Dynamic return-based classification of European mutual funds

April 30, 2015

Philipp Gerlach (corresponding author)
Finance Department, Goethe University
Grueneburgplatz 1 (Uni-PF. H 23)
Frankfurt am Main, Germany
gerlach@finance.uni-frankfurt.de

Raimond Maurer
Finance Department, Goethe University
Grueneburgplatz 1 (Uni-PF. H 23)
Frankfurt am Main, Germany
maurer@finance.uni-frankfurt.de

This research was conducted with support from the German Investment and Asset Management Association (BVI). Opinions and errors are solely those of the authors and not of the institutions with whom the authors are affiliated.

Dynamic return-based classification of European mutual funds

Philipp Gerlach Raimond Maurer

Abstract

In this study, we analyze the monthly fund migrations in a return-based classification scheme for European mutual funds. Accounting for the time-varying commonalities in returns, we derive a classification scheme consisting of four layers with two, six, 15, and 20 groups respectively. In the two-group solution fund migrations are low while in the classification with six groups migrations are low during normal times but high during the financial crisis for stock funds and during the debt crisis for bond funds. Fund migrations in the classification with 15 and 20 groups are high overall. In line with prior research, we confirm that return-based classifications are better able to explain the cross-sectional differences in mutual fund returns than existing industry classifications based on portfolio holdings and self-declared investment objectives.

JEL Classification: C63, G23

Keywords: Fund Migrations, Hierarchical K-Means, Mutual Funds, Return-Based Classification, Transition Probabilities

1 Introduction

Prior studies show that return-based classification schemes for mutual funds are superior to traditional classifications based on portfolio holdings and self-declared investment objectives for three reasons. First, return-based classification schemes are not prone to misclassification, i.e., the allocation of a fund to a peer group with distinct attributes, as a result of fund managers trying to get a superior ranking relative to their peer group, or of ambiguous classification schemes, or because changes in the investment objective of the fund do not enter the classification scheme (e.g. diBartolomeo and Witkowski, 1997; Brown and Goetzmann, 1997, 2003; Dor et al., 2006). Second, return-based classifications are able to derive the optimal number of groups endogenously. Therefore, return-based classifications uncover distinct return profiles within a group of funds having the same investment objective (e.g. Brown and Goetzmann, 1997, 2003; Gerlach and Maurer, 2014) and they identify redundant return profiles of groups with different investment objectives (e.g. Moreno et al., 2006). Third, return-based classifications are better able to explain return differences among the funds (e.g. Brown and Goetzmann, 1997, 2003; Gerlach and Maurer, 2014).

Yet, these studies are silent about the dynamics of return-based classification schemes even though this is an important aspect of the reliability of these classification schemes. If return-based classification schemes were static over time, they would provide no information about changes in the return profiles of the funds. On the other hand, ongoing reporting would be difficult, and expectations about the future classification structure could not be formed if these approaches exhibited little continuity over time. Therefore, for a practical application, classification schemes should show a high degree of consistency over time and should adapt to changes in the return profiles of the funds.

The dynamics of classification schemes for mutual funds include two components: changes in the number of categories or groups and changes in the allocation of the funds to these groups. In this study, we focus on the second component.

We use the hierarchical K-means algorithm (Gifford, 2014; Lamrous, 2006; Lee, 2012; Nister and Stewenius, 2006), a recursive application of the standard K-means algorithm, to classify a sample of 38,073 mutual funds registered for sale in Europe over the period of January 2000 – January 2014. Accounting for the time-varying commonalities in mutual fund returns, we find evidence that a return-based classification scheme for European mutual funds consists of four layers with two, six, 15, and 20 groups. The investment objectives and the return time series of the funds indicate that the first layer separates stable from variable return funds, while the second layer differentiates between the major asset classes (bond, equity, and money market), each asset class either with a European or a global focus. The fourth and fifth layers divide the asset classes and economic areas into finer subgroups, but they also form a group of funds following active strategies that aim for downside protection. To analyze the dynamics of the developed classification scheme, we estimate the monthly transition probabilities of funds remaining in their respective group or migrating to a different group. We find that the two-group solution is highly stable, with a 99% average probability that the funds remain in their respective groups; therefore, almost no fund migrations occur over time. Even though the average probability of remaining significantly decreases to 93% in the six-group solution, the monthly fund migrations show how the return-based classification adapts to the current market situations by relocating funds between the groups. The 15- and 20-group solutions, with an average probability of remaining of 82% and 77% respectively, are unstable. In almost every month, there is at least one entire group of funds that migrates to another group. We also estimate the quality of the classification scheme to reflect the differences in fund returns. Compared to an existing industry classification scheme based on self-declared investment objectives, our return-based classification scheme is better able to explain the cross-sectional variation of fund returns, and it also provides more adequate peer-group benchmarks.

The prior literature on fund classification can be divided into three streams. In the first stream, the number of fund categories and the fund memberships are estimated from observable fund attributes such as realized returns or portfolio holdings. Then the estimated classifications are compared to existing industry classifications, which are based on the investment objectives as stated by the fund managers. These studies address the quality of existing industry classifications as provided by rating agencies. While Bailey and Arnott (1986) and Lisi and Otranto (2008) show that the differences in the fund returns are well reflected by industry classifications, Gibson and Gyger (2007), Kim et al. (2000), LeClair (1974), and Marathe and Shawky (1999) show that a significant percentage of funds have attributes different from their peer groups with the same investment objectives. Furthermore, Marathe and Shawky (1999) and Moreno et al. (2006) find that industry classifications have redundant categories; even though funds claim to have different investment objectives, they have similar attributes. In the second stream, existing industry classifications are refined without changing the number of categories, by reassigning misclassified funds to a more adequate peer group. The studies by diBartolomeo and Witkowski (1997) and Dor et al. (2006) confirm the problem of misclassified funds. Besides the ambiguity of industry classifications as a potential source for misclassification, diBartolomeo and Witkowski (1997) find evidence of systematic misclassification. Dor et al. (2006) show that after misclassified funds are reassigned, they remain in their proper category. In the third stream, fund classifications are estimated based on historical fund returns. Then, the return indices of the groups are used for further style analysis and performance attribution. These approaches follow the idea that funds with similar return series hold the same types of assets and are managed according to the same trading strategy (Brown and Goetzmann, 1997, 2003; Fung and Hsieh, 1997; Gruber, 2001). Therefore, investment styles that try to beat a specific asset class index and, more importantly, styles that represent dynamic trading strategies can be extracted directly from publically available fund returns.

We contribute to the existing literature on fund classification in several ways: First, we analyze an extensive data set of European mutual funds that has not been studied before. Second, we introduce an intuitive return-based classification approach that is more capable of handling extensive mutual fund samples. Third, this is the first study to estimate the dynamics of the fund memberships in a return-based classification scheme over time.

2 Methodology and Data

2.1 Methodology

We apply the hierarchical K-means algorithm to classify mutual funds based on their realized return series. The algorithm stepwise partitions a data set into finer classifications by recursively applying the standard K-means algorithm. It is designed to classify high dimensional data and is therefore highly suitable in our context of classifying a large sample of mutual funds based on their return time series. The hierarchical K-means clustering is appealing for its simplicity, low processing time, and resulting classification quality. Different variants of the hierarchical K-means algorithms exist in the literature (Gifford, 2014; Lamrous, 2006; Lee, 2012; Nister and Stewenius, 2006). Therefore, we briefly specify our approach as follows:

First, we select the group of funds with the highest sum of squared errors as a measure for the most heterogeneous group. Second, we partition the selected group into subgroups by applying the standard K-means algorithm. These two steps of selecting and dividing are applied recursively to refine the classification.

The standard K-means algorithm is one of the most common classification approaches (Jain, 2010; Steinley, 2006). Given a set of $\mathrm{i} = 1, \dots, N$ funds, where each fund is characterized by its historical time series of monthly returns $\sum_{t=0}^{t-59} r_{i,t}$, the K-means algorithm classifies the funds into $\mathrm{k} = 1, \dots, K$ groups, $K \in \{1,2, \dots, 20\}$ such that the sum of squared errors (SSE) of the group return indices μ_k and its member funds is minimized

$$SSE(K) = \sum_{t=0}^{t-59} \sum_{k=1}^{K} \sum_{\substack{i=1 \\ i \in k}} \left(r_{i,t} - \mu_{k,t}\right)^2 \rightarrow min! \tag{1}$$

where

$$\mu_{k,t} = \frac{1}{|k|} \sum_{\substack{i=1 \\ i \in k}} r_{i,t}.$$

The K-means algorithm starts with an initial classification and iteratively relocates funds between the groups such that the sum of squared errors converges to a local minimum. The algorithm requires two parameter specifications: the number of groups K and an initial classification to start the relocation process. Among the proposed initialization approaches in the literature, the most prevalent technique is to use multiple runs of the K-means algorithm, each one with random starting points. This approach suffers from an extensive computer processing time when the size of the dataset is large. Therefore, we follow Steinley and Brusco (2007) and use the solution resulting from a Ward's hierarchical clustering procedure as an initial starting point (Ward, 1963). To estimate the appropriate number of groups, we follow Lamrous and Taileb (2006) and Gibson and Gyger (2007) by using the silhouette statistic (Rousseeuw, 1987).

