European Equity Fund Managers: Luck or Skill?!

Juliana CAICEDO-LLANO *   Enareta KURTBEGU †

June 18, 2013

Abstract

Seeking persistent abnormal portfolio performance has been a key question for academic and practitioners in the asset management industry. One of the uses is the construction of Fund-of-Funds (FoF) by the selection "skilled" managers. We suggest to use a procedure based on the managers' estimated *alphas* that filters the "lucky" managers taking them out of the selection. The standard tests to identify funds with non-zero *alphas* do not adequately account for the presence of "lucky" funds. The Family-Wise Error Rate ($FWER$) approach and the False Discovery Rate (FDR) method are the two procedures proposed recently for excluding the managers which have significant estimated *alphas*, while their true ones are equal to zero. We estimate the managers' *alphas* using different asset pricing models. The performance of the FoF created with these two selection procedures is analyzed not only in-sample (long run *versus* short run) but also out-of-sample (rolling window) as we analyze the performance of notorious portfolios. To measure the portfolios performance we use traditional Sharpe ratio and a measure of cumulative losses as the maximum draw-down among others. We find that both methods are useful to improve the performance of FoF, the results being sensitive to the sample and asset pricing model.

Keywords: Fund-of-Funds, Factor Models, Performance.

JEL Classification: C12, C14, G11.

* Ph.D. Lecturer, University of Evry-Val d'Essonne, Department of Economics, EPEE, TEPP and EONOS Investment Technologies, France. Address: 4, Boulevard Francois Miterrand, 91025, Evry Cedex, France. Tel: +33 1 69 47 70 78. E-mail: juliana.caicedollano@univ-evry.fr

† Ph.D. Student, Department of Economics, University of Evry-Val d'Essonne, EPEE and TEPP. Address: 4, Boulevard Francois Miterrand, 91025, Evry Cedex, France. Tel: +33 1 69 47 71 57. E-mail: enareta.kurtbegu@univ-evry.fr

1 Introduction

Market efficiency, persistence of abnormal returns and managers selection are among the key questions related to the topic of portfolio performance measurement which have been studied in recent years by practitioners and financial academics. The reason of creating Fund-of-Funds (FoF) is the reduced level of risk compared to investing in an individual fund and the possibility of including different portfolio management approaches. Multi-managers face the task of selecting a small number of funds with attractive return properties based on past track records while the main challenge is to select the funds which are managed by the "skilled" managers. We consider as "skilled" managers, the ones who statistically outperform their benchmark. Therefore, we name "lucky" managers the ones who will pass the filter of the selection procedure while being selected "by accident" as funds outperforming the benchmark, while in reality they will not have a good performance. The selection procedure is based on the *alphas* of the managers and the goal is not to leave any "lucky" one to be part of this selection.

Two statistical techniques have been recently proposed to achieve this selection, the Family-Wise Error Rate ($FWER$) approach proposed by Wolf and Wunderli (2009) [22] and the False Discovery Rate (FDR) approach proposed by Barras *et al.* (2010)[14]. Using the funds' returns and the corresponding benchmark, based on the historical data, these two methods manage to select only the funds delivering true significant positive performance and leaving away the ones which are not predicted to have a high future performance. Of course, we cannot observe the true *alpha* of each fund in the population. Therefore, a reasonable way to estimate the prevalence of "skilled" fund managers is to test for having a number of funds with sufficiently high estimated *alphas*. The limit level of outperforming the benchmark is called the threshold among "lucky" and "skilled" managers.

Differently from the existing literature which focuses on non-European data such as US and UK, the only analysis in the existing literature which uses European data but quite country specific restricted is based on German equity mutual fund industry. There exist different statistical bootstrap methods used for p-value correction. Therefore, in our study we consider not only the bootstrap selection FDR method but also the stepwise bootstrap as in $FWER$ selection approach. The comparison of these two selection procedures is not studied so far, in the existing literature.

Based on the two selection procedures, first, we study the link between the proportion of "lucky" *versus* "skilled" managers and the influence of the number of bootstraps and the "confidence" threshold in the performance of the FoF. The variation of the threshold toward luck does not always imply a decrease in the performance because of the increase in the diversification while more funds are taken in consideration, using a simple asset pricing model.

Secondly, we consider alternative unconditional asset pricing factor models and discuss the consequences in terms of FoF performance. The $CAPM$, the three factor Fama-French model together with the conditional Carhart model are compared. We show that different *alpha* estimates based on different factor model estimations modify the discussion on the results.

Thirdly, once the two bootstrap methods select the high performance funds, we construct the FoF using different kinds of portfolios such as equally-weighted (EW), minimum-variance (MV), equally-risk (ER) among others. We analyze the performance of the resulting funds using the traditional measures such as total/average return and volatility, and the Sharp ratio and also more robust measures such as the L-moments and the $L1$-performance, which take into account the presence of higher moments as well as a cumulative risk measure such as the maximum draw-dawn.

Finally, the two main approaches implemented to create the FoF with highest performance are analyzed in-sample in the short and long run and out-of-sample using a rolling window strategy. A comparison among the performances of the created FoFs based on the two selection procedures and notorious portfolios is done.

We find that both selection procedures reduce the number of selected funds in the sample and often improve significantly the performance of FoF. Other important results show the stability of the bootstrap procedure and the comparison of different asset pricing methods to estimate *alpha*. Our main conclusion is that the FDR and $FWER$ are both appropriate methods to identify mutual funds with non-zero *alphas*.

The reminder of this paper is organized as follows. Section 2 consists on the literature review. A quick methodology description of the theoretical framework based on the two fund selection procedures, the asset pricing models and the notorious portfolios is presented in section 3. Moreover we introduce the dataset description and the main descriptive statistics in section 4. Section 5 summarizes the paper findings and results including the main in-sample (including the "back-test" procedure) and out-of-sample ("observe one year, invest three months" strategy) conclusions differences between the two bootstrap procedures as we change the threshold delta not only in short-run but also in the long-run. Moreover, we discuss the impact of different asset pricing methods as estimation of the parameter *alpha* on the portfolio performance, while section 6 concludes.

2 Literature review

There are two main axes interfering in this literature. On one hand there exists a question according to the sign of the average risk adjusted abnormal fund performance. On the other side, the issue is whether the abnormal performances can be *ex-ante* identified and the challenge is to implement methods that select funds with superior future performance. Even if many research articles (Jensen 1968 [12], Carhart 1997 [5] among others) suggest negative average fund *alphas*, recent papers show that managers have selection skills. Studies done from Pastor and Stambaugh (2002*b*) [17] and Avramov and Wermers (2006) [3] use the Bayesian perspective to show the benefits of investing in managed funds, Romano and Wolf (2005)[19] follow a stepwise bootstrap procedure, while the Kosowski *et al.* (2006) [18] use a bootstrap procedure to select the *ex-post* outperformed funds. Based on the later bootstrap method, first Benjamini and Hochberg (1995) [4] introduced a statistical method named the False Discovery Rate (FDR) used to qualify the impact of luck on mutual fund performance, where the FDR itself measures the proportion of lucky funds among the significant ones.

The US studies of mutual funds of Lakonishok *et al.* (1992)[10], Grinblatt *et al.* (1995) [15], Carhart (1997) [5], Wermers (2000) [21], Pastor and Stambaugh (2002) [17], Barras *et al.* (2010) [14], Fama and French (2010) [8] among others conclude for no superior performance but somehow stronger evidence of under-performance. The results of Wolf and Wunderli (2009) [22], are in line with previous papers, while the study is done using US hedge fund data. Using the UK equity mutual funds data as in Blake *et al.* (1999) [7] and Cuthbertson *et al.* (2008) [13] the results are quite similar while the power properties of standard tests of abnormal performance are quite low even for high level of abnormal performance. Their analysis rejects the hypothesis that most poor performing funds are unlucky. Moreover, they conclude that the positive performance among onshore funds is due to "skill", while the positive performance among offshore funds is due to "luck". The only study in this literature done using European data is the specific case in the German Equity long-only mutual fund industry for a period of 20 years by Cuthbertson and Nitzsche (2010) [6]. Using the FDR bootstrap method in the three factor model with market timing model result in a large increase of in the proportion of truly positive *alpha* funds and it is the only factor whose inclusion varies the results, while different sample periods, the alternative factor models and investment in German or non-German firms keeps the results invariant.

The literature in fund selections focuses on non-European data such as US and UK, the only study using European country specific data is the German Equity mutual fund data. Differently from the existing literature in our paper, we discuss a large European equity mutual fund database for a seven year data period. Moreover, we are interested in the selection results of not only the FDR but also in the $FWER$ selection model. A comparison between these two methods is not yet empirically studied. Both methods use a bootstrap selection but in a slightly different framework. In the previous cited papers, researchers compare in-sample and out-of-sample strategies based on EW portfolios or MV ones. Thus, the goal of our paper is to go beyond the classical portfolio strategies and to consider among them also other portfolios such as the ERC and suggest other methods to build portfolios taking the fund selection methods into account.

