was a bit out of scope for our research thesis but could be an interesting point to go further.

A Bond Asset Pricing Results

A.1 Coefficient Distributions from the linear regression of Equation 4.2 with Model 2 variables

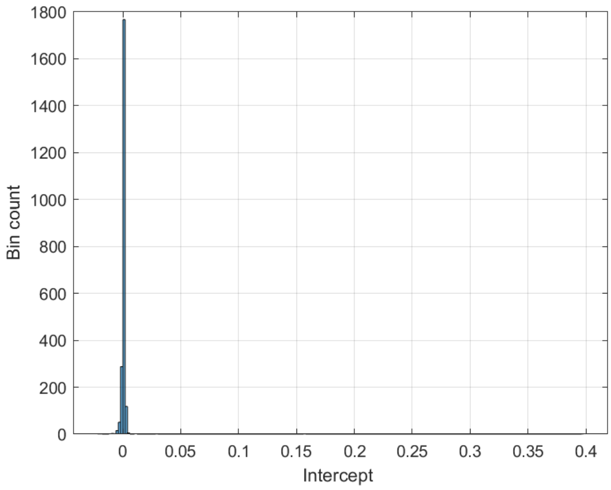

Figure 11: Intercept Distribution

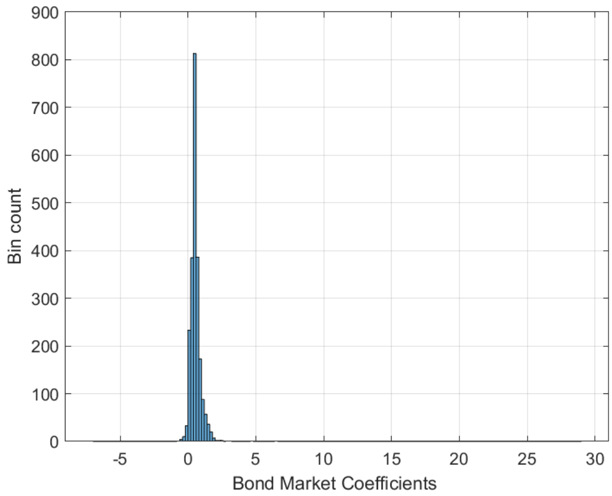

Figure 12: Bond Market Coefficients distribution

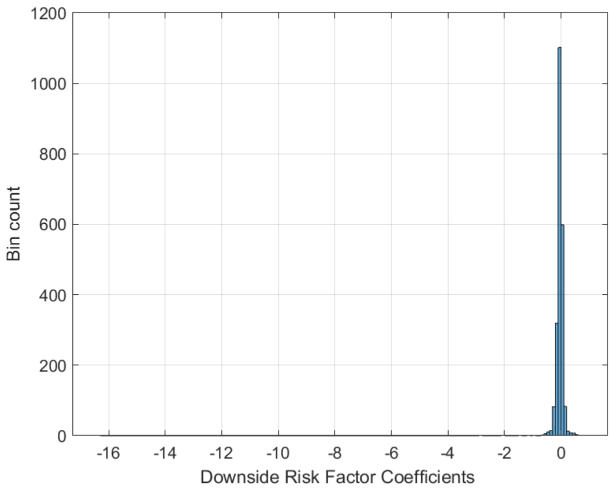

Figure 13: Downside Risk Factor Coefficients Distribution

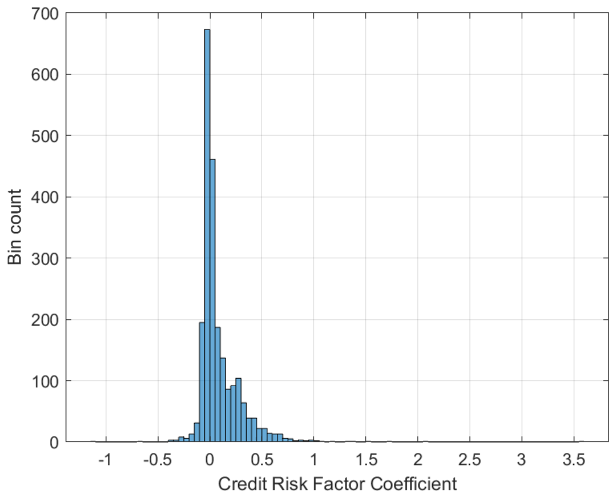