The K-means objective of minimizing the sum of squared errors is scale-dependent (Everitt, 2011). Therefore, we test the effect of raw returns, normalized returns, where the returns of each fund are scaled by the inverse of the estimated standard deviation, and standardized returns, where the returns of each fund are demeaned and then scaled by the inverse of the estimated standard deviation, on the hierarchical K-means classification results. We also analyze the effect of the return time series length with 24, 36, and 60 months on the hierarchical K-means classification results. Finally, we choose a time series length of 60 months and standardize the return time series of each mutual fund such that the returns have a mean of zero and a standard deviation of one. This removes return differences that result from leverage and ongoing costs differences. This data preparation results in most stable classifications.

The hierarchical K-means algorithm produces multiple classification solutions, a valid classification in each step.[1] To estimate which classifications are most appropriate to describe the differences in the mutual fund returns, we use the proportional reduction in errors. Let SSE_s be the sum of squared errors in step s; the proportional reduction in error is

$$PRE_s = 1 - \frac{SSE_s}{SSE_{s-1}}. \quad (2)$$

The proportional reduction in error measures the marginal gain of an additional K-means run in terms of the decrease in the sum of squared errors for the cost of an increase in the number of groups. We identify the classifications after which the proportional reduction in errors drops considerably to be appropriate for explaining the differences in the fund returns.

2.2 Mutual fund data

We classify a sample of mutual funds registered for sale in Europe. Our sample contains 38,073 primary share classes of active, merged, and liquidated mutual funds.[2] The data are provided by Lipper, a Thomson Reuters Company. Table 1 presents the distribution of the 13 different investment objectives listed by funds in our sample as reported in January 2014.

[Insert Table 1 here]

Of the funds examined, 1,255 (3%) are *absolute return* funds. These funds aim for absolute positive returns irrespective of market conditions. *Alternative funds* make up 590 (2%) of the funds. These funds follow the same strategies as hedge funds, but they are subject to stricter regulations regarding the exposures, leverage etc. Additionally, 6,846 funds (18%) investing in fixed income securities with an average maturity of more than one year are *bond* funds. With a predominant exposure to commodities, either directly by investing in physical commodities or indirectly by investing in structured securities or derivatives, 141 funds (<1%) are *commodity* funds. Of the remaining funds, 13,692 funds (36%) investing in stock markets are *equity* funds, and 1,516 funds (4%) are *guaranteed* funds, which guarantee the

[1] Industry classification schemes based on fund investment objectives also consist of multiple layers. For example, classifying funds into equity, bond, money market, and other is valid as a finer classification, where each asset class is further divided into the regional focus such as equity Europe, equity US, etc.

[2] To account for potential data errors, we remove all monthly returns below the 0.001th and above the 99.999th percentile. From the 3,003,281 initial returns entries, we remove 31 entries with values lower than -54.80% and 31 entries with values higher than 65.15%. We observed three types of errors in the dataset: a). monthly returns with values above the thousands, b).values around 90% after the launch of a fund, and c). values around -90% before a fund was closed.

principal and/or the dividend/interest at one or more predetermined dates. Only 4 funds (<1%) are *hedge* funds, which are typically less regulated. Also, 5,959 funds (16%) are *mixed asset funds*, which have a strategic mixture of variable income and fixed income securities, and 1,903 funds (5%) investing in fixed income securities with an average maturity less than a year are *money market* funds. Of the funds, 1,331 (3%) are *protected* funds, aiming at a minimum return whilst protecting from a downside risk. Additionally, 69 funds (<1%) investing in physical land property are *real estate* funds, and 638 funds (2%) are *target maturity* funds, which aim to maximize the total return at a predetermined date and are normally liquidated at the maturity date. Finally, 4,129 funds (11%) are unclassified, undisclosed, or have empty investment objective data entries. We subsume these funds as *unclassified*.

Figure 1 shows the total number of active funds (right axis) and the proportions of investment objectives (left axis) over time.

[Insert Figure 1 here]

The number of active funds almost doubles from 10,592 in January 2000, to 19,489 in January 2014; however, the number of funds does not increase monotonically over time, i.e., the number of funds increases to a peak of 20,889 in January 2009 and decreases subsequently. The proportions of the investment objectives are almost constant over time. Yet, the percentage of active funds that are unclassified decreases over time due to liquidations or mergers. In addition, absolute return funds, alternative funds, and target maturity funds increase their share in the European mutual fund market over time. Absolute return funds increase from 45 funds (0.42%) in January 2000 to 663 (3.40%) funds in January 2014, alternative funds increase from four funds (0.04%) to 399 funds (2.05%), and target maturity funds increase from 27 funds (0.25%) to 438 funds (2.25%). Still, the share of each of these three categories is low when compared to the traditional asset-class related categories such as bond, equity, or mixed asset.

3 Empirical results

The commonalities underlying mutual fund returns are time variant; therefore, a classification that reflects these commonalities is time variant as well. This is clear when we classify the mutual fund sample at two different dates.

Figure 2 shows the four initial steps of the hierarchical K-means algorithm performed in December 2004.

[Insert Figure 2 here]

In the first step, the algorithm divides the funds into two groups, which we label as stable returns and variable returns, based on the investment objectives of the member funds. In the following step, the algorithm refines the classification by separating the group of variable return funds into developed markets and emerging markets. In the third step, the group of funds with stable returns is divided into funds globally investing in bonds, funds investing in Euro-denominated bonds or bonds from European issuers, and money market funds. In the fourth step, the algorithm divides the developed markets group into 17 subgroups: equity Europe, equity Europe, equity core Europe, equity Europe/global, equity Nordic countries, equity Norway/Latin America/Denmark, equity UK, equity UK/global, equity Italy, equity Spain, equity US/global, equity US/Global, equity global/US, equity technology sector, equity pharma sector, equity financials sector, and mixed asset conservative.

Figure 3 shows the four initial steps of the hierarchical K-means algorithm performed in January 2014.

[Insert Figure 3 here]

In the first step, the algorithm divides the sample into funds with stable returns and funds with variable returns. In the second step, the classification process refines the variable returns group into developed markets and emerging markets. In the third step, the stable returns group is divided into bond and money market funds. In the fourth step, the developed markets group disaggregates into equity Europe and equity US/Japan.

Comparing these two disaggregation processes, as performed by the hierarchical K-means algorithm in December 2004 and January 2014, we see that the disaggregation rate differs. While the first two steps are identical, the subsequent steps differ in the number and composition of subgroups formed.

The time dependency of return-based classifications has two implications. The first implication is when the number of groups and the fund memberships are allowed to vary over time, return-based classification schemes can adapt optimally to current market trends. The second implication is when we hold the number of groups constant over time, it is necessary to account for changes in the commonalities in the mutual fund returns and hence for variations in the optimal number of fund groups. In the following analysis, we restrict the classification to have a constant number of groups over time.

3.1 Optimal classification scheme

To estimate the optimal disaggregation process on average, we run a series of monthly hierarchical K-means procedures, without pre-specifying the number of subgroups formed in each classification step. The algorithm chooses the optimal number of subgroups formed at each step based on the silhouette statistic. Table 2 presents the technical coefficients resulting from the 110 rolling, hierarchical K-means classification procedures.

[Insert Table 2 here]

Each row shows the stepwise, average classification results with the standard deviation in parentheses. The first column shows the stepwise disaggregation of the fund universe into an increasing number of groups. In the first classification step, the algorithm splits the fund universe into two groups – even though we allow the algorithm to choose up to 20 groups at each step. In the second and third steps, one group is divided into three subgroups. As indicated by the low standard deviations, these results are stable over time. In the subsequent steps, the disaggregation rate and its standard deviation increase. The last two columns show the effect of the stepwise classification refinements on the sum of squared errors. The explained variation measures the reduction in the sum of squared errors of a classification with K groups relative to the unclassified sample, whereas the proportional reduction in error measures the reduction in the sum of squared errors of a classification with K groups relative to the preceding classification of $K^*<K$ groups. We see that the marginal increase in the explained variation declines after each classification step. In addition, the effect of an additional classification step on the proportional reduction in error is decreasing. However, the reduction in the proportional reduction in error is not gradual. To identify the number of groups suitable for classifying the funds, we read the last column top-down and choose those classifications, after which the proportional reduction in error decreases considerably. Based on the proportional reduction in error, we identify four suitable classifications with two, six, 15, and 20 groups.

3.2 Characteristics of classification scheme

We re-run the series of 110 rolling classifications, this time setting the stepwise disaggregation in each month to equal the derived average disaggregation. To give the return-based classification scheme, i.e., the four suitable classifications, an economic interpretation, we use the self-declared investment objectives of the mutual funds and the four central moments of the group return indices.[3]

Table 3 presents the cross-tabulation of the declared fund investment objectives and the funds membership averages.