3 Methodology

Both methods consist on multi-testing selection procedures, while selecting the "skilled" managers and excluding the "lucky" ones at the same time. Based on the *alphas* of the fund managers, the two methods proceed in slightly different ways. The $FWER$ consists in detecting whether there is at least one fund with non-zero *alpha*. Thus, it is defined as the probability of yielding at least one "lucky" fund among the N tested ones. In contrast to the $FWER$ method, the FDR approach is designed to measure the proportion of "lucky" funds among a given set of significant funds. Despite the differences, both of these methods are based in the *alpha* of each fund manager. In this paper we will start giving results using alphas estimated from a Capital Asset Pricing Model $CAPM$ which is based on a single market factor proxied here by the excess return of a market index over the risk-free rate. Then, we compare these results based on *alphas* estimated with the Fama-French three-factor model which includes a size and a style factor as well as the market factor and with the four-factor Carhart model which includes a momentum factor.

3.1 The Family-Wise Error Rate ($FWER$) Procedure

The Family-Wise Error Rate ($FWER$) approach proposed by Wolf and Wunderli (2009)[22] is one of the methods used to provide a robust selection process. The fund selection is based on the track records of the individual managers. The manager is considered to be "skilled" if his *alpha* with respect to a suitable benchmark is proved statistically to be positive beyond a certain threshold of doubt. The significance level of the test is represented by the doubt level which considers the possibility that a "lucky" manager passes the test while he is wrongly identified as "skilled". The existence of the "lucky" managers in the Fund-of-Funds (FoF) structure may lead to a decrease of its performance. Thus, it happens that the "skilled" managers will continue to outperform, while the "lucky" ones will not, inducing in this way a decrease of the performance of the overall FoF.

Since the selection process of the "skilled" managers consists of a procedure which takes in consideration the whole managers at the same time, the chance of the "lucky" managers to be selected and considered as "skilled" accumulates. Let us consider N funds where for each we consider a return history of T observations. Following the previous notations, for $n = [1, , N]$ the *alpha* of a given fund manager with respect to his benchmark is denoted by α_n. For each of the individual managers we look at individual hypotheses of the form:

$$\begin{cases} H_0 : \alpha_n \leq 0 \\ H_1 : \alpha_n > 0. \end{cases} \tag{1}$$

The null hypothesis where the *alpha* is considered to be negative or zero corresponds to the "non-skilled" manager, while the positive *alpha* reflects the "skilled" manager. There exists the possibility for the "non-skilled" managers to pass the statistical test by chance. A manager n is considered as "skilled" by the statistical method used if H_0 is rejected. In this situation two possibilities exist. On one side, the false discovery situation may be caused, where the H_0 hypothesis is actually true, so we make a mistake by considering the "skilled" manager as being "non-skilled". On the other side, it may happen that H_0 hypothesis is actually false so while rejecting it we take the right decision. In this case we have to do with the statistical phenomena called true discovery. The aim of the $FWER$ approach is to account for the possibility of even one "lucky" manager to make even one single false discovery. We denote by F the number of false discoveries. The $FWER$ method consists in determining the probability of making even one false discovery. It is expressed as the following:

$$FWER = prob\,\{F > 0\} \tag{2}$$

The $FWER$ is the probability of rejecting at least one H_0, where the fund n belongs to the set of the selected funds. We have to do with multiple testing because the testing procedure is done for some managers at the same time. The aim is to keep the probability of rejecting at least one H_0 lie below some certain threshold. In this way the probability that even one "lucky" manager passes the test is limited. Therefore, the goal is to keep holding true the following inequality:

$$FWER < \gamma \tag{3}$$

So, if we consider the level of precision $\gamma = 5\%$, then after applying the $FWER$ selection approach we can argue that we are 95% confident that all the identified managers are truly skilled. Thus, the FoF portfolio will consist of skilled managers with a high probability. Details about the estimation procedure are given in Appendix A 7.

3.2 The False Discovery Rate (FDR) Approach

The FDR approach is used to qualify the impact of "luck" on fund performance. It was first introduced by Benjamini and Hochberg (1995) [4] and now it is widely used to measure the proportion of "lucky" funds among the funds with significant estimated *alphas*. Barras *et al.* (2010)[14] propose a slightly different approach under the same name. They consider separately the positive and the negative estimated *alphas* and compute the FDR according to this pre-selection. The main positive effect of this method is that it is easy to compute from estimated p-values of fund *alphas*. The presence of different performances (positive and negative *alphas*) is tested for each of the N funds in the population. We consider a fund to have significant estimated *alpha* if its p-value is smaller than the conventional significance level set before denoted by δ. We define the "lucky" manager as the one whose fund has a significant estimated *alpha* different from zero while his true *alpha* is equal to zero. The difficulty of the method is that the tests are repeated N times because of the multiple testing framework. Thus, luck cannot be measured by the significance level applied in each fund separately. Keeping the previous notations, the null hypothesis H_0 indicates the fund n achieves no performance while the alternative hypothesis H_1, if it is true, identifies the funds with differential performance, such as:

$$\begin{cases} H_0 : \alpha_n = 0 \\ H_1 : \alpha_n > 0 \; or \; \alpha_n < 0. \end{cases} \tag{4}$$

The null hypothesis is rejected for all funds with estimated p-values smaller than δ share, implying significant estimated *alphas*. Testing the null hypothesis *versus* the alternative one, there are some possible outcomes. The total number of funds N is composed by the ones with no performance (W_δ) and the ones with significant differential performance (R_δ). Among the (R_δ) significant funds, there are funds with no performance which are incorrectly classified as significant funds (F_δ) of them which are considered "lucky". Therefore, the correct number of funds which have true differential performance is calculated as the difference between R_δ and F_δ is the number of significant funds which truly yield differential performance (Q_δ). Denoting the number of funds with no performance as W_δ, one can mention the existence of funds which have differential performance but which are incorrectly classified as funds with zero *alphas* (A_δ). Thus, to get the funds with no performance we subtract A_δ from the total number of funds with no performance W_δ and we get the funds with no performance who are considered as funds with zero *alpha* (M_δ).

The approach states that while increasing the significance level δ, we will have an increase in both the number of significant funds R_δ and the number of "lucky" funds F_δ. The increase of R_δ is only due to the detection of new funds with differential performance, thus it cannot capture the increase because of the inclusion of "lucky" funds. This is a main problem that the standard approach (Benjamini and Hochberg; 1995 [4]) cannot solve.

Based on the notations given above, the FDR is defined as the expected proportion of "lucky" funds among the significant funds for a given predefined significance level δ. The FDR_δ is calculated as the conditional expectation of the proportion of the "lucky" funds knowing that they have positive performance. Thus the FDR, which is the probability of "lucky" funds, is given as the proportion of "lucky" funds times the significance level, multiplied by the weighted sum of the "lucky" funds under the null and the alternative hypothesis. The funds detected to have differential performance consist of either positive or negative *alphas*. Based on the sign of the estimated *alphas* we consider two groups for the significant funds R_δ. The first group contains the R_δ^+ funds with positive estimated *alphas* and the second group contains the R_δ^- funds with negative estimated *alphas*. The same idea is used to denote the "lucky" funds among the best ones and among the worst ones, respectively denoted F_δ^+ and F_δ^-. To measure the relative importance of F_δ^+ and F_δ^-, the FDR is computed separately among the best and the worst funds. Authors assume that since the null hypothesis of no performance is a two-sided test with equal-tailed confidence level, then there exists half of the "lucky" funds with positive estimated *alphas* and half with negative estimated *alphas*. Thus, denoting the corresponding FDR of best and worst funds respectively as $FDR^+(\delta)$ and $FDR^-(\delta)$, we can write the following:

$$\begin{cases} FDR^+(\delta) = E\left[F_\delta^+(R_\delta^+)^{-1} |\, R_\delta^+ > 0\right] = E\left[1/2F_\delta(R_\delta^+)^{-1} |\, R_\delta^+ > 0\right]; \\ FDR^-(\delta) = E\left[F_\delta^-(R_\delta^-)^{-1} |\, R_\delta^- > 0\right] = E\left[1/2F_\delta(R_\delta^-)^{-1} |\, R_\delta^- > 0\right]. \end{cases} \tag{5}$$

These measures are designed to deal with quantifying separately the proportion of "lucky" funds in the right tail of the cross-sectional *alpha* distribution and the proportion of "lucky" funds in the corresponding left tail. Details about the estimation procedure are given in Appendix B 8.

3.3 Portfolio construction and performance measurement

Differently from the portfolio manager's objectives that have to deal with asset allocation decisions, the managers are interested in how the risk of individual investment affects the overall risk of the entire fund. The "skilled" funds are selected to create the FoF whose performance is measured using the Sharpe Ratio, Max drawdown, L-moments and $L1$-performance. This new fund performance measure, called the $L1$-performance, is conceptually close to the conventional power moments, but provides more detailed information about the extremes. Since the aim of this of this selection process is to check the extremes and to pick up only under a certain security threshold, the use of L-moments is appropriate. The FDR and $FWER$ methods determine only which funds to select, delivering true excess returns, in order to build a portfolio with the selection of funds. In order to build the Fund-of-Funds (FoF), once the selection process is complete, a method to determine the weight to allow to each fund has to be determined. We consider three standard methods to determine the weights of the portfolio: An equal weights portfolio (EW), a minimum variance portfolio following the optimization methods used in modern portfolio theory (MV) and an equal contribution to risk portfolio (ERC) as in Maillard *et al.* (2010) [20]. We also suggest to calculate the optimal weights based on the bootstrapped *alphas* used during the selection process (*alpha*).