Figure 14: Credit Risk Factor Coefficients Distribution

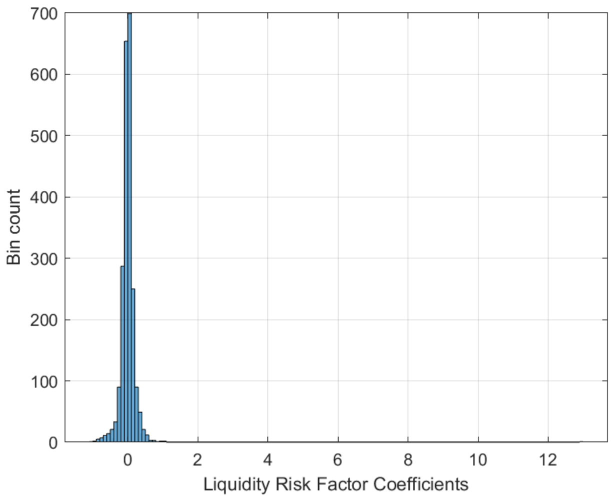

Figure 15: Liquidity Risk Factor Coefficients Distribution

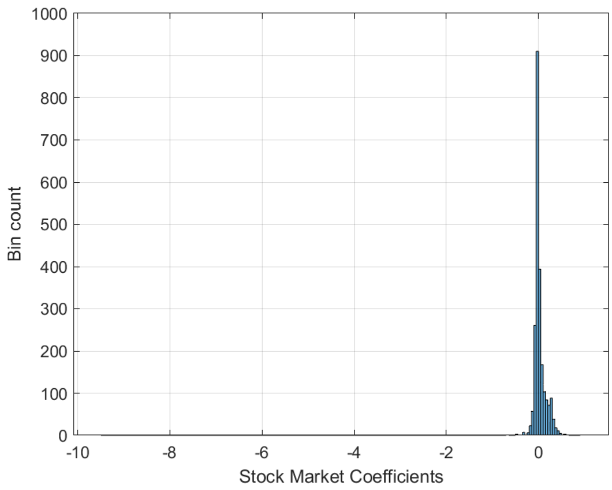

Figure 16: Stock Market Coefficients Distribution

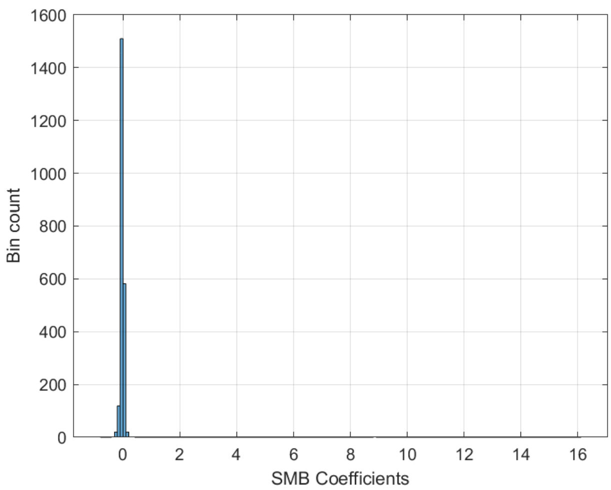

Figure 17: SMB Coefficients Distribution

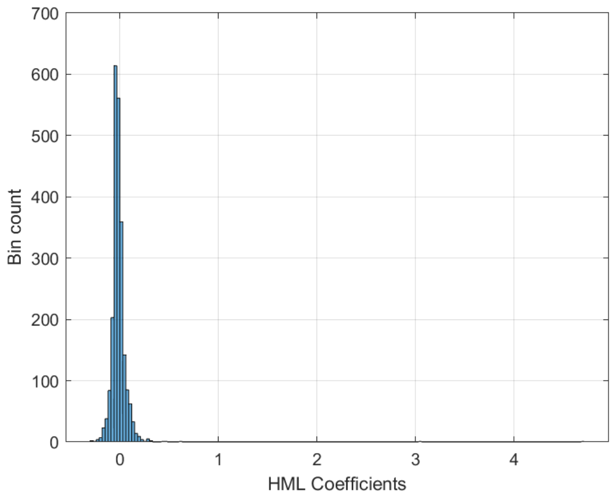

Figure 18: HML Coefficients Distribution

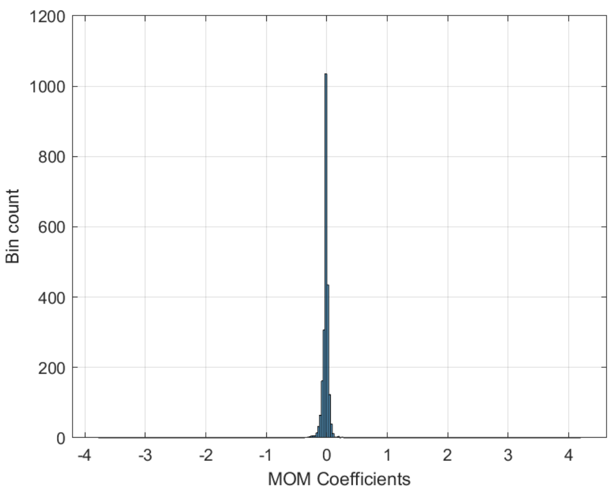

Figure 19: MOM Coefficients Distribution

B Skill and Scale distributions