[Insert Table 3 here]

The variation of the investment objectives within the return-based classification groups, as reported in parentheses, has two potential reasons. The first results from variations in the sample: new funds are launched and existing funds are merged or liquidated. The sample variation is indicated in parentheses in the last column of the table. The second reason results from variations of the fund membership in the return-based classification: if a fund exhibits a change in the return profile and therefore will be more similar to another group of funds, it will be reassigned by the return-based classification algorithm. Therefore, the variation of the investment objectives within the groups has to be evaluated with caution.[4]

Table 4 presents the four central moments of the group return indices. The group indices are calculated from January 2000 to January 2014, as the simple return averages of their member funds.

[Insert Table 4 here]

Classification solution with two groups

The roughest classification solution separates the fund universe into two groups. We begin to interpret the two groups based on the return characteristics presented in Table 4, Panel A. As indicated by the standard deviation, kurtosis, and skewness of the group return time series, the first classification solution separates the funds with stable returns from funds with variable returns: Compared to group II, the return index of group I is characterized by a high standard deviation, negative skewness, and positive kurtosis. The interpretation that the classification solution with two groups separates the fund universe into stable and variable returns funds is

[3] A technical issue arises because the group indices $k = 1,2, \ldots, K$ in each month are independent of the previous month. Even if a group contains the same funds over time, the index may vary over time. To align the group indices over time, we identify the consecutive group indices by iteratively searching for the maximal intersection of member funds $|k_t \cap k_{t-1}^*| = \{i: i \in k_t \wedge i \in k_{t-1}^*\} \rightarrow max!$.

[4] The variation in the return-based classification can also result from changes in the investment objective. Yet, we are not able to capture this effect since Lipper does not report changes in the fund objectives.

supported by the investment objectives within the two groups. Table 3, Panel A shows that group I contains an average of 8,202 funds, with 65% equity, 19% mixed asset, and 8% bond funds, while group II contains an average of 2,901 funds, with 61% bond funds and 24% money market funds.

Classification solution with six groups

The second suitable classification solution separates the fund universe into six groups. Table 3, Panel B shows two groups primarily consisting of bond funds, two groups consisting of equity funds, one group consisting of bond and equity funds, and one group consisting of money market funds. In detail, group I contains an average of 1,234 funds, composed of 37% bond funds, 35% equity funds, 12% mixed asset funds, and 7% money market funds. The investment objectives of the equity funds are emerging markets, Japan, Europe, and global. Among the bond funds, and in contrast to groups V and VI, there is a high percentage investing in high yield or corporate bonds. Therefore, we label this group "Bond High Yield/Equity Global". Group II contains an average of 2,356 funds, with 61% equity and 24% mixed asset funds. We label this group "Equity Global" because there is no regional concentration of the funds and the majority of the equity funds invest globally. Group III contains an average of 4,612 funds, with 76% equity and 19% mixed asset funds. We label this group "Equity Europe" because most of the equity funds invest in Europe. Group IV contains an average of 733 funds, with 82% money market and 10% bond funds, whereby most bond funds are short-term; we label this group "Money Market". Group V contains an average of 1,408 funds, with 83% bond funds. The majority invest in euro-denominated fixed income securities or in fixed income securities from Eurozone issuers. Therefore, we label this group "Bond Eurozone". Group VI contains an average of 760 funds, composed of 71% bond, 7% money market, and 5% mixed asset funds. We label this group "Bond Global" because most of the bond funds invest globally and there is no predominant currency exposure. The characteristics of the group return indices presented in Table 4, Panel B reflect the fund's investment objectives comprising the six groups. There is a considerable co-movement in the returns of groups I, II, and III comprising equity funds and of groups V and VI comprising bond funds.

Classification solution with 15 groups

The third suitable classification solution separates the fund universe into 15 groups. Table 3, Panel C shows three groups primarily consisting of bond funds, 10 groups consisting of equity funds, and one group consisting of money market funds. One group has a significant percentage of protected funds.

Group I contains an average of 1,460 funds and is composed of 73% equity and 22% mixed asset funds. We label this group “Equity Eurozone” because most of the equity funds invest in Europe, particularly in the Eurozone. Group II contains an average of 1,179 funds, with 65% equity and 28% mixed asset funds. Most of the equity funds invest globally; therefore, we label this group “Equity Global”. Group III contains an average of 733 funds, composed of 82% money market and 10% bond funds, whereby most of the bond funds invest in short-term bonds. We label this group “Money Market”. Group IV contains an average of 737 funds, composed of 71% bond, 9% mixed asset and guaranteed funds, 7% money market, and 5% mixed asset funds. We label this group “Bond Global” because most of the bond funds invest globally and there is no predominant currency exposure. Group V contains an average of 335 funds, with 72% equity, 15% mixed asset, and 6% bond funds. Most of the equity funds invest in Japan or Europe. Therefore, we label this group “Equity Japan/Europe”. Group VI contains an average of 404 funds, with 78% equity, 11% mixed asset, and 7% bond funds. Most of the equity funds invest in Japan or Asia. We label this group “Equity Japan/Asia”. Group VII contains an average of 362 funds, composed of 69% equity, 16% bond, and 10% mixed asset funds. We label this group “Equity Emerging Markets” because the majority of the equity funds invest in the emerging markets. Group VIII contains an average of 816 funds, with 49% bond, 15% mixed asset, 14% equity, and 9% money market funds. In contrast to groups IV and XI, which also primarily comprise bond funds, there is significant percentage of high yield and corporate bonds in this group. Therefore, we label this group “Bond High Yield”. Group IX contains an average of 478 funds, composed of 71% equity, 18% mixed asset, and 5% bond funds. Because the majority of the equity funds invest in the emerging markets, we label this group “Equity Emerging Markets”. Group X contains an average of 1,076 funds, with 76% equity and 18% mixed asset funds. The majority of the equity funds invest in Europe. Therefore, we label this group “Equity Europe”. Group XI contains an average of 1,380 funds, with 83% bond funds. In contrast to groups IV and VIII, there is a significant percentage of bond funds focusing on the Eurozone. Therefore, we label this group “Bond Eurozone”. Group XII contains an average of 551 funds and is composed of 76% equity and 17% mixed asset funds. One-third of the equity funds invest in the UK, while most of the remaining equity funds invest globally or in Europe. Among the mixed asset funds, the majority have the reference currency GBP. We label this group “Equity UK/Global”. Group XIII contains an average of 598 funds, with 85% equity and 10% mixed asset funds. The majority of the equity funds invest either in the US or globally. Therefore, we label this group "Equity US/Global". Group XIV contains an average of 370 funds, composed of 31% mixed asset, 21% equity, 14% protected, 12% bond, 8% guaranteed funds, and 4% absolute return funds. Interestingly, this group mainly comprises funds that aim for a non-linear return profile. We label this group "Protected" because the majority of the funds belonging to that group aim to limit the downside risk. Group XV contains an average of 634 funds, with 72% equity and 22% mixed asset funds. Most of the equity funds invest in the UK, followed by funds investing globally and funds investing in Europe. Most of the mixed asset funds have the reference currency GBP. Therefore, we label this group "Equity UK/Global".

Classification with 20 groups

The finest classification solution separates the fund universe into 20 groups. Table 3, Panel D shows five groups primarily consisting of bond funds, 12 groups consisting of equity funds, and one group consisting of money market funds. One group consists equally of bond and equity funds, and one group comprises funds following dynamic strategies to limit the downside risk.