4 Data description

4.1 Descriptive statistics

We consider 89 European equity mutual funds for the period from January 2005 to February 2012 observed at a daily basis. We have calculated average descriptive statistics for this set of mutual funds (See Table 1). The funds present an average return for the whole period of 1.4% and an average annualized volatility of 21.7%. We have also tested for normality of the returns with the JB statistic. We decomposed the sample into two subperiods of equal length, the first one (S_1) corresponding to the pre-crisis period (February 2005-August 2008) and the second (S_2) considering the crisis and post-crisis period (August 2008-February 2012). We observe the difference in the statistics before and after the financial crisis. The average annual return drops from nearly 4% to -1% and the volatility increases 73% from 15% to 26%.

Table 1: Descriptive Statistics Mutual Funds

	Mean	St. Dev	Skew	Kurt	JB-pvalue
All data	1.39	21.71	0	0	0
< 2008	3.75	15.21	0	0	0
$>= 2008$	-0.99	26.38	0	0	0

Source: Descriptive statistics of the dataset of 89 European equity mutual funds during Jan 2005-Feb 2012 (All data), during the pre-crisis period (Jan 2005-Aug 2008) and during the post-crisis period (Aug 2008-Feb 2012); computations by the author.

We use the series of the $EURIBOR$ one month as risk-free returns while the market risk factor is proxied by the $MSCI$ Europe Index. Other risk factors needed to estimate *alpha*, are the Fama-French factors Size and Style. For the size factor (SMB)[1], we calculated the excess performance of the $MSCI$ Small Cap Index over the performance of an Index composed with only 50 big companies from the European region [2]. As for the style factor (HML), we computed the extra performance of the $MSCI$ Value Index over the $MSCI$ Growth Index instead of using the approach of Fama and French who estimated the style factor using portfolios associated with different fundamental criteria [3]. These factors have the following descriptive statistics (see Table 2):

Table 2: Descriptive Statistics MSCI Index and Fama-French factors

						Correlation		
	TotR	AvgR	Vol	IR	MDD	MSCI	SMB	HML
MSCI Europe	−8.6	−1.217	26.1	0	−96.3	1	−0.4614	0.5242
SMB	86.3	8.374	12.2	0.7	−24.4	−0.4614	1	−0.3943
HML	−18.3	−2.719	8.6	−0.3	−32.7	0.5242	−0.3943	1

Source: Descriptive statistics of the MSCI Index (benchmark) and the Fama-French factors, Small-Minus-Big (SMB) and High-Minus-Low (HML), from Jan 2005 to Feb 2012; Descriptive statistics contain: total return, average return, Sharpe Ratio, Volatility, Max Drawdown, Correlations among variables; computations by the author.

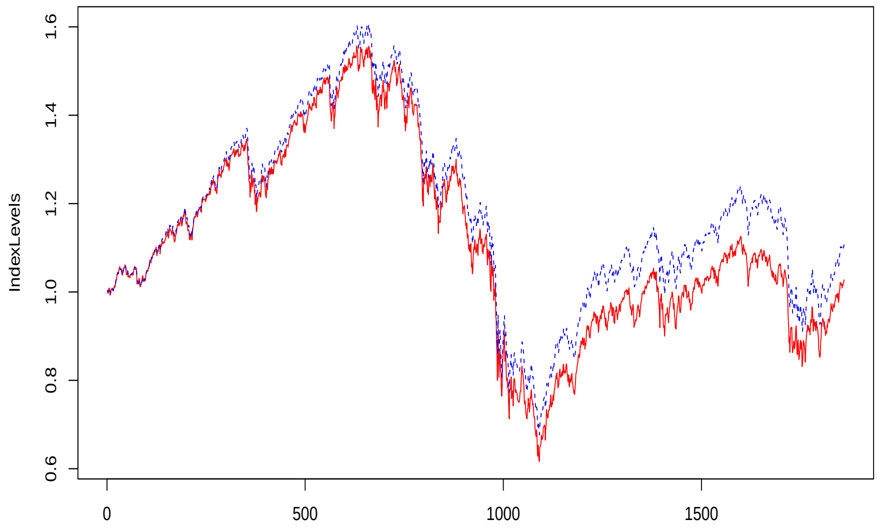
This database is quite interesting because it basically contains two sub-samples. The first one corresponds to a bull market period while the second one corresponds to the 2008 global financial crisis which is mainly a bear market over the period $2008 - 2012$. We can therefore test our method based on three different samples: one considering the whole sample, one capturing only the bullish period and finally a sample capturing the market crash followed by a recovery period. The following graph presents the MSCI Index used as benchmark and the equally weighted portfolio of all the 89 funds in the database (see Graph 1), one could generalize that a similar trend is visible in each fund of the database.

[1]SMB: Small minus Big and HML: High minus Low factors are usually calculated as the differential performance of sets of portfolios with small and big companies as well as high and low valuation ratios.

[2]The $MSCI$ Big cap Index has only a history of 3 years. It has been replaced by the Eurostoxx 50 Index.

[3]Fama and French use the Market to Book ratio to build the HML factor.

Figure 1: Benchmark (MSCI Index) and EW portfolio.



Source: Based on MSCI Europe Index; daily data from (Jan 2005-Feb 2012) in EUR; We represent in this figure the Benchmark portfolio (in red) and the Equally-Weighted portfolio among the totality of 89 European Mutual fund database (in blue). The x-axis represents the time and in the y-axis are denoted the index prices. Computations by the authors.

4.2 Factor models estimates

We estimated two different factor models (CAPM and Fama-French three factors model) individually for each fund in the database. The estimated coefficients were averaged and we present the main average statistics for these regressions in Table 3. For the CAPM model the average estimated alpha is close to zero but not significant on average and the beta coefficient for the market factor is equal to 0.63. The average R^2 coefficients for these regression is equal to 0.57. We compare these results to the three factors estimation where the average sensitivity to the market factor is 0.66, and the coefficients for the size factor and the style factor are equal to 0.26 and 0.21, the adjusted R^2 being equal to 0.6.

Table 3: Average estimates for CAPM and Fama-French 3 factor models

	Estimate	t-stat	$R^2/\hat{R}^2$
alpha-CAPM	0.00	0.21	0.57
MKT	0.63	78.78	0.57
alpha-FF3	-0.00	-0.05	0.61
MKT	0.66	66.83	0.60
SMB	0.26	7.84	
HML	0.21	5.29	

Source: Newey–West t-statistic are used to overcome autocorrelation and heteroskedasticity in the error terms in the models. Calculations by the authors.

Based on the CAPM and Fama-French three factors estimations, we have made a selection of significant funds considering different levels of significance. For each level of significance, we constructed an equally weighted portfolio based on this fund selection procedure. Only 4 funds have positive significant alphas based on a CAPM estimation at 5% significance, 11 funds with a significance level of 10% and 16 funds for 20%. The resulting equally weigthed portfolios based on these selections have Sharpe ratios close to 0.2, corresponding average annual returns ranging from 3.4% to 4.1% and annualized volatilities close to 18.5% as presented in Table 4.

Table 4: Statistics for different portfolio based on CAPM alphas

	0.05	0.1	0.15	0.5	0.5+	0.9	0.95	0.975
Nb	2.00	3.00	4.00	34.00	55.00	16.00	11.00	4.00
Avg Alpha	-1.56	-5.82	-4.65	-9.84	-3.99	-1.12	0.64	1.71
Tot Ret	5.98	0.65	0.03	9.71	18.25	23.31	24.44	24.34
Avg Ret	1.56	0.17	0.01	2.49	4.51	5.63	5.88	5.86
Vol	13.74	11.01	11.68	9.36	10.78	12.18	11.88	12.89
IR	0.10	0.01	0.00	0.22	0.34	0.37	0.39	0.35
MDD	-35.92	-34.71	-37.92	-37.54	-34.98	-31.90	-31.55	-35.57

Source: Selection of significant funds considering different levels of alpha (t-statistics) significance; model: CAPM. Calculations by the authors.

Similar results are calculated for the Fama-French three factors estimation. A portfolio constructed with funds having significant alphas at 10%, has an average return of 5.5%, volatility of 19.4% and Sharpe ratio of 0.3 as it is presented in Table 5.

Table 5: Statistics different portfolio based on Fama-French alphas

	0.05	0.1	0.15	0.5	0.5+	0.9	0.95
Nb	10.00	18.00	27.00	69.00	20.00	4.00	3.00
Avg Alpha	-5.55	-7.90	-7.73	-7.77	-0.89	-0.68	-0.34
Tot Ret	8.84	10.28	9.74	13.06	21.53	34.86	35.15
Avg Ret	2.28	2.63	2.50	3.30	5.24	8.04	8.10
Vol	12.57	12.55	12.10	9.88	12.38	11.56	12.01
IR	0.15	0.17	0.17	0.27	0.34	0.55	0.52
MDD	-37.68	-37.35	-37.64	-37.32	-31.75	-29.73	-33.51

Source: Selection of significant funds considering different levels of alpha (t-statistics) significance; model: FF. Calculations by the authors.