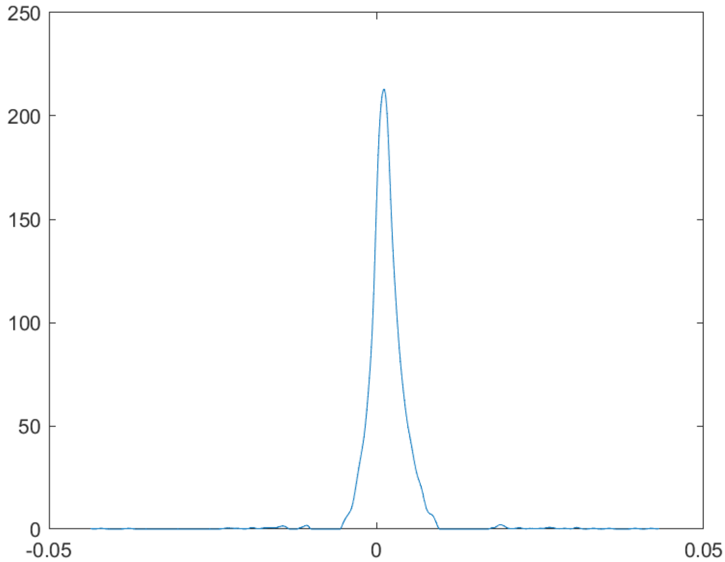

Figure 20: Bias-adjusted skill distribution, $\tilde{\phi}(a)$

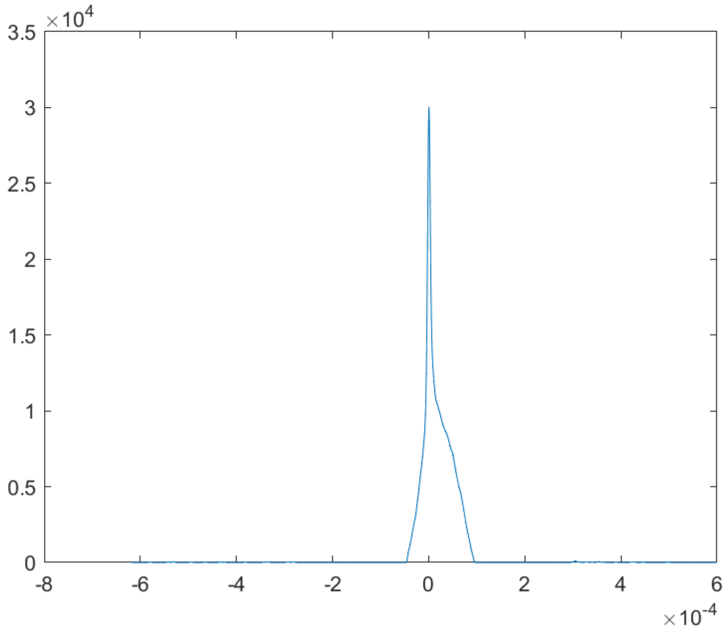

Figure 21: Bias-adjusted scale distribution, $\tilde{\phi}(b)$

C Optimal value added

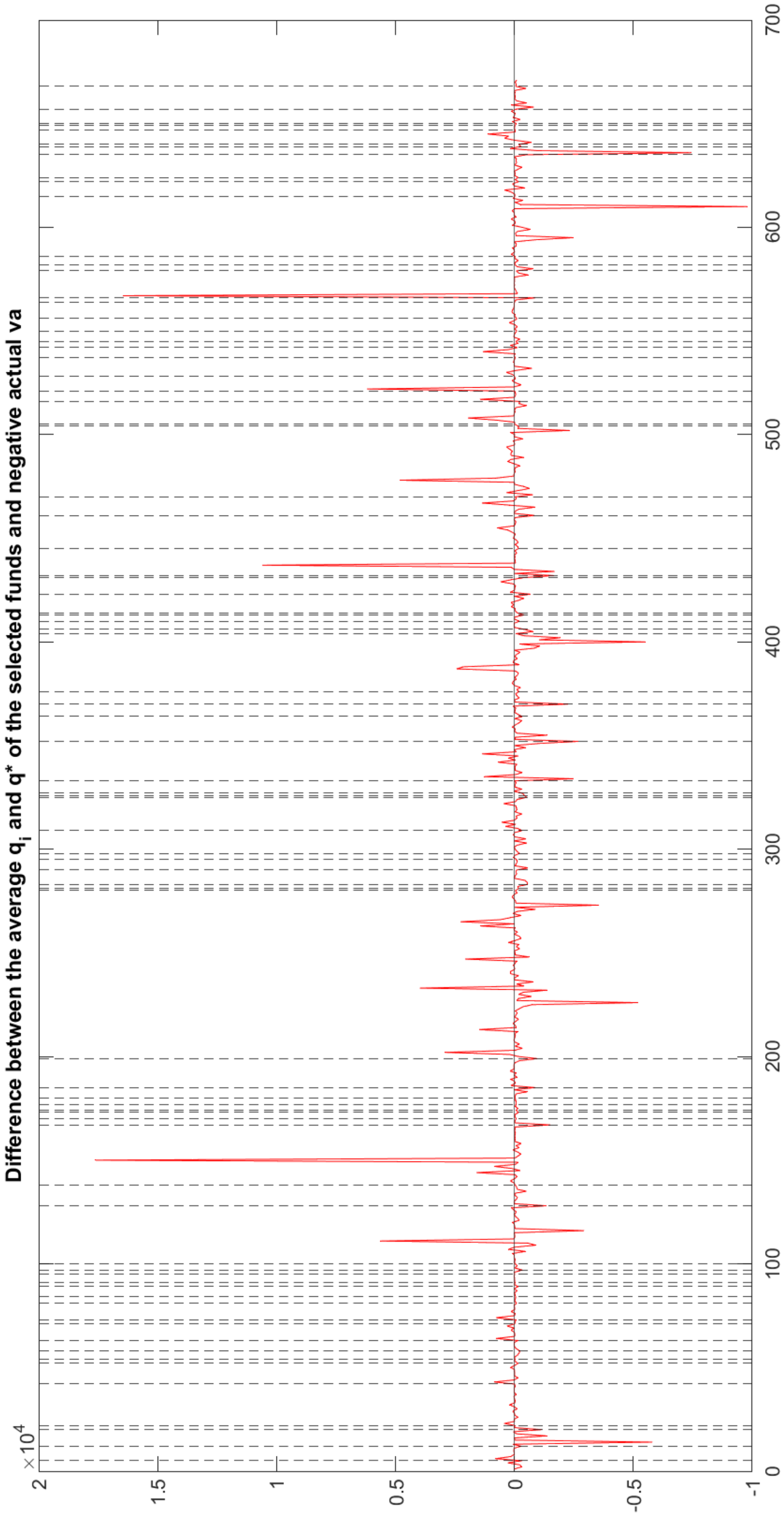

Figure 22: **Plot of the difference between q_i^* and q_i with negative va_i illustrated.** The sizes measures are given in $M in terms of the base year 2000.

List of resource persons

Prof. Dr. Julien Hambuckers, Professor of Finance at HEC Liège Management School.
Prof. Dr. Santi, Assistant Professor in International Finance at HEC Liège.
Philippe Hübner, teaching assistant of Prof. Hambuckers and doctoral student in the field of Finance.

References

Amihud, Y. and Hurvich, C. M. (2004). Predictive regressions: A reduced-bias estimation method. *The Journal of Financial and Quantitative Analysis*, 39(4):813–841.

Avramov, D., Barras, L., and Kosowski, R. (2013). Hedge fund return predictability under the magnifying glass. *The Journal of Financial and Quantitative Analysis*, 48(4):1057–1083.