Group I contains an average of 1,469 funds, with 74% equity and 22% mixed asset funds. Both, the equity and the mixed asset funds, invest in Europe, with a majority focusing on the Eurozone. Therefore, we label this group "Equity Eurozone". Group II contains an average of 733 funds, composed of 82% money market and 10% bond funds, whereby the bond funds primarily invest in short term bonds. We label this group "Money Market". Group III contains an average of 689 funds, with 55% bond, 13% mixed asset, 12% equity, and 8% money market funds. Most of the bond funds either invest globally or in Euro denominated fixed-income securities. In contrast to group IX, the other Bond Global group, there is a significant percentage of funds focusing on corporate or high yield bonds. Therefore, we label this group "Bond Global High Yield". Group IV contains an average of 638 funds, with 69% equity, 22% mixed asset, and 5% bond funds. We label this group "Equity UK" because the majority of the equity and mixed-asset funds invest in the UK. The remaining funds invest in Europe or globally. Group V contains an average of 263 funds, composed of 85% equity and 10% bond funds. Almost all of the equity funds invest in Japan. Therefore, we label this group "Equity Japan". Group VI contains an average of 289 funds, composed of 44% equity, 35% bond, and 11% mixed asset funds. The major focus of the equity funds is US equity, followed by the technology and pharma sectors. More than half of the bond funds focus on high yield bonds. We label this group "Equity US Technology and Pharma/Bond High Yield". Group VII contains an average of 376 funds, with 63% equity, 15% mixed asset, and 12% bond funds. We label this group “Equity Emerging Markets” because the majority of the equity funds invest in the emerging markets. Group VIII contains an average of 315 funds, with 52% equity, 22% bond, and 17% mixed asset funds. The majority of the equity funds invest in the emerging markets. Either the bond funds focus on EUR or USD denominated fixed-income securities or invest globally. Therefore, we label this group “Equity Emerging Markets”. Group IX contains an average of 198 funds, composed of 56% bond, 18% equity, 9% mixed asset, and 6% money market funds. We label this group “Bond Global” because there is no regional or currency concentration. Group X contains an average of 979 funds, with 83% bond funds. Most of the bond funds invest in Euro-denominated funds and a significant percentage of them focus on the Eurozone. We label this group “Bond Eurozone”. Group XI contains an average of 270 funds, with 60% equity, 25% bond funds, 5% mixed asset, and 6% guaranteed or protected funds. The bond funds in USD or Euro-denominated securities or to invest globally. Because more than two-thirds of the equity funds invest in emerging markets, we label this group “Equity Emerging Markets”. Group XII contains an average of 289 funds, composed of 24% bond, 22% mixed asset, 16% protected, 13% equity, 10% guaranteed, 5% absolute return, and 4% money market funds. The mixed asset funds are primarily flexible or conservative. We label this group “Protected”. Group XIII contains an average of 341 funds, with 50% bond, 18% equity, 13% mixed asset 9% money market, and 6% protected or guaranteed funds. Most of the bond funds invest in euro-denominated fixed income securities, followed by bond funds investing globally or in the Eurozone. Because there is a considerable percentage of bond funds focusing on short term bonds, we label this group “Bond EUR Short Term”. Group XIV contains an average of 302 funds, with 61% equity, 24% bond, and 9% mixed asset funds. The majority of the equity funds invest in the emerging markets, followed by sector, US, and global equity funds. The sector funds invest in technology, pharma, or natural resources. We label this group “Equity Emerging Markets/US”. Group XV contains an average of 559 funds, with 58% equity, 20% mixed asset, and 14% bond funds. Half of the equity funds invest in the UK, and the others invest globally or in Europe. Half of the mixed asset funds have GBP reference currency, and the other half of the mixed asset funds have euro reference currency and invest globally. We label this group “Equity UK”. Group XVI contains an average of 395 funds, with 50% bond, 17% mixed asset, 14% equity, and 5% money market funds. Among the bond funds, there is no emphasis on a region. In addition, the bond funds’ currency exposures are EUR, followed by GBP and US. Therefore, we label this group “Bond”. Group XVII contains an average of 202 funds, with 44% equity, 25%

bond, 15% mixed asset, 6% guaranteed, and 5% protected funds. The investment objectives of the equity funds are Europe, followed by technology and natural resources, and emerging markets. Half of the equity emerging market funds focus on European emerging markets. We label this group “Equity Emerging Market Europe”. Group XIII contains an average of 1,077 funds, with 76% equity and 18% mixed asset funds. We label this group “Equity Europe” because most equity funds invest in Europe. Group IXX contains an average of 1,111 funds, composed of 65% equity and 28% mixed asset funds. The majority of the equity and the mixed asset funds invest globally. Therefore, we label this group “Equity Global”. Group XX contains an average of 609 funds, with 85% equity funds and 10% mixed asset funds. The major regional focus of the equity funds is the US. Therefore, we label this group “Equity US”.

3.3 Dynamics of fund memberships

To evaluate the dynamics of the four return-based classification solutions over time, we estimate the fund transition probabilities. Following the cohort approach, which is the industry standard in credit rating analysis (e.g. Jafry and Schuermann, 2004; Schuermann, 2008; Emery et al., 2008; Tennant, 2008; Chambers and Gurwitz, 2014), we estimate the monthly probability that a fund in group k migrates to group k^* as the number of funds assigned to group k in month $t-1$ and to group k^* in the subsequent month t, relative to all funds of funds in group k in month $t-1$[5]

$$p(t)_{k,k^*} \equiv \frac{|k_t^* \cap k_{t-1}|}{|k_{t-1}|}. \tag{3}$$

Table 5 presents the average monthly transition probabilities for the funds to stay in a given group (diagonal) or to migrate to a different group (off-diagonal).

[Insert Table 5 here]

Overall, the high values on the diagonal indicate that the return-based classification scheme is stable over time. Increasing the number of groups into which the funds are classified generally reduces the probability of remaining in the respective group. However, there are differences between the groups: while the fund memberships to some groups are instable over time, the

[5] In contrast to credit quality analysis, we do not include an absorbing group that contains funds which are withdrawn from the sample within a month.

memberships to others, particularly the group comprising money market funds, are highly stable. These differences in the transition probabilities grow with an increasing number of groups.

Classification solution with two groups

Table 5, Panel A shows the transition probabilities for the classification with two groups. The probability to remain is 99.7% for group I, “Variable Returns”, and 99.2% for group II, “Stable Returns.”

Classification solution with six groups

Table 5, Panel B shows the transition probabilities for the classification with six groups. The entries on the diagonal show that overall, the dynamics of the classification increase significantly, but there are differences across the groups: the lowest probability to remain is 89.9% for group I, “Bond High Yield/Equity Global,” while the highest is 97.6% for group IV, “Money Market”. In detail, the probabilities to remain are 89.9% for group I (“Bond High Yield/Equity Global”), 92.1% for group II (“Equity Global”), 95.3% for group III (“Equity Europe”), 97.6% for group IV (“Money Market”), 94.7% for group V (“Bond Eurozone”), and 90.8% for group VI (“Bond Global”). The off-diagonal elements indicate fund migrations among the first three groups, comprising equity funds and among the last two groups, comprising bond funds.

Classification solution with 15 groups

Table 5, Panel C shows the transition probabilities for the classification with 15 groups. The span of the diagonal entries widens from 71.7% for group XIV (“Protected”) to 97.6% for group III (“Money Market”). In detail, the probabilities to remain are 85.9% for group I (“Equity Eurozone”), 84.2% for group II (“Equity Global”), 97.6% for group III (“Money Market”), 90.6% for group IV (“Bond Global”), 75.8% for group V (“Equity Japan/Europe”), 78.3% for group VI (“Equity Japan/Asia”), 81.0% for group VII (“Equity Emerging Markets”), 76.7% for group VIII (“Bond High Yield”), 72.1% for group IX (“Equity Emerging Markets”), 77.6% for group X (“Equity Europe”), 94.5% for group XI (“Bond Eurozone”), 81.3% for group XII (“Equity UK/Global”), 85.3% for group XIII (“Equity US/Global”), 71.7% for group XIV (“Protected”), and 77.0% for group XV (“Equity UK/Global”). The highest pairwise migrations are among group I (“Equity Eurozone”) and group X (“Equity Europe”), among group IX (“Equity Emerging Markets”) and XIV

("Protected"), and from Group IX ("Equity Emerging Markets") to group II ("Equity Global"). The off-diagonal row sums quantify the monthly percentage of funds that leave the respective group, while the off-diagonal column sums quantify the monthly percentage of funds that migrate to the respective group. On average, there is a migration tendency of funds toward group I ("Equity Eurozone"), group XI ("Bond Eurozone"), group II ("Equity Global"), and group IV ("Bond Global").

Classification with 20 groups

Table 6, Panel D shows the transition probabilities for the classification with 20 groups. The span of the diagonal entries widens further from 60.8% for group XIII ("Equity Emerging Markets") to 97.6% group II ("Money Market"). In detail, the probabilities to remain are 87.6% for group I ("Equity Eurozone"), 97.6% for group II ("Money Market"), 79.1% for group III ("Bond Global High Yield"), 74.5% for group IV ("Equity UK"), 89.3% for group V ("Equity Japan"), 75.8% for group VI ("Equity US Technology and Pharma / Bond High Yield"), 69.8% for group VII ("Equity Emerging Markets"), 60.8% for group VIII ("Equity Emerging Markets"), 61.1% for group IX ("Bond Global"), 87.1% for group X ("Bond Eurozone"), 76.0% for group XI ("Equity Emerging Markets"), 73.2% for group XII ("Protected"), 69.5% for group XIII ("Bond EUR Short Term"), 76.7% for group XIV ("Equity Emerging Markets/US"), 81.2% for group XV ("Equity UK"), 71.0% for group XVI ("Bond"), 61.1% for group XVII ("Equity Emerging Market Europe"), 78.3% for group XVIII ("Equity Europe"), 83.5% for group IXX ("Equity Global"), and 88.8% for group XX ("Equity US"). The off-diagonal row and column sums show, on average, that there is a considerable migration of funds toward group X ("Bond Eurozone") and group IXX ("Equity Global") while, contrarily, there is a migration trend away from the "Equity Emerging Markets" groups (VII, VIII, and XVII) and from the "Bond Global" group (IX).

To analyze if the monthly fund migrations show moderate variation over time or exhibit a time dependent volatility, we estimate the monthly maximum migration rates of one group, i.e. $1 - \min(p(t)_{k,k})$. Figure 4 shows the maximal monthly probability of funds in one group to migrate to a different group.