In addition, we consider Carhart pricing model (4-Factors) which is constructed based on the Fama and French's 3-factor model (1993)[8] plus an additional factor capturing the one year momentum anomaly (Jegadeesh and Titman, 1993 [11]). The market inefficiency due to slow reaction to information is called momentum anomaly. We construct it in the same way as Carhart (1997)[5], here we build this factor based on one month and one year past performance using the top and worst 30% performers.

5 Findings and Results

5.1 FDR Findings

In-sample results

We first start with the in-sample analysis where we consider the 89 funds in the database for the whole period from January 2005 to February 2012 in daily basis. We initially focused on the standard capital asset pricing model ($CAPM$), where the only factor is the market factor[4]. The Newey West estimator is used in order to overcome autocorrelation and heteroskedasticity in the error terms. The t-statistic using Newey West varies between -2.16 and 4.18. We applied the selection procedure as described in section 3.2, this selection method based on a bootstrap procedure requires the specification of the number of samples needed to have stable estimates. After testing for different numbers of simulations, we established that the selection procedure becomes stable once we reached $B = 2000$ where B is the number of bootstrap. Beyond this threshold, the sets of selected funds is

[4]estimated as excess return of the $MSCI$ Europe Index aver the risk free rate

always the same. The size of the time-series is crucial in deciding the number of bootstraps needed to have a set of selected funds which converges[5]. Considering a sub-sample of 250 observations we conclude that the convergence toward a stable set of funds is reached with 500 samples. Furthermore, the FDR procedure is quite sensitive to the confidence threshold δ. Table 6 presents the funds selected[6] and the total number of funds chosen with the FDR procedure while we increase the significance level δ[7]. The result goes in line with our expectation. An increase in delta increases the number of selected funds which in this case varies between 4 and 19. We also observe that the funds selected for a given confidence threshold include the selection for lower levels of δ.

Table 6: The variation of the FDR fund selection for the whole sample ($CAPM$)

delta	The selected Funds	Nr. selected funds
0.05	5 8 16 54	4
0.1	5 8 16 25 41 54 56 76 83 85 86	11
0.15	5 8 16 17 25 28 41 54 56 76 77 82 83 85 86	15
0.2	5 8 16 17 25 28 41 46 54 56 76 77 82 83 85 86	16
0.25	5 8 16 17 25 28 38 41 46 54 56 76 77 82 83 85 86	17
0.3	5 6 8 16 17 25 28 38 41 46 54 56 58 76 77 82 83 85 86	19

Source: Fund selection done among a sample of 89 European Equity Mutual Funds daily quotes for a period from Jan 2005-Feb 2012. Asset pricing model: $CAPM$ (1Factor). The selection method used: FDR. Number of bootstraps: $B = 2000$. $\delta \in [0.05, 0.3]$; computations by the author.

This study will be focused also in the long-term and short-term comparison. Since the daily database we study contains the financial crisis of 2008, we focus our short term comparison in two sub-samples before and after the financial crisis. Considering this sub-sample choice is quite interesting knowing that the first sub-sample (S_1) is characterized by a bullish market while the second sub-sample (S_2) consists on a bear market during the first half followed by an increase starting by the end of 2011. By consequence the number of funds selected by the FDR method is not of the same amplitude for both sub-samples S_1, S_2 (see Table 7 and 8).

[5] We repeated the bootstrap procedure using a different number of bootstraps for each sample. For the whole sample (1960 observations) we considered $B = 1000, 2000, 5000, 10000$. For half samples (930 observations) we used $B = 500, 1000, 2000$ and for rolling samples of one year (250 observations), $B = 250, 500, 1000$. We presented the results for the B value that ensures a stable data set selection.

[6] The 89 funds in the database have a numeric identifier. The 2^{nd} column in Table 9 present the code for the selected funds from our database for each level of delta. The fund name and $ISIN$ code is found in the Appendix C 9.

[7] Results for $B = 2000$.

Table 7: The variation of the FDR fund selection for the pre-crisis sample (CAPM)

delta	The selected Funds	Nr. selected funds
0.05	5 8 25 28 41 86	6
0.1	5 8 22 25 28 41 86	7
0.15	5 8 16 22 25 28 41 76 86	9
0.2	5 8 12 16 19 22 25 28 41 54 76 86	12
0.25	5 8 12 16 19 20 22 25 28 41 46 54 67 76 82 86	16
0.3	5 6 8 12 16 17 19 20 22 23 25 28 41 46 48 54 67 76 82 83 86	21

Source: Fund selection done among a sample of 89 European Equity Mutual Funds daily quotes for a period from Jan 2005-Aug 2008. Asset pricing model: $CAPM$ (1Factor). The selection method used: FDR. Number of bootstraps: $B = 2000$. $\delta \in [0.05, 0.3]$; computations by the author.

During the bull period there are more funds with high *alpha* inducing a higher number of selected funds contrary to post-crisis sub-sample for each level of *delta*. The long-term in-sample results stay in between the short-term in-sample results this because of the crisis which is affected in the value of *alpha*, decreasing it in general. Thus, the number of selected funds differs comparing to S_1 and S_2 sub-sample results.

Table 8: The variation of the FDR fund selection for the post-crisis sample (CAPM)

delta	The selected Funds	Nr. selected funds
0.05	54 85	2
0.1	8 54 56 58 85	5
0.15	8 54 56 58 77 85	6
0.2	5 8 16 17 54 56 58 77 85	9
0.25	5 8 16 17 54 56 58 77 80 83 85	11
0.3	5 8 16 17 54 56 58 73 76 77 80 83 85	13

Source: Fund selection done among a sample of 89 European Equity Mutual Funds daily quotes for a period from Aug 2008-Feb 2012. Asset pricing model: $CAPM$ (1Factor). The selection method used: FDR. Number of bootstraps: $B = 2000$. $\delta \in [0.05, 0.3]$; computations by the author.

We observe in the previous tables that the same group of funds are selected at each test confirming the robustness of the FDR selection procedure. The FDR procedure selects not only the funds with obvious positive high *alpha* but also those funds who do not have *ex-ante* high *alpha* (t-statistic) but that will *ex-post* show a quite good performance. Selection based not on a bootstrapped procedure selects only a part of the funds that

will *ex-post* perform well. The bootstrap procedure make it possible to capture also other information of the fund (other than the total estimated t-statistic estimated). Thus, it selects more funds considering also a probability δ of having chased some lucky ones.
Moreover, we tested the importance of the bootstrapping procedure of the residuals by replacing it by a simple procedure of constructing residuals from a normal random variable distributed normally with mean zero and volatility the same as the residuals of the regression. The results show that there is not a high difference among these two different ways of creating the new returns. The funds selected are almost the same with a change of 1 or 2 funds in some cases. This change may be due to the loss of the information we induce by introducing different residual values even why they are distributed in the same way. We also checked the normality of the residuals and we concluded that they are normally distributed (Jarque-Bera test). Furthermore, we tested for autocorrelation (Durbin-Watson test) and heteroskedasticity using the Engel test for residual heteroskedasticity (Autoregressive Conditional Heteroskedasticity, $ARCH$). There is rejection of the null hypothesis for all the fund regressions done inducing no heteroskedasticity effect. Additionally, the results of the autocorrelation test reject H_0 at 5% significance for most of the regressions but not for the totality, implying the existence of the autocorrelation in some fund regressions.

We analyzed the features of different portfolios composed by FDR selctions and compared them to the Benchmark and to an equally weigthed portfolio including all the funds in the database. This increase in delta means, on one side, that there is a higher probability of having selected "involuntary" funds with low *alphas* while on the other side it means that the portfolio diversification has increased. Table 9 describes some performance characteristics of the the benchmark (1^{st} column), the EW portfolio among all set of funds (2^{nd} column) and the equally weighted porfolios corresponding to the selected funds by the FDR method for different values of significant levels (3^{rd}-8^{th} column). Comparing to the benchmark performance, creating the FoF by selecting *ex-ante* a certain number of funds, induces out-performance, and a decrease in volatility almost 50%. The maximum drawdown of the FoF decreases by almost 50% compared to the result of the benchmark. These indicators remains almost the same while comparing the FoF by selected funds for different deltas and the portfolio of all 89-fund dataset. Furthermore the EW-all outperforms all other FoF. The variation of delta and the selection of more funds implies both consequences, diversification and "luck". The in-sample FoF constructed based on $delta = 0.15$ is the most optimal selection in terms of total performance and risk adjusted performance.