Bai, J., Bali, T. G., and Wen, Q. (2019). Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, 131(3):619–642.

Bali, T. G. (2019). Turan G. Bali - Data Working Papers.

Bali, T. G., Brown, S. J., and Caglayan, M. O. (2014). Macroeconomic risk and hedge fund returns. *Journal of Financial Economics*, 114(1):1–19.

Bao, J., Pan, J., and Wang, J. (2011). The illiquidity of corporate bonds. *Journal of Finance*, 66:911–946.

Barras, L., Gagliardini, P., and Scaillet, O. (2021). Skill, Scale, and Value Creation in the Mutual Fund Industry. *The Journal of Finance*, 77(1):601–638.

Barras, L., Scaillet, O., and Wermers, R. (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The Journal of Finance*, 65(1):179–216.

Berk, J. B. and Green, R. C. (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112(6):1269–1295.

Berk, J. B. and van Binsbergen, J. H. (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics*, 118(1):1–20.

Blake, C. R., Elton, E. J., and Gruber, M. J. (1993). The performance of bond mutual funds. *The Journal of Business*, 66(3):371–403.

Cai, B., Cheng, T., and Yan, C. (2018). Time-varying skills (versus luck) in u.s. active mutual funds and hedge funds. *Journal of Empirical Finance*, 49:81–106.

Chen, J., Hong, H., Huang, M., and Kubik, J. D. (2004). Does fund size erode mutual fund performance? the role of liquidity and organization. *The American Economic Review*, 94(5):1276–1302.

Elton, E. J., Gruber, M. J., and Blake, C. R. (1995). Fundamental economic variables, expected returns, and bond fund performance. *The Journal of Finance*, 50(4):1229–1256.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

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

Federal Reserve Bank of St. Louis (1991). Interest Rates | FRED | St. Louis Fed.

Fernando, D., Klapper, L. F., Sulla, V., and Vittas, D. (2003). *The global growth of mutual funds*. The World Bank.

French, K. R. (1993). Kenneth R. French - Data Library.

FRENCH, K. R. (2008). Presidential address: The cost of active investing. *The journal of finance.*, 63(4).

Gagliardini, P., Ossola, E., and Scaillet, O. (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica*, 84(3):985–1046.

Greene, W. (2008). *Econometric Analysis Global Edition.* Pearson-prentice Hall, 8th edition edition.

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

Hansen, B. E. (2009). Lecture Notes on Nonparametrics. Technical report.

MathWorks (2022). Time Series Regression VIII: Lagged Variables and Estimator Bias - MATLAB Simulink Example - MathWorks Benelux.

Mordor-Intelligence (2021). *US Mutual Funds Industry - growth, trends, Covid-19 impact, and forecasts (2022 - 2027).* Consulted the 12th of May, 2022 on https://www.mordorintelligence.com/industry-reports/us-mutual-funds-industry#:~:text=The%20US%20shares%20grew%20186,the%20mutual%20fund%20segment%20globally.

Pollet, J. M. and Wilson, M. (2008). How does size affect mutual fund behavior? *The Journal of Finance*, 63(6):2941–2969.

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, 39(1):119–138.

Thune, K. (2021). Learn More About the Various Share Class Types of Mutual Funds.

US Inflation Calculator (2022). Current US Inflation Rates: 2000-2022 | US Inflation Calculator.

Executive Summary

This research thesis aims at applying the performance measures of Barras et al. (2021) on US bond mutual funds from July 2004 to December 2020 unlike most studies which concern equity mutual funds. Performance can be assessed in several ways but the most common measure of performance stays the abnormal return (alpha). However, this thesis supports that assessing the managers' skill the same way as assessing the fund's performance might not deliver the right conclusions especially regarding the managerial skills. Thereby, the gross alpha has been broken down into a skill and a scale coefficient. Mutual funds suffering from diseconomies of scale, the managers might be skilled but their performance could be undermined by the size of the fund. The results show that most US bond mutual funds are actually skilled, confirming the previous statement. Then, the performance of the fund is assessed based on the value added it brings to the market which is calculated by multiplying the gross alpha by the average fund's size. The densities of the skill, scale and value added measures could be deduced thanks to a non-parametric approach. As a result, it has been possible to obtain some reliable bias-adjusted densities from which some conclusions can be drawn, i.e. 87% of the bond funds create value in the market between 2004 and 2020.

Words count= 14 816