[Insert Figure 4 here]

In the two-group solution, there are almost no migrations of funds among the groups, indicating that the separation between variable and stable return funds is highly stable over time. Only in September and October 2008 do the migration rates reach 6% and 10% respectively. The six-group solution shows a pattern reflecting the situation of the financial markets. During periods of market turbulence, such as during the financial crises in 2007-2008 and the sovereign debt crisis in the Euro-area that began in 2008 and peaked in 2012, the maximum migration rates increase significantly while the migration rates remain low in gentle market environments. In the 15- and 20-group solutions, the level of fund migrations among the groups is 76% and 88%, indicating that the fund memberships of these classification solutions are highly unstable. In almost every month, a whole group of funds migrates to other groups.

The previous analysis indicates that the classification solutions with two and six groups respectively are the most stable of the classification scheme with four layers. The findings resulting from the fund migration analysis, that fund migrations increase in a turbulent market environment and there are a few migration rate peaks close to one, raise the questions about the origin and goal of the fund migrations and the types of funds that are migrating. To investigate this issue, we show the monthly pairwise migrations for the two- and six-group solutions and provide further information about the investment objectives of the migrating funds. Figure 5 shows the monthly pairwise fund migrations for the two-group solution.

[Insert Figure 5 here]

In the two-group solution, funds primarily migrate from group II, “Stable Returns”, to group I, “Variable Returns”; the opposite is rare. Panel A shows that this rare event of funds migrating from group I to group II happens during the four last months in 2013. Half of the 793 funds that migrate are bond funds, and the other half are equity, mixed asset, or money market funds. Panel B shows the significant peak in September and October 2008. Within these two months, 552 bond and money market funds migrate to group I (“Variable Returns”). The majority of these bond funds invest in emerging markets, corporates, or euro-denominated short term bonds, and the majority of these money market funds invest in euro-denominated money market instruments. Figure 6 shows the monthly pairwise fund migrations for the six-group solution.

[Insert Figure 6 here]

Panels A, B, and C show that funds of group I (“Bond High Yield/Equity Global”), group II (“Equity Global”), and group III (“Equity Europe”) migrate among these three groups but not to the other groups. Similarly, Panels E and F show that funds generally migrate among groups V ("Bond Eurozone") and VI ("Bond Global") but not to the other groups. Exemptions are September and October 2008 for group VI ("Bond Eurozone"). In these two months, funds also migrate to groups I ("Bond High Yield/Equity Global") and IV ("Money Market"). Group IV ("Money Market") is different to the other groups in two ways: First, it is the only group that, outside of August 2008, loses funds but does not attract funds from other groups over time. Second, the average migration rate is low compared to the other five groups. Figure 6 also shows that the periods characterized by high fund migrations differ for groups I, II, and III, comprising equity and high yield funds, and for groups V and XI, comprising bond funds. For groups I, II, and III, the periods with high migration rates are February 2005, May and June 2006, October 2007 to May 2009, April 2010, and August to December 2013. In February 2005, all funds of group I, by this time almost exclusively equity funds investing in Europe or global, are replaced by equity funds investing in Japan or Asia Pacific. The separation of Japanese and Asia/Pacific equity funds indicate a distinct return pattern of the Japanese equity market. In May 2006, all Japan and Asia Pacific equity funds migrate from group I to group II and move back to group I in June 2006. A potential reason for this temporary relocation of the Japanese and Asia Pacific equity funds is the temporary market correction in May and June 2006 that was most pronounced in Japan and the euro area (OECD, 2006). The temporary migration of all member funds away from group I is offset by a relocation of equity funds investing globally, in the US, and in the technology or pharma sector to group I. In October 2007 – May 2009, during the financial crisis, the fund migrations between the first three groups significantly increase reflecting the equity markets' turmoil. The classification algorithm relocates 1,255 funds per month on average, primarily equity funds investing in Europe, globally, the UK, or in the technology, real estate, or financial sectors. In April 2010, 2,357 funds, the majority investing in equity Europe but some investing in the real estate or financial sector, migrate from group II ("Equity Global") to group III ("Equity Europe"). In August – December 2013 2,012 funds per month on average migrate between the first three groups; most of them are equity funds investing in emerging markets, globally, in the US, the UK, Asia Pacific, indicating changes in the return profiles of these markets.

For groups V and VI, the periods characterized by high fund migrations are August to October 2008, June to November 2009, and October 2011 to May 2012. During the financial crisis in August – October 2008, the classification algorithm relocates 563 funds per month on average between the last two groups. The relocated funds are bond funds, with a significant portion focusing on corporate bonds. In June – November 2009, on average, 370 funds per month migrate between the last two groups; the majority invest in euro-denominated bonds, in Eurozone bonds, or in European bonds. Retrospectively, these fund migrations reflect the beginning pressure in the bond market of the Eurozone. During the peak of the euro crisis, October 2011 – May 2012, the algorithm relocates 574 funds per month on average between the last two groups. The majority of the relocated funds invest in euro-denominated bonds, Eurozone bonds, or globally in bonds; there is a considerable number of relocated funds that invest in GBP or USD denominated bonds or aim for a guaranteed return, absolute return, or downside protection. In August 2008, 208 funds, mostly money market funds but also some real estate funds, migrate from group IV ("Money Market") to group VI ("Bond Global"), indicating that the risk in the returns of these funds drift away from the stable returns of their former peer group.

The few migration rate peaks close to one hundred percent in the six-group solution indicate that the investment objectives comprising the groups significantly change over time. Figure 7 provides detailed information about the size and the composition of the groups over time, whereby each color represents one of the 134 self-declared investment objectives, including unclassified.

[Insert Figure 7 here]

It is most obvious that the last three groups exhibit a high continuity over time, while the first three groups show a few structural breaks in terms of their sizes and their compositions. The development of group I ("Bond High Yield/Equity") shows five stages that we label based on the self-declared investment objectives of the funds: "Equity Europe" in December 2004 – January 2005, "Equity Japan" in February 2005 – September 2007, "Equity Global" in September 2007 – September 2008, "Bond Global High Yield" in October 2008 – April 2013, and "Bond Global High Yield/ Equity Emerging Markets" in May 2013 – January 2014. Group II ("Equity Global") also shows five stages, but compared to Group I ("Bond High Yield/Equity"), the composition changes at different dates: "Equity Emerging Markets/ Bond Global High Yield" in December 2004 – September 2007, "Equity Japan" in October 2007 – November 2008, "Equity Global" in December 2008 – March 2010, and "Equity Global" in April 2010 – January 2014, whereby the overall size and percentage of European equity funds significantly decrease. Group III ("Equity Europe") shows three stages: December 2004 – April 2009, May 2009 – April 2010, and May 2010 – January 2014. The investment objectives comprising group III do not change over time, but the size drops in the second stage. The last three groups do not exhibit stages, indicating that the funds within each group exhibit similar return time series.

3.4 Quality of classification scheme

Fund classification schemes should reflect differences in the funds' attributes, in our case, the fund returns. Funds with similar returns should belong to the same group, while funds with different returns should be in different groups.

To estimate the explanatory power of the return-based and Lipper classification schemes, we regress the monthly cross-section of fund returns $r_{i,t}$ against a set of group dummies $D_{i,k}$, $k = 1, \ldots, (K-1)$ which take the value of one if fund i belongs to group k and zero otherwise

$$r_{i,t} = a_t + \sum_{k=1}^{K-1} b_k D_{i,k} + e_{i,t}\,. \qquad (4)$$

Table 6 presents the average adjusted R^2 (standard deviation in parentheses) resulting from regressing the monthly cross-section of fund returns against the alternative classifications.[6]

[Insert Table 6 here]

Generally, increasing the number of groups into which the funds are categorized results in a higher explanatory power. Comparing the in-sample adjusted R^2 of the return-based and the Lipper classification schemes shows two issues that we want to point out. First, in line with Brown and Goetzmann (1997, 2003) and Gerlach and Maurer (2014), we find reasons that the return-based classification has greater explanatory power compared to the Lipper classification, based on declared investment objectives. To explain, on average, 25% of the monthly cross-sectional return variation, the Lipper classification requires thirteen groups (12 plus the unclassified group), while the hierarchical K-means classification needs less than six groups. Second, we find that the Lipper classification is highly successful in explaining the differences in funds. The finest classification comprising 286 categories explains 63% of the monthly, cross-sectional return variation on average. We also estimate the predictive power of the return-based classification, by regressing cross-section of fund returns against the classification of the preceding month. Since Lipper does not provide a history of its

[6] In contrast to the Lipper classification, the hierarchical K-means classification only includes funds with a return times-series of 60 months. To avoid potential biases resulting from different samples, we match the two samples at each month by restricting the Lipper sample to those funds that also enter the hierarchical K-means classification.

classification, we are only able to derive the out-of-sample results for the hierarchical K-means classification. The out-of-sample R^2 are not significantly lower than the in-sample R^2, indicating that the estimated classification based on historical returns has a high predictive power of the return differences in the subsequent month.