Table 9: $EW - FoF$ for different deltas (S) - $CAPM$ (1Factor)

	MSCI	EW-all	2k5	2k10	2k15	2k20	2k25	2k30
TotR	2.72	10.91	39.63	37.85	34.84	34.09	33.78	32.14
AvgR	0.36	1.39	4.49	4.31	4.02	3.94	3.91	3.75
Vol	21.28	15.15	19.43	18.67	18.91	18.67	18.48	18.63
IR	0.02	0.09	0.23	0.23	0.21	0.21	0.21	0.20
MDD	-83.93	-82.21	-77.13	-68.93	-69.38	-68.75	-69.11	-69.64

Source: Fund selection done among a sample of 89 European Equity Mutual Funds daily quotes for a period from Jan 2005-Feb 2012. Asset pricing model: $CAPM$ (1Factor). The selection method used: FDR. Number of bootstraps: $B = 2000$. $\delta \in [0.05, 0.3]$. Performance indicators: total return, average return, Sharpe Ratio, Volatility, Max Drawdown, L-moments; computations by the author.

These results are somehow ambiguous and sometimes contrary to the intuition, but they are explained by the fact that this is an in-sample selection procedure for a very long (7-years) daily data containing a crisis period. Meanwhile, these results become more clear when we discuss the sub-sample results (see Table 10 and 11). The bull market (S_1) implying higher positive *alphas*, and more selected funds, concludes that constructing $EW - FoF$ out-performs the benchmark and the EW-all for each value of significance level delta. Moreover, the optimal delta to be considered is $\delta = 0.2$. It is remarkable the increase in performance reaching 38.5%, an increase in Sharpe Ratio, and a decrease in volatility more than 50% compared to the benchmark and the EW-all. At the same time, results for the maximum drawdown are optimistic as well. The existence of the selection of "lucky" fund managers imply the decrease in the indicators of performance as the results for $\delta = 0.15$ and $\delta = 0.25$. In all the cases, the FDR selection produces portfolios with better risk-adjusted performance measures.

Table 10: $EW - FoF$ for different deltas (S_1)- $CAPM$ (1Factor)

	MSCI	EW-all	2k5	2k10	2k15	2k20	2k25	2k30
TotR	8.39	14.91	29.56	28.01	28.20	25.96	23.44	21.94
AvgR	2.17	3.74	6.96	6.64	6.68	6.20	5.66	5.33
Vol	15.12	10.03	13.49	13.74	12.99	12.89	12.91	12.99
IR	0.12	0.30	0.42	0.39	0.41	0.38	0.35	0.33
MDD	-37.75	-36.00	-29.78	-30.92	-31.52	-32.29	-32.52	-32.43

Source: Fund selection done among a sample of 89 European Equity Mutual Funds daily quotes for a period from Jan 2005-Aug 2008. Asset pricing model: $CAPM$ (1Factor). The selection method used: FDR. Number of bootstraps: $B = 2000$. $\delta \in [0.05, 0.3]$. Performance indicators: total return, average return, Sharpe Ratio, Volatility, Max Drawdown, L-moments; computations by the author.

The picture is not the same as we refer to the post-crisis period (S_2) where the face of the market index has dropped and is quite volatile. Investing in $EW - FoF$ is not a good choice, neither in terms of performance nor of Sharpe ratio, but it induces a decrease in volatility of more than 50%, and a decrease in the maximum drawdown compared to the benchmark.

Table 11: $EW - FoF$ for different deltas (S_2)- $CAPM$ (1Factor)

	MSCI	EW-all	2k5	2k10	2k15	2k20	2k25	2k30
TotR	-5.88	-3.57	14.64	13.67	15.06	13.15	11.37	10.23
AvgR	-1.63	-0.98	3.68	3.45	3.78	3.33	2.90	2.62
Vol	26.03	18.95	22.49	22.69	23.16	23.34	23.45	23.18
IR	-0.07	-0.06	0.20	0.18	0.19	0.17	0.15	0.13
MDD	-55.62	-54.34	-49.87	-47.51	-48.51	-47.78	-47.37	-46.97

Source: Fund selection done among a sample of 89 European Equity Mutual Funds daily quotes for a period from Aug 2008-Feb 2012. Asset pricing model: $CAPM$ (1Factor). The selection method used: FDR. Number of bootstraps: $B = 2000$. $\delta \in [0.05, 0.3]$. Performance indicators: total return, average return, Sharpe Ratio, Volatility, Max Drawdown, L-moments; computations by the author.

Every portfolio selection based on the FDR procedure has better risk adjusted measures of performance compared to the benchmark and the equally weigthed portfolio composed of all the funds in the database.

The pricing model used to calculate alpha implies slightly different results. The increasing number of factors to three, induces that some part of the value of *alpha* calculated by the $CAPM$ model is now captured by Fama-French factors. The decrease in its value implies a smaller number of selected funds by the FDR procedure for the entire period, sample S (see Table 12). The increase in delta induces a softer increase in the number of selected funds but not in the same amplitude.

Table 12: FDR fund selection - whole sample (Fama-French)

delta	The selected Funds	Nr. selected funds
0.05	25	1
0.1	16 25 54	3
0.15	16 25 41 54	4
0.2	16 25 41 54	4
0.25	16 25 41 54	4
0.3	16 25 41 54	4

Source: Fund selection done among a sample of 89 European Equity Mutual Funds daily quotes for a period from Jan 2005-Feb 2012. Asset pricing model: *Fama-French* (3Factor). The selection method used: FDR. Number of bootstraps: $B = 2000$. $\delta \in [0.05, 0.3]$; computations by the author.

Results for the pre-crisis sub-sample selection (S_1) are quite similar to the corresponding $CAPM$, but they variate to a smaller range (4-17) selected funds while the $CAPM$ results variate in the range (6-21) selected funds for the same *delta* variation. Moreover, the set of the considered ”skilled” fund managers is more or less the same. In line with the previous results, the increase of the ”significance” is followed by an increase in the number of the selected funds.

Table 13: FDR fund selection - pre-crisis sample (Fama-French)

delta	The selected Funds	Nr. selected funds
0.05	5 8 25 41	4
0.1	5 8 25 28 41 86	6
0.15	5 8 16 23 25 28 41 86	8
0.2	5 8 16 20 23 25 28 41 76 86 88	11
0.25	5 6 8 12 16 20 23 25 28 41 67 76 82 83 86 88	16
0.3	5 6 8 12 16 20 23 25 28 41 54 67 76 82 83 86 88	17

Source:Fund selection done among a sample of 89 European Equity Mutual Funds daily quotes for a period from Jan 2005- Aug 2008. Asset pricing model: *Fama-French* (3Factor). The selection method used: FDR. Number of bootstraps: $B = 2000$. $\delta \in [0.05, 0.3]$; computations by the author.

The picture is slightly different in the Fama-French post-crisis sample (S_2) where the *alphas'* values are mostly negative because of the market performance of this period and quite small. Thus the p-values are higher than 0.35 leading to the selection of no funds for *delta* smaller than 35%. In analogy to the previous results according to the performance of the $FoFs$, portfolios with the totality of the funds and the benchmark using the 3-factor model one could conclude the same results as the ones induced by a $CAPM$ model. The EW-all portfolio is the one that out-performs even the benchmark, while the FoF for *delta* higher than 10% outperforms the benchmark, decreases the maximum drawdown and reduces the volatility. The increase in delta implies a relative increase in performance and Sharpe ratio as well as a decrease in the maximum drawdown.

Table 14: $EW - FoF$ for different deltas (S) - Fama-French (3Factors)

	MSCI	EW-all	2k5	2k10	2k15	2k20	2k25	2k30
TotR	2.72	10.91	60.65	58.65	54.47	54.47	54.47	54.47
AvgR	0.36	1.39	6.37	6.20	5.84	5.84	5.84	5.84
Vol	21.28	15.15	25.82	20.04	19.35	19.35	19.35	19.35
IR	0.02	0.09	0.22	0.30	0.29	0.29	0.29	0.29
MDD	-83.93	-82.21	-60.63	-67.93	-64.17	-64.17	-64.17	-64.17

Source: Fund selection done among a sample of 89 European Equity Mutual Funds daily quotes for a period from Jan 2005- Feb 2012. Asset pricing model: *Fama-French* (3Factor). The selection method used: FDR. Number of bootstraps: $B = 2000$. $\delta \in [0.05, 0.3]$. Performance indicators: total return, average return, Sharpe Ratio, Volatility, Max Drawdown, L-moments; computations by the author.

The performance for the $EW - FoFs$ in the sub-sample S_1 is shown in table 14 is completely in line with results for $CAPM$. The best FoF performance starts for a level of significance $\delta = 15\%$[8]. Moreover, the increase in *delta* does not decrease the FoF performance. It is remarkable but not unexpected the level of maximum drawdown for *delta* being 5%, because of the fact that for this significance level, in this sample, we select only one fund.