Investors use mutual fund classifications to evaluate the performance of an individual fund relative to its peers following the same investment style. These peer group comparisons are only adequate when the return profiles of the fund and the peer group are similar. To estimate the adequacy of the peer group of the return-based and Lipper classification schemes, we run a series of rolling time series regressions. In each month, we regress the historical 60-month return time series of each fund $r_{i,t}$ against its respective group return index $\mu_{k,t}$

$$r_{i,t} = a_i + b_{i,k}\mu_{k,t} + e_{i,t}\,. \tag{5}$$

Accounting for the differences in the ongoing costs captured by the intercept a_i and the fund specific exposure to style k, $b_{i,k}$, the resulting R^2 measures the similarity of the historical fund returns and peer group average.

Table 7 presents the resulting average adjusted R^2 (standard deviation in parentheses).

[Insert Table 7 here]

In general, even though the hierarchical K-means classification scheme sorts the funds in only a few groups, the higher adjusted R^2 indicate that the funds within these groups are more similar than the funds within groups by Lipper. Therefore, the return-based classification provides more adequate peer group benchmarks than the classification based on self-declared investment objectives.[7]

4 Conclusion

Return-based classification schemes better reflect the time-varying commonalities in mutual fund returns than traditional classification schemes based on self-declared investment objectives. We apply the hierarchical K-means algorithm to derive an endogenous, return-based classification scheme for an extensive sample of mutual funds registered for sale in Europe. We find a classification scheme that has four layers with two, six, 15, and 20 groups

[7] In the Lipper classifications, several groups consist of a single fund. Therefore, the regression results in a perfect fit. Table 7 includes the R^2 for the Lipper groups consisting of a single fund. Excluding the R^2 of the groups with a single fund does not change the results.

is appropriate to reflect the average return commonalities. In the first layer, mutual funds are divided into stable and variable returns; in the second layer, into the major asset classes (bond, equity, and money market), each asset class either with a European or a global focus. The last two layers refine the classification of asset classes and economic areas and form one group of funds aiming for market participation combined with downside protection. We estimate the dynamics of our classification scheme using the monthly probabilities of each fund to remain in its respective group, or to migrate to a different group. The two-group solution is highly stable with almost no fund migrations. In the six-group solution, the monthly probabilities of funds remaining in their groups decreases significantly. The six-group solution also shows the strength of our return-based classification scheme to adapt to changes in in the market environment by relocating funds between the groups. In the 15- and 20-group solutions, the fund migrations are high: In almost every month, at least one entire group is replaced by funds from other groups. We compare how well each approach explains the cross-sectional return variation and estimate the adequacy of their group return indices for performance evaluations. We find evidence that the return-based classification scheme requires fewer groups to explain the same percentage of the cross-sectional return variation, and it also provides more adequate benchmarks for peer group comparisons.

Our study shows that only a few groups are needed to reflect most of the differences in mutual fund returns. Most of the return differences can be explained by separating stable from variable returns. We find that six groups provide the best compromise between explanatory power and stability of the classification scheme. Even though return-based classifications seem to be superior to schemes based on self-declared investment objectives, they are purely statistical approaches that require an economic interpretation either through intensive style analysis techniques or by the more convenient way using the declared investment objectives. Therefore, both the return-based and the investment objective-based classification schemes complement one another.

5 References

Bailey, J. V. and Arnott, R. D. (1986): Cluster analysis and manager selection. *Financial Analysts Journal* 42(6), 20–28.

Brown, S. J. and Goetzmann, W. N. (1997): Mutual fund styles. *Journal of Financial Economics* 43(3), 373–399.

Brown, S. J. and Goetzmann, W. N. (2003): Hedge funds with style. *Journal of Portfolio Management* 29(2), 101–112.

Chambers, J. B. and Gurwitz, Z. R. (2013): Sovereign defaults and rating transition Data, 2013 Update. *Standard & Poor's Financial Services*.

diBartolomeo, D. and Witkowski, E. (1997): Mutual fund misclassification: Evidence based on style analysis. *Financial Analysts Journal* 53(5), 32–43.

Dor, A. B., Dynkin, L., and Gould, T. (2006): Style analysis and classification of hedge funds. *Journal of Alternative Investments* 9(2), 10–29.

Emery, K., Ou, S., and Tennant, J. (2008): Corporate default and recovery rates, 1920-2007. *Moody's Global Corporate Finance*.

Fung, W. and Hsieh, D. A. (1997): Empirical characteristics of dynamic trading strategies: The case of hedge funds. *The Review of Financial Studies* 10(2), 275–302.

Gerlach, P. and Maurer, R. (2014): Return-based classification of absolute return funds. *SSRN Working Paper Series*.

Gibson, R. and Gyger, S. (2007): The style consistency of hedge funds. *European Financial Management* 13(2), 287–308.

Gifford, H.: Hierarchical k-Means for unsupervised learning. http://www.andrew.cmu.edu/user/hgifford/projects/k_means.pdf.

Gruber, M. J. (2001): Identifying the risk structure of mutual fund returns. *European Financial Management* 7(2), 147–159.

Jafry, Y. and Schuermann, T. (2004): Measurement, estimation and comparison of credit migration matrices. *Recent Research on Credit Ratings* 28(11), 2603–2639.

Jain, A. K. (2010): Data clustering: 50 years beyond K-means. *Pattern recognition letters* 31(8), 651–666.

Kim, M., Shukla, R., and Tomas, M. (2000): Mutual fund objective misclassification. *Journal of Economics and Business* 52(4), 613–634.

Lamrous, S. and Taileb, M. (2006): Divisive hierarchical k-means. *Computational Intelligence for Modelling, Control and Automation, 2006 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce*, 18–118.

LeClair, R. T. (1974): Discriminant analysis and the classification of mutual funds. *Journal of Economics and Business* 26(3), 220–224.

Lee, C. and Chu, K.: CUDA-accelerated hierarchical k-means. http://cml18.csie.ntu.edu.tw/~cclee/CUDA-acceleratedHierarchicalK-means.pdf.

Lisi, F. and Otranto, E. (2008): Clustering mutual funds by return and risk levels. *Working Paper CRENoS*.

Marathe, A. and Shawky, H. A. (1999): Categorizing mutual funds using clusters. *Advances in Quantitative Analysis of Finance and Accounting* 7(1), 199–204.

Moreno, D., Marco, P., and Olmeda, I. (2006): Self-organizing maps could improve the classification of Spanish mutual funds. *European Journal of Operational Research* 174(2), 1039–1054.

Nister, D. and Stewenius, H. (2006): Scalable recognition with a vocabulary tree. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2, 2161–2168.

OECD (2006): Highlights of recent trends in financial markets. *Financial Market Trends 91*.

Schuermann, T. (2008): Credit migration matrices. Forthcoming in *Encyclopedia of Quantitative Risk Analysis and Assessment*.

Steinley, D. (2006): K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology* 59 (1), 1–34.

Steinley, D. and Brusco, M. J. (2007): Initializing k-means batch clustering: A critical evaluation of several techniques. *Journal of Classification* 24(1), 99–121.

Tennant, J. (2008): Corporate one-to-five-year rating transition rates. *Moody's Investors Service*, Report Number 108154.

Table 1
Fund investment objectives in the sample

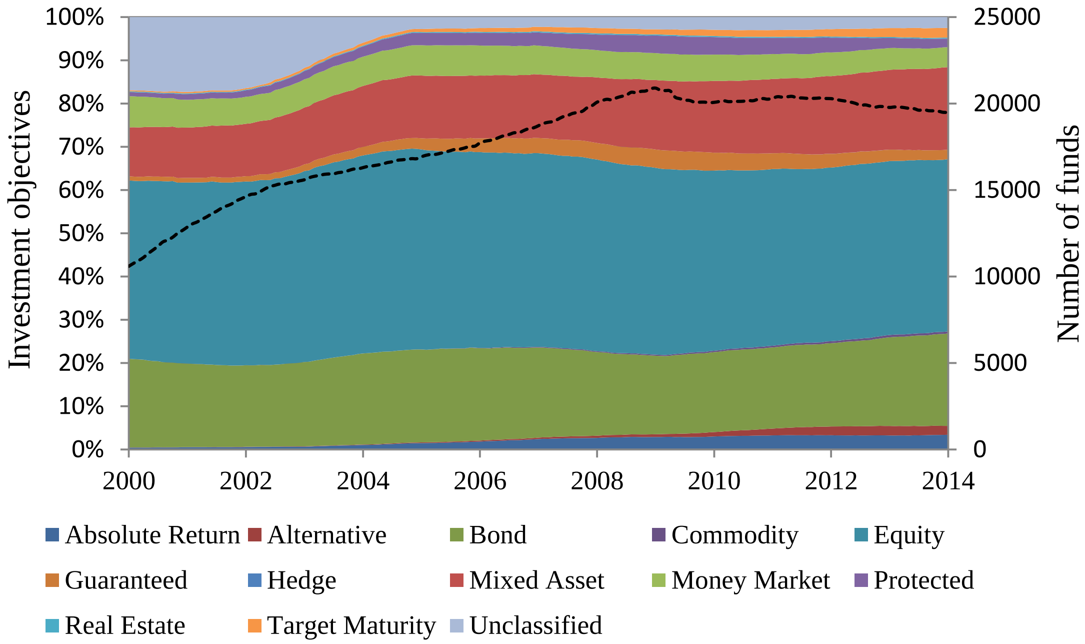
This table presents the investment objectives of all primary share classes of active, merged, and liquidated mutual funds registered for sale in Europe, as reported by Lipper at the end of January 2014.

	Funds
Absolute Return	1,255
Alternative	590
Bond	6,846
Commodity	141
Equity	13,692
Guaranteed	1,516
Hedge	4
Mixed Asset	5,959
Money Market	1,903
Protected	1,331
Real Estate	69
Target Maturity	638
Unclassified	4,129
Total	38,073

Source: Lipper, a Thomson Reuters Company.

Figure 1
Investment objectives of active funds

This figure shows the number and investment objectives of active mutual funds in Europe for each month, from January 2000 to January 2014. The investment objectives are as reported at the end of January 2014.



Source: Lipper, a Thomson Reuters Company.

Figure 2
Classification process in December 2004

This figure shows the first four classification steps in December 2004. The subsample of 8,764 mutual funds with a return series from January 2000 to December 2004 is classified using the hierarchical K-means clustering approach. The labels of the groups are given by the authors based on the investment objectives of the funds assigned to the groups.

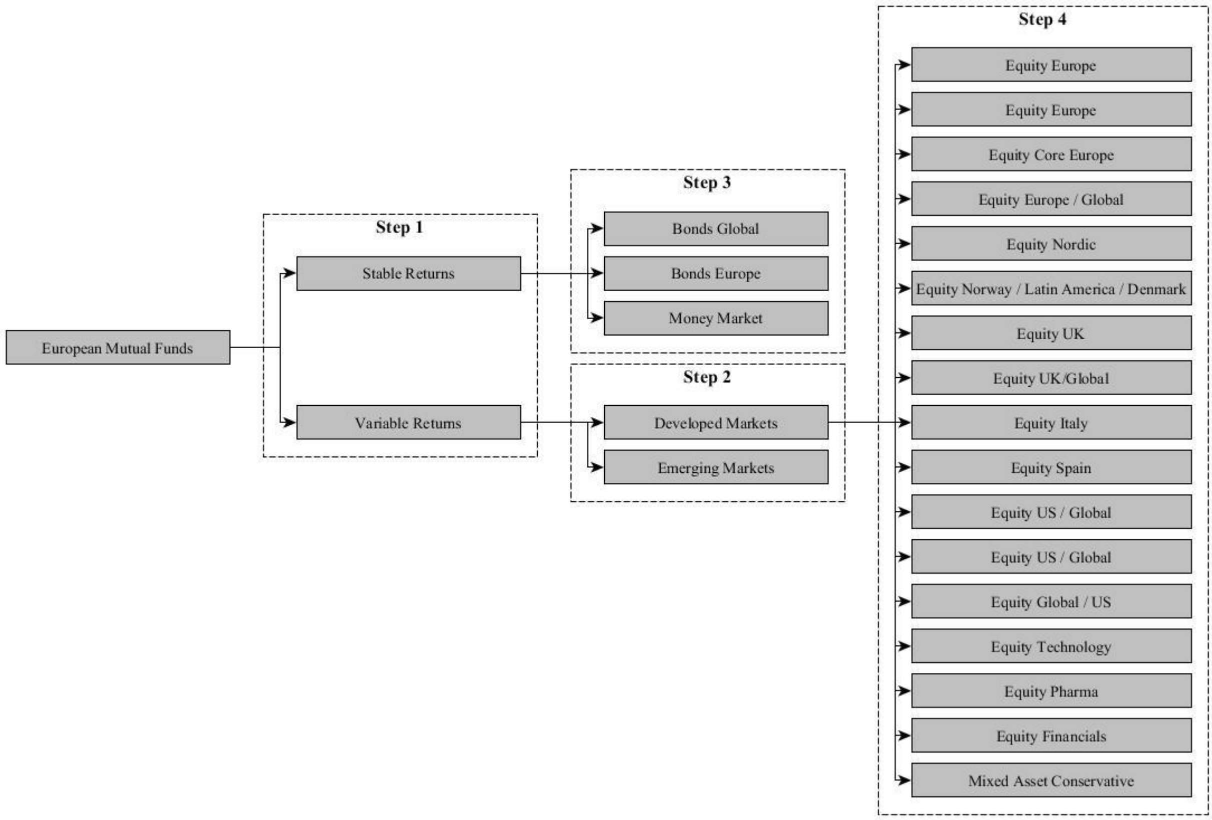


Figure 3
Classification process in January 2014

This figure shows the first five classification steps in January 2014. The subsample of 12,365 mutual funds with a return series from February 2009 to January 2014 is classified using the hierarchical K-means clustering approach. The labels of the groups are given by the authors based on the investment objectives of the funds assigned to the groups.

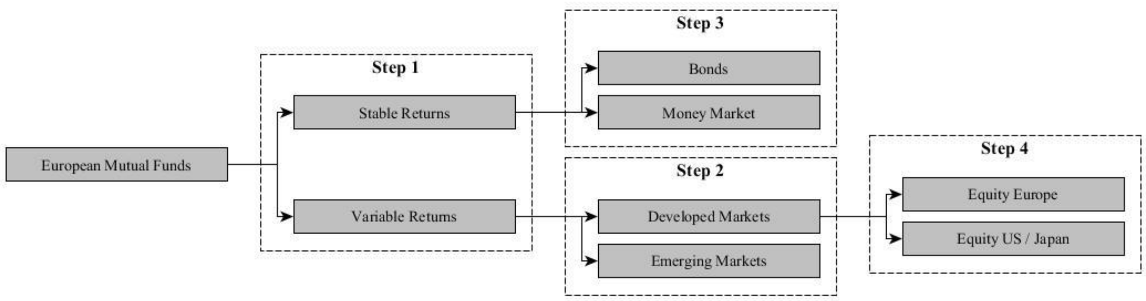


Table 2
Technical summary of rolling classifications

This table presents the technical coefficients resulting from a series of monthly hierarchical K-means procedures with overlapping windows. In each month, the classification procedure iteratively divides one group into multiple subgroups. The first column presents the average number of groups at each classification step (standard deviation in parentheses) and the second column presents the corresponding, average sum of squared errors (standard deviation in parentheses). The explained variation in column four measures the reduction in the sum of squared errors relative to the unclassified sample. The proportional reduction in error, presented in the last column, measures the reduction in the sum of squared errors relative to the preceding classification step.

	Number of groups	Sum of squared errors	Explained variation	Proportional reduction in error
	1	428,749		
	(0.00)	(16,970)		
Step 1:	2.45	263,188	0.39	0.39
	(0.50)	(31,181)		
Step 2:	3.59	245,107	0.43	0.07
	(0.60)	(29,685)		
Step 3:	5.67	220,649	0.49	0.10
	(2.88)	(34,370)		
Step 4:	9.72	203,151	0.53	0.08
	(4.69)	(34,679)		
Step 5:	14.68	186,739	0.56	0.08
	(5.34)	(33,148)		
Step 6:	19.76	175,576	0.59	0.06
	(7.07)	(32,122)		
Step 7:	22.71	168,854	0.61	0.04
	(7.22)	(30,067)		
Step 8:	26.67	162,680	0.62	0.04
	(7.74)	(29,521)		
Step 9:	30.76	157,488	0.63	0.03
	(9.40)	(29,135)		

Table 3
Cross-tabulation of the hierarchical K-means classification and the funds' investment objectives

This table presents the average investment objectives (standard deviation in parentheses) within each group of the estimated hierarchical K-means classification. The fund objectives are as reported by Lipper at the end of January 2014.