Table 15: $EW - FoF$ for different deltas (S_1) - Fama-French (3Factors)

	MSCI	EW-all	2k5	2k10	2k15	2k20	2k25	2k30
TotR	8.39	14.91	31.07	29.56	28.21	25.26	22.31	22.35
AvgR	2.17	3.74	7.27	6.96	6.68	6.05	5.41	5.42
Vol	15.12	10.03	13.31	13.49	13.17	12.98	13.18	13.00
IR	0.12	0.30	0.44	0.42	0.41	0.38	0.33	0.34
MDD	-37.75	-36.00	-29.79	-29.78	-31.34	-31.90	-31.97	-32.06

Source: Fund selection done among a sample of 89 European Equity Mutual Funds daily quotes for a period from Jan 2005-Aug 2008. Asset pricing model: *Fama-French* (3Factor). The selection method used: FDR. Number of bootstraps: $B = 2000$. $\delta \in [0.05, 0.3]$. Performance indicators: total return, average return, Sharpe Ratio, Volatility, Max Drawdown, L-moments; computations by the author.

We have also analyzed another pricing model based on the 3 factors of Fama and French and a momentum factor such as in Carhart (1997) [5]. We observe that the number of selected funds taking the whole dataset is reduced compared to other pricing models while the number of funds selected for the first sample is higher than the previous models.

We analyzed the performance of different portfolios based on these selection in tables 14, 15 and 17, and we conclude that the risk adjusted performance measures of these portfolios are higher than the benchmark index. The drawdown measures are also improved and all the funds present larger positive performance.

[8] For other levels of delta we found the same FDR selction and thus the same portfolios till $\delta = 30\%$

Table 16: The variation of the FDR fund selection for the pre-crisis sample (Carhart)

delta	The selected Funds	Nr. selected funds
0.05	1	1
0.1	1	1
0.15	1 50	2
0.2	1 41 50	3
0.25	1 3 41 50	4
0.3	1 3 41 48 50	5

Source: Fund selection done among a sample of 89 European Equity Mutual Funds daily quotes for a period from Jan 2005-Feb 2012. Asset pricing model: *Carhart*(4Factor). The selection method used: FDR. Number of bootstraps: $B = 2000$. $\delta \in [0.05, 0.3]$; computations by the author.

Table 17: $EW - FoF$ for different deltas (S_1) - Carhart (4Factors)

	MSCI	EW-all	2k5	2k10	2k15	2k20	2k25	2k30
TotR	8.39	14.91	13.00	13.00	7.69	15.83	13.97	15.55
AvgR	2.17	3.74	3.29	3.29	1.99	3.95	3.52	3.88
Vol	15.12	10.03	11.99	11.99	9.69	9.53	10.35	10.95
IR	0.12	0.30	0.22	0.22	0.17	0.35	0.28	0.29
MDD	-37.75	-36.00	-36.82	-36.82	-39.07	-32.16	-34.85	-34.69

Source: Fund selection done among a sample of 89 European Equity Mutual Funds daily quotes for a period from Jan 2005-Feb 2012. Asset pricing model: *Carhart* (4Factor). The selection method used: FDR. Number of bootstraps: $B = 2000$. $\delta \in [0.05, 0.3]$. Performance indicators: total return, average return, Sharpe Ratio, Volatility, Max Drawdown, L-moments; computations by the author.

Back-testing results As a continuation of the in-sample analysis we consider another in-sample exercise with rolling portfolios. First of all we use the FDR procedure to select the "skilled" fund managers during the whole available period. The selected funds are used to create the EW, ERC and MV portfolios actualizing the respective weights every 3 months and using one year past observations to estimate the parameters needed for each portfolio construction. We fix the parameter δ to 0.2 and we build the different portfolios based on the FDR selection. The set of funds to build those portfolios is thus the same. We conduct this back-test type procedure in order to see the impact of the different portfolio construction methods and to be able to compute performance measures for funds constructed with the FDR selection method. There is no difference between the $EW-FoF$ in-sample and any EW-back-testing portfolio but the MV-portfolio and ERC-portfolios are different if the rebalancing period and the window of estimation changes. Table **??** gives performance indexes of the FoF constructed based on different portfolio strategies based on the selected funds by FDR and *alphas* generated by $CAPM$, Fama-French (FF) and Carhart (CA) pricing models. Focusing in the whole sample results. In the $CAPM$ pricing model, The best strategy is the $EW-FoF$ for both, the $CAPM$ (Table **??**a) and Fama-French (Table **??**d). There is a higher performance in terms of total return and average return which are positive in terms of $CAPM$ model and negative but still higher than the other strategies for the Fama-French model. The volatility for the $EW-FoF$ is almost 50% lower than the $ERC-FoF$ while the Sharpe ratio is higher. The S_1 sub-sample $ERC-FoF$ has a higher performance compared to the other $FoFs$ for both pricing models. Compared to $EW-FoF$ and $MV-FoF$ the average performance of the three portfolios is similar but the volatility decreases by more than 50% as well as the maximum drawdown which is three times lower, therefore the Sharpe ratio increases by more than 50%. The total performance ($TotR$) of the three portfolios is higher than the benchmark. Based on the most difficult sub-period of our sample, the post-crisis sub-sample, the $EW-FoF$ portfolio remains having the highest performance followed by the $MV-FoF$. It is impossible to conclude in the case of Fama-French asset pricing model because of the selection of none of the funds during this period by the FDR procedure. This result seems logical as the possibilities to generate alphas decrease during the crisis period, no selection within the three factors framework suggest that it is not possible to generate any positive sustainable return during the crisis period that is not explained by the three main factors of the model.

Tables: Performance of back-testing EW-FoF, ER-FoF and MV-FoF

Table 18: tab: CAPM-S

	EW	ER	MV
TotR	34.09	34.02	57.47
AvgR	3.94	3.94	6.10
Vol	18.67	18.64	15.07
IR	0.21	0.21	0.40
MDD	-68.75	-68.63	-56.10

Table 19: tab: CAPM-S1

	EW	ER	MV
TotR	25.96	26.13	38.41
AvgR	6.20	6.24	8.74
Vol	12.89	13.00	10.93
IR	0.38	0.38	0.63
MDD	-32.29	-31.89	-25.90

Table 20: tab: CAPM-S2

	EW	ER	MV
TotR	13.15	13.22	27.55
AvgR	3.33	3.34	6.55
Vol	23.34	23.26	19.69
IR	0.17	0.17	0.38
MDD	-47.78	-48.00	-35.68

Table 21: tab: Fama-French-S

	EW	ER	MV
TotR	54.47	55.55	61.84
AvgR	5.84	5.94	6.47
Vol	19.35	19.34	17.16
IR	0.29	0.30	0.37
MDD	-64.17	-63.79	-58.11

Table 22: tab: Fama-French-S1

	EW	ER	MV
TotR	25.26	25.41	36.84
AvgR	6.05	6.09	8.43
Vol	12.98	12.98	10.89
IR	0.38	0.38	0.62
MDD	-31.90	-31.30	-25.61

Table 23: tab Carhart - S1

	EW	ER	MV
TotR	15.83	16.78	21.96
AvgR	3.95	4.17	5.34
Vol	9.53	8.87	8.38
IR	0.35	0.39	0.53
MDD	-32.16	-30.93	-26.91

Source: Back-testing in-sample exercise, with rolling window if 1 year, for sample S, S_1 and S_2. Asset pricing model: $CAPM$ (on the left) *Fama-French* and *Carhart* (on the Right). The selection method used: FDR. Number of bootstraps: $B = 2000$. Delta = 20%. Portfolio strategy: Equally-Weighted (EW), Equally-Risk-Contribution (ERC) and Minimum-Variance (MV). Performance indicators: total return, average return, Sharpe Ratio, Volatility, Max Drawdown, L-moments; computations by the author.

Out-of-sample results

The in-sample analysis provided some useful conclusions as for the different types of portfolio strategies and different pricing model to use for each sample. But these results cannot be used in practice because there are ex-post results. We conduct therefore an out-of-sample study for the most interesting results in-sample. We considered the estimation of alphas with $CAPM$ selection procedure with a parameter *delta* set to 0.1 and 0.2 and a number of bootstraps of 500. We used 250 observations to estimate the asset pricing models and determine the set of FDR selection of funds at a given date. An equally weighted portfolio is constructed based on the set of selected funds. This procedure is repeated every three months (75 observations) in order to update the FDR selection based on the three types of asset pricing models. The performance of these strategies is evaluated and the results are presented in table 5.1.

Table 24: Out-of-sample results

	CAPM10	CAPM20	FF3.10	FF3.20
TotR	4.08	6.75	-0.33	9.10
AvgR	0.63	1.04	-0.05	1.38
Vol	24.07	26.59	24.01	24.62
IR	0.03	0.04	-0.00	0.06
MDD	-33.90	-42.23	-34.21	-34.59

Source: Out-of-sample exercise, with rolling window strategy: "observe one year, invest during three months" for the total sample S. Asset pricing model: $CAPM$ (1factor). The selection method used: FDR. Number of bootstraps: $B = 2000$. Delta= 10% and Delta= 20%. Portfolio strategy: Equally-Weighted (EW). Performance indicators: total return, average return, Sharpe Ratio, Volatility, Max Drawdown, L-moments; computations by the author.

We observe that for EW portfolio based on an FDR from $CAPM$ with a *delta* of 0.1 has a total positive return for the period of 4.1% and a volatility of 24%. For the case with parameter delta= 0.2, the performance is higher 6.7 probably due to the inclusion of a higher number of funds at each rebalancing day and a similar volatility of 26.6%. It is interesting the low performance of the EW portfolio constructed out-of-sample using the Fama-French pricing model with delta being 0.1. This low performance is due to the very low number of funds selected during the post-crisis period. Moreover, during the last period of end 2011, the selection procedure does not filter any fund at all. Thus, the total out-of-sample portfolio is quite poor due to the miss of diversification. The results become quite comparable to the expectations for delta being 0.2 where the portfolio moreover out-performs other constructed portfolios of the same kind. These results compared to the in-sample results presented in previous sections show that the out-of-sample results deliver a lower performance which was expected as in the in-sample estimation the selection procedure is based on an *ex-post* analysis. The positive performance of the out-of-sample strategies compared to the negative performance of the Benchmark for the same sample for similar levels of risk let us conclude that these fund selection procedure could be useful anyway in the multi-management investment industry.

5.2 FWER Findings

We test the in-sample and out-of-sample results by using not only the FDR but also the $FWER$ procedure. This stepwise procedure intends to capture the funds with high performance situated in the tails of the distribution. The confidence level γ differently form the FDR is used to fix a percentage of the positive tail of *alpha* distribution above which the funds are selected. The $FWER$ procedure is much more restrictive than the FDR. The number of the selected funds is smaller but the set of the selected ones is in line with the FDR selections. For this reason, we focus our study on the results of the FDR selection since more funds selected makes more sense and allows us to calculate the notorious portfolios and compare them easily. Moreover, the $FWER$ being a bootstrap stepwise procedure is more time costly procedure than the FDR.

6 Conclusion

We conducted a FDR and $FWER$ selection procedure to determine the set of fund managers having a "true" positive *alpha* return or being "skilled" fund managers. Both methods are based on bootstrap procedures and we conducted different tests to determine the optimal number of bootstrap needed to find a stable set of selected funds as well as to determine the optimal parameter that should be used within our database to select an optimal number of funds. The FDR selection was based on the t-statistic of the *alphas* estimated from different asset pricing models, mainly the $CAPM$ and the Fama and French three factor model were compared and, as expected, the number of selected funds based on a model with more factors was reduced. With the different FDR and $FWER$ selections we used different methods to build optimal portfolios and we compared several performance measures for the computed portfolios. All our tests were repeated for different static samples, as well as for rolling samples in order to study the stability of the portfolios that were generated and their performance. Our goal was to evaluate the FDR and $FWER$ selection procedures and to determine if these methods could be useful to be applied to select funds for the FoF industry. The main results of this study include the stability of FDR sets and the increasing number of funds as the parameter delta is increased, the higher performance of several portfolio strategies compared to the benchmark and the lower performance of out-of-sample strategies compared to in-sample. Moreover, the fund selection procedures were useful not only to exclude lucky managers from the basic estimation process but also to selected missed skilled ones. We conclude that both techniques are useful to select funds with "true" *alphas*. The FDR method being preferred for being more flexible and easy to compute and for selecting a larger number of funds which was useful to evaluate different methods of portfolio construction. This study can be completed with other asset pricing models as well as other ways to build portfolio in order to have more information about the resulting portfolios based on these selection procedures. We could also apply this methodology to a larger set of European funds or other databases and it could also be useful to extend the type of funds in the database to Hedge Funds as a way to verify robustness check of our results. Results in terms of more robust portfolio performances, L-moments, are yet to be implemented. Finally, we are still working in trying to check whether these selection procedures holding on some particular information could be used as investment strategies.

References

[1] D. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 49:817–858, 1991.

[2] D. Andrews and J. Monahan. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 60:953–966, 1992.

[3] D. Avramov and R. Wermers. Investing in mutual funds when returns are predictable. *Journal of Financial Economics*, 81:339–377, 2006.

[4] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57(1):289–300, 1995.

[5] M. Carhart. On persistence in mutual fund performance. *Journal of Finance*, 52:57–82, 1997.

[6] K. Cuthbertson and D. Nitzsche. Performance, stock selection and market timing of the german equity mutual fund industry. *Working Paper, Cass Business School*, 2010.

[7] B. Lehmann D. Blake and A. Timmermann. Performance measurement using multi-asset portfolio data: A study of uk pension funds. *Pension Institute London*, pages 1986–1994, 1999.

[8] E. Fama and K. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56, 1993.

[9] E. Fama and K. French. Lucky versus skill in the cross-section of mutual fund returns. *Working Paper*, 2010.

[10] A. Shleifer J. Lakonishok and R. Vishney. The structure and performance of the money management industry. *Brookings Papers on Economic Activity*, pages 339–391, 1992.

[11] N. Jegadeesh and S. Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91, 1993.

[12] M. Jensen. The performance of mutual funds in the period 1945-1964. *Journal of Finance*, 23:389–416, 1968.

[13] D. Nitzsche K. Cuthbertson and N. Sullivan. Uk mutual fund performance: Skill or luck? *Journal of Empirical Finance*, 15:613–634, 2008.

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

[15] S. Titman M. Grinblatt and R. Wermers. Momentum investment strategies, portfolio performance and herding: A study of mutual fund behavior. *American Economic Review*, 85:1088–1105, 1995.

[16] W. Newey and K. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708, 1987.

[17] L. Pastor and R. Stambaugh. Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics*, 63:315–350, 2002.

[18] R. Wermers R. Kosowski, A. Timmermann and H. White. Can mutual fund stars really pick stocks? new evidence from a bootstrap analysis. *Journal of Finance*, 61:2551–2595, 2006.

[19] J. Romano and M. Wolf. Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282, 2005.

[20] T. Roncalli S. Maillard and J. Tieletche. The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management*, 36(4):60–70, 2010.

[21] R. Wermers. Mutual fund performance: an empirical decomposition into stock picking talent, style, transactions costs, and expenses. *Journal of Finance*, 55:1655–1703, 2000.

[22] M. Wolf and D. Wunderli. Fund-of-funds construction by statistical multiple testing methods. *Working Paper, Institute of Empirical Research in Economics 445, Forthcoming in Journal of Behavioral Finance*, pages 1655–1703, 2009.

7 Appendix A

The Estimation Procedure of the FWER Approach

Using the raw *alpha* estimate to measure the individual performance and to rank the funds may create problems because it does not account for the varying risks taken by various fund managers. This problem is reduced while instead of considering the raw *alpha* estimations, we consider the corresponding t-statistic. In this case the "student" test statistic considers varying risks, since a larger risk will be reflected by a larger standard error $\hat{\sigma}_n$, HAC standard error employing kernel estimation techniques proposed by Andrews and Monahan (1992) [2] is used. The condition of the $FWER$ approach in terms of t-statistic test is the following:

$$\begin{cases} H_0 : \hat{t}_n > \hat{d} \\ H_1 : \hat{t}_n \leq \hat{d}. \end{cases} \tag{6}$$

The sketch of the algorithm for estimating the critical value $\hat{d}$ is given as such:

1. generate B bootstrap vectors for each fund, which represent the artificial returns. Denote them as $\left(r_n^{*,1}, , r_n^{*,B}\right)$.

2. from each of the bootstrap data vector $r_n^{*,b}$, where $1 \leq b \leq B$ we compute the *alphas* and the standard errors for each fund. Thus, for one bootstrapped vector $r_n^{*,b}$ the *alphas* and their standard errors are denoted respectively as $\left(\alpha_{n,1}^{*,b}, , \alpha_{n,N}^{*,b}\right)$ and $\left(\sigma_{n,1}^{*,b}, , \sigma_{n,N}^{*,b}\right)$.

3. repeat step 1 and 2 for each fund available. So, at the end of this step, one should have the respective matrices of *alphas* and *sigmas.*

4. for $1 \leq b \leq B$, compute the maximum over the funds of one bootstrapped matrix $\mathbf{r}^{*,b}$ such as:
$$max_n^{*,b} = \max_{n \in \mathbf{N}} \left\{ \frac{\hat{\alpha}_n^{*,b} - \hat{\alpha}_n}{\hat{\sigma}_n^{*,b}} \right\}. \tag{7}$$

5. we compute $\hat{d}$ as the $(1 - \delta)$ quantile of the B values $\overset{*,1}{\underset{n}{\max}}, \ldots, \overset{*,B}{\underset{n}{\max}}$.

It consists in a stepwise procedure which first detects the "skilled" managers over the total number of funds. Then the procedure is repeated but this time the difference between the total number of funds and the ones managed by skilled managers is considered. As the number of funds decreases in the stepwise procedure, the test statistic t_n will remain the same. However, the $\hat{d}$ will be at most as large as the one of the previous step of the iteration. This allows us to do some further rejections. The stepwise selection procedure of the "skilled" managers is repeated again until no further rejections result any more. The price one has to pay for replacing d by $\hat{d}$ is that control of the $FWER$ is replaced by asymptotic control of the $FWER$.

8 Appendix B

The Estimation Procedure of the FDR Approach

As in the $FWER$ approach, the t-statistic used is not calculated as ratio of *alpha* and *sigma* coming both from the OLS regression. Here the Newey-West (1987) [16] is used to provide an estimate of the covariance matrix of the parameters of a regression-type model when this model is applied in situations where the standard assumptions of regression analysis do not apply. The estimator is used to try to overcome autocorrelation, or correlation, and heteroskedasticity in the error terms in the models. This often corrects the effects of correlation in the error terms in regressions applied to time series data. As for the estimation of the alpha Barras *et al.* (2010) [14] propose to use the OLS regression results. The bootstrap procedure approximates the distribution of $\left(\hat{t}_n - t_n\right)$ by the distribution of $\left(\hat{t}^*_n - \hat{t}_n\right)$ where t_n is the fund t-statistic and the $\hat{t}_n$ is the bootstrapped t-statistic. Let us consider $\hat{\epsilon}_{n,t}$ the estimated residuals that we get from the b bootstrap iterations with replacement $(b = [1, , B])$. Then after re-sampling the residuals $\hat{\epsilon}^{*,b}_{n,t}$, we create a new time series of excess returns $\hat{r}^{*,b}_{n,t}$, by imposing that r_n is zero. We compute the $\alpha^{*,b}_n$ and $\sigma^{*,b}_{\alpha_n}$, by regressing $r^{*,b}_{n,t}$,on the factors. So that we can obtain the bootstrap t-statistic $\hat{t}^{*,b}_n$. The procedure explained above is repeated B times. Since we consider the two-sided, equal-tailed test, the bootstrapped p-value of each fund n is computed as follows:

$$\hat{p}_n = 2\min\left(B^{-1}\sum_{b=1}^{B}\#\left\{\hat{t}^{*b}_n > \hat{t}_n\right\}, B^{-1}\sum_{b=1}^{B}\#\left\{\hat{t}^{*b}_n < \hat{t}_n\right\}\right), \tag{8}$$

Thus, we consider as selected funds the ones satisfying the following: $\hat{p}_n < \gamma and \alpha_n > 0$ Beyond the selection procedure, Barras *et al.* (2010) [14] calculate the following estimate:

$$\widehat{FDR}^{+}(\delta) = N\hat{\pi}_0(\lambda)\delta\left[\#\left\{\hat{p}_n < \gamma\right\}\right]^{-1} = \hat{F}_\delta\left[\hat{R}_\delta\right]^{-1}, \tag{9}$$

In order to compute the $\hat{\pi}_0$ we can exploit this information, specifying the exact distribution under H_1. The estimate $\hat{\pi}_0(\lambda)$ of non-performing funds is calculated in the approach such as:

$$\hat{\pi}_0(\lambda) = \#\left\{\hat{p}_n < \lambda\right\}\left[(1-\lambda)N\right]^{-1} = \hat{W}(\lambda)\left[(1-\lambda)N\right]^{-1}, \tag{10}$$

where $\hat{W}_n$ is the number of estimated p-values bigger than the threshold. The estimator $\hat{W}_n$ depends on the parameter γ that is determined by a bootstrapping, proposed by Storey (2002) [?]. A bootstrap algorithm is used to determine the optimal threshold γ.

9 Appendix C

caption: The variation of the FDR fund selection for the pre-crisis sample $S1$ (Carhart)

Source:

caption: Performance of Benchmark, EW and $EW - FoF$ for different deltas ($S1$) - Carhart (4 Factors)

Source: Fund selection done among a sample of 89 European Equity Mutual Funds daily quotes for a period from Jan 2005-Aug 2008. Asset pricing model: *Carhart* (4Factor). The selection method used: FDR. Number of bootstraps: $B = 2000$. $\delta \in [0.05, 0.3]$. Performance indicators: total return, average return, Sharpe Ratio, Volatility, Max Drawdown, L-moments; computations by the author.

Table 25: List of the European Mutual Funds contained in the database used in this study (EUR)

Nr.	Fund Name	ISIN
1	All Europe T	AT0000721444
2	NBG Intl Fds Socially Responsible A	LU0165283150
3	Sstpankki Eurooppa A	FI0008806591
4	ANM Anima Europa	IT0001415287
5	AXA Rosenberg Pan-Eurp Enh Idx Alp A	IE0033609839
6	Amundi Actions Europe P Acc	FR0010013763
7	MEAG EuroInvest A	DE0009754333
8	SSgA Europe Enhanced Equity Fund	FR0000986747
9	Bankia Bolsa Valor Europea FI	ES0138840030
10	Allianz Invest Aktienfonds A	AT0000823299
11	Santander Solidario Dividendo Europa FI	ES0114350038
12	Invesco Actions Euro E	FR0010135871
13	E. Rothschild Europe Value A	LU0112689434
14	Montepio Accioes FI	PTYMGCLM0009
15	Essor Europe	FR0000286411
16	BG Long Term Value Z	FR0010137646
17	DNCA Value Europe C	FR0010058008
18	AXA Europa	DE0009775643
19	AXA Rosenberg Pan-Eurp Eq Alp A EUR	IE0008365730
20	UniValueFonds: Europa A	LU0126314995
21	Montsegur Perspectives C	FR0010109140
22	Epsilon QValue	IT0001496097
23	Athena European Equity Acc	BE0156531700
24	Seligson Co Eurooppa- indeksirahasto A	FI0008801774
25	AC Inversion Selectiva FI	ES0106949037
26	Pioneer Fds (A) Select Europe Stock A	AT0000856042
27	Pro Fonds (LUX) Premium B	LU0106484834
28	Medivalor Europeo FI	ES0162022034
29	CompAM European Equity A	LU0165045302
30	BIM Azionario Europa	IT0003391676
31	Kapitalfonds LK Aktien Europa G	LU0172200718
32	Nordea Eurooppa Kasvu	FI0008800446
33	Natixis Actions Europe Conviction I	FR0010346429
34	H&A Lux Equities - VALUE Invest B	LU0100177426
35	Allianz RCM Deep Value Europe A EUR	DE0008479544
36	ESPA Stock Europe-Value A	AT0000659230
37	Nouvelle Europe II A	AT0000856950
38	Nemesis Eurp Val Euro Adv Acc	IE00B3THSZ36
39	Konzept Europa plus	DE0009780288

Nr.	Fund Name	ISIN
40	Rinascimento GAMES Arbit Mom Appr A	LU0091602218
41	BBV Invest Union	DE0009750018
42	MPC Competence-Europa Methodik AMI P(t)	DE0007248627
43	Kempen (Lux) European High Div A	LU0427929939
44	FBG Europe Equity	CH0008249739
45	Multifondo Europa FI	ES0138614039
46	Oddo Valeurs Rendement A	FR0000989758
47	Consultinvest Azione	IT0001076626
48	Barclays IF (Lux) Eurp ex-UK Alpha A Inc	LU0012163928
49	Handelsbanken Eurooppa Indeksi A	FI0008805742
50	Barclays Bolsa Europa FI	ES0138596038
51	Macquarie MS Equities Western Europe T	AT0000819792
52	Fideuram Equity Europe	LU0096628044
53	HSBC Valeurs Haut Dividende A Acc	FR0010043216
54	MFS Meridian Europ Value A1 EUR	LU0125951151
55	Santander Aces Europa FI	PTYSADLM0008
56	Odey Pan European	IE0032284907
57	Kempen European High Dividend	NL0000293348
58	UBS (CH) EF European Opportunity P	CH0002791702
59	Danske Invest Europe High Dividend A	LU0123484957
60	JB EF Europe-EUR A	LU0026740760
61	AZ Fd1 European Trend - AZ Fund A	LU0107996786
62	Monceau Slection France Europe	FR0007016720
63	Postbank Europa	DE0009770289
64	Ofi Leader ISR I	FR0000981441
65	BGF European Value A2 EUR	LU0072462186
66	Metropole Slection A	FR0007078811
67	R Conviction Europe C	FR0010784835
68	JPM Europe Eq A (dist)-EUR	LU0053685029
69	Fundquest Europe Multimanagers Acc	FR0010376830
70	XT Europa T	AT0000697065
71	SG Actions Europe Slection	FR0010311993
72	PACTO Azionario Europa A	IT0003028948
73	BPER Intl SICAV Equity Europe	LU0085741386
74	CPR European Opportunities P	FR0010447573
75	Santander E.F. Cclico FI	PTYSAPLM0004
76	UniInstitutional European MinRisk Eq	DE0009750554
77	4Q-European Value Fonds Universal	DE0009781989
78	Cahispa Europa FI	ES0124541030
79	Santander SICAV European Dividend A	LU0082927103
80	Pioneer Azionario Val Eurp a dist. A	IT0001029864
81	DVG Merkur Fonds 1	DE0008493370

Nr.	Fund Name	ISIN
82	UniEuropa -net-	DE0009750232
83	BNP Paribas Europe Dividende Acc	FR0010077362
84	Ofi Palmars Equity	FR0007041546
85	Pictet-European Sust Eq-I EUR	LU0144509550
86	Mediolanum Europa R.V. S FI	ES0165128036
87	Raiffeisen-Europa-Aktien R A	AT0000986377
88	Banesto Dividendo Europa FI	ES0113109039
89	Degroof Eqs Europe Behavioral Val B	LU0006098676