Panel A: two-group solution

	Hierarchical K-means group		
	I	II	Total
Absolute Return	109	37	146
	(58.94)	(16.23)	(74.06)
Alternative	24	9	33
	(17.83)	(10.53)	(28.05)
Bond	637	1,781	2,418
	(286.46)	(223.79)	(96.74)
Commodity	10	1	11
	(12.11)	(0.60)	(12.07)
Equity	5,359	15	5,375
	(332.30)	(16.92)	(338.52)
Guaranteed	61	82	143
	(23.82)	(38.61)	(42.61)
Hedge	0	0	0
	(0.00)	(0.00)	(0.00)
Mixed Asset	1,596	69	1,665
	(206.49)	(30.81)	(234.37)
Money Market	105	699	804
	(59.46)	(55.38)	(46.62)
Protected	124	73	197
	(30.65)	(26.72)	(51.90)
Real Estate	7	22	30
	(4.42)	(3.30)	(6.84)
Target Maturity	62	32	94
	(25.22)	(14.32)	(36.38)
Unclassified	106	81	188
	(42.15)	(25.49)	(66.13)
Total	8,202	2,901	11,103
	(924.59)	(219.73)	(845.59)

Panel B: six-group solution

	Hierarchical K-means group						
	I	II	III	IV	V	VI	Total
Absolute Return	35	38	36	7	16	14	146
	(31.77)	(18.29)	(16.80)	(3.51)	(6.69)	(9.36)	(74.06)
Alternative	11	7	6	4	2	4	33
	(9.77)	(3.06)	(6.67)	(4.12)	(2.94)	(5.35)	(28.05)
Bond	458	106	74	76	1,167	538	2,418
	(349.00)	(109.07)	(54.85)	(23.16)	(360.52)	(196.55)	(96.74)
Commodity	9	1	0	0	0	0	11
	(11.39)	(0.92)	(1.38)	(0.39)	(0.36)	(0.50)	(12.07)
Equity	435	1,431	3,494	2	3	11	5,375
	(535.23)	(644.97)	(587.79)	(6.63)	(5.13)	(11.20)	(338.52)
Guaranteed	8	44	10	8	33	41	143
	(9.80)	(25.53)	(8.09)	(3.27)	(9.64)	(44.01)	(42.61)
Hedge	0	0	0	0	0	0	0
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mixed Asset	152	562	882	4	29	36	1,665
	(168.42)	(265.25)	(162.16)	(3.43)	(20.24)	(15.15)	(234.37)
Money Market	82	17	7	604	42	52	804
	(64.35)	(16.91)	(5.37)	(75.42)	(38.06)	(41.52)	(46.62)
Protected	15	76	34	1	42	30	197
	(16.95)	(46.43)	(22.21)	(1.66)	(22.77)	(21.55)	(51.90)
Real Estate	2	4	1	14	1	7	30
	(1.99)	(2.54)	(1.23)	(3.98)	(1.56)	(4.79)	(6.84)
Target Maturity	4	29	28	0	23	9	94
	(4.86)	(26.39)	(10.63)	(0.10)	(10.31)	(7.72)	(36.38)
Unclassified	24	42	41	12	51	18	188
	(19.14)	(24.79)	(31.37)	(7.05)	(22.13)	(14.99)	(66.13)
Total	1,234	2,356	4,612	733	1,408	760	11,103
	(753.78)	(933.60)	(743.28)	(86.40)	(347.99)	(268.29)	(845.59)

Panel C: 15-group solution

	Hierarchical K-means group									
	I	II	III	IV	V	VI	VII	VIII	IX	X
Absolute Return	8	10	7	13	4	4	5	28	11	9
	(4.81)	(11.99)	(3.51)	(8.73)	(6.73)	(7.03)	(5.99)	(25.79)	(12.75)	(7.80)
Alternative	1	1	4	4	1	1	1	9	1	2
	(2.21)	(2.50)	(4.12)	(5.06)	(2.37)	(1.37)	(1.64)	(8.14)	(2.60)	(2.96)
Bond	13	9	76	525	20	27	57	402	22	27
	(13.69)	(11.91)	(23.16)	(197.50)	(43.52)	(59.27)	(59.75)	(333.45)	(35.48)	(22.50)
Commodity	1	0	0	0	0	0	0	6	0	0
	(3.96)	(0.78)	(0.39)	(0.50)	(0.28)	(2.31)	(0.43)	(9.56)	(0.41)	(0.10)
Equity	1,072	764	2	10	241	314	248	115	337	814
	(310.76)	(320.27)	(6.63)	(10.73)	(141.53)	(147.11)	(251.48)	(120.46)	(239.76)	(352.90)
Guaranteed	5	3	8	38	3	1	3	10	4	2
	(4.68)	(6.93)	(3.27)	(41.40)	(6.50)	(1.81)	(4.47)	(9.71)	(13.17)	(3.07)
Hedge	0	0	0	0	0	0	0	0	0	0
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mixed Asset	327	333	4	35	49	45	36	122	85	191
	(102.34)	(158.55)	(3.43)	(15.01)	(72.40)	(73.59)	(58.10)	(76.07)	(97.04)	(118.35)
Money Market	2	2	604	50	5	3	3	74	3	2
	(1.55)	(4.11)	(75.42)	(39.57)	(12.76)	(3.75)	(6.86)	(66.27)	(6.89)	(1.52)
Protected	14	7	1	29	7	4	4	17	5	8
	(11.67)	(5.26)	(1.66)	(20.51)	(13.00)	(5.01)	(6.96)	(17.08)	(9.90)	(11.78)
Real Estate	0	1	14	6	1	0	0	2	1	1
	(0.65)	(2.42)	(3.98)	(4.59)	(0.99)	(0.72)	(0.45)	(2.04)	(1.41)	(1.01)
Target Maturity	6	30	0	9	0	1	2	3	2	8
	(4.25)	(13.43)	(0.10)	(7.70)	(0.73)	(4.31)	(4.55)	(3.07)	(3.45)	(7.80)
Unclassified	11	19	12	17	5	4	3	26	5	11
	(8.26)	(16.08)	(7.05)	(14.64)	(11.38)	(7.04)	(4.09)	(20.98)	(8.86)	(13.49)
Total	1,460	1,179	733	737	335	404	362	816	478	1,076
	(404.15)	(500.11)	(86.40)	(267.56)	(249.61)	(159.98)	(247.33)	(437.01)	(238.20)	(488.61)

Panel C: 15-group solution (continued)

	Hierarchical K-means group					
	XI	XII	XIII	XIV	XV	Total
Absolute Return	16	4	4	16	6	146
	(6.82)	(5.88)	(4.68)	(12.99)	(5.52)	(74.06)
Alternative	2	2	1	3	1	33
	(2.94)	(2.62)	(1.59)	(2.60)	(1.73)	(28.05)
Bond	1,142	18	11	46	23	2,418
	(355.28)	(33.77)	(17.11)	(87.27)	(44.03)	(96.74)
Commodity	0	1	1	0	0	11
	(0.34)	(4.70)	(3.29)	(0.33)	(2.77)	(12.07)
Equity	2	419	511	78	448	5,375
	(5.10)	(184.24)	(161.87)	(100.32)	(317.48)	(338.52)
Guaranteed	32	1	1	29	2	143
	(10.34)	(3.86)	(1.03)	(30.73)	(2.54)	(42.61)
Hedge	0	0	0	0	0	0
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mixed Asset	28	96	62	115	136	1,665
	(20.55)	(69.68)	(41.00)	(101.23)	(96.51)	(234.37)
Money Market	42	2	1	11	1	804
	(38.60)	(4.45)	(0.92)	(24.51)	(1.90)	(46.62)
Protected	41	2	2	52	4	197
	(22.86)	(6.13)	(2.02)	(48.17)	(6.54)	(51.90)
Real Estate	1	1	0	2	0	30
	(1.53)	(2.37)	(0.31)	(2.45)	(0.13)	(6.84)
Target Maturity	23	1	3	3	2	94
	(10.31)	(3.67)	(2.81)	(3.05)	(3.53)	(36.38)
Unclassified	51	3	3	15	2	188
	(21.97)	(4.22)	(5.26)	(20.79)	(3.44)	(66.13)
Total	1,380	551	598	370	624	11,103
	(347.02)	(249.09)	(200.11)	(144.57)	(413.27)	(845.59)

Panel D: 20-group solution

	Hierarchical K-means group									
	I	II	III	IV	V	VI	VII	VIII	IX	X
Absolute Return	8	7	24	6	1	4	6	8	3	11
	(3.46)	(3.51)	(26.57)	(4.16)	(3.25)	(7.79)	(6.13)	(11.92)	(3.99)	(7.39)
Alternative	1	4	7	1	0	1	1	1	1	1
	(2.03)	(4.12)	(8.76)	(1.31)	(0.83)	(2.28)	(1.55)	(2.49)	(1.48)	(1.82)
Bond	13	76	381	31	27	100	45	70	111	815
	(13.73)	(23.16)	(285.56)	(50.26)	(61.92)	(103.05)	(54.76)	(89.23)	(61.12)	(325.16)
Commodity	0	0	6	0	0	0	0	0	0	0
	(0.00)	(0.39)	(9.19)	(0.32)	(2.51)	(0.27)	(0.41)	(0.30)	(0.32)	(0.23)
Equity	1,084	2	85	440	223	127	238	165	36	1
	(282.97)	(6.63)	(128.70)	(264.09)	(112.71)	(184.66)	(236.63)	(188.94)	(116.04)	(1.84)
Guaranteed	5	8	6	4	1	3	5	3	5	18
	(4.69)	(3.27)	(5.59)	(5.15)	(1.27)	(4.92)	(8.78)	(4.99)	(5.98)	(15.39)
Hedge	0	0	0	0	0	0	0	0	0	0
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mixed Asset	325	4	90	139	5	33	57	53	17	18
	(103.39)	(3.43)	(90.09)	(82.46)	(9.80)	(61.61)	(78.51)	(87.21)	(45.06)	(20.45)
Money Market	2	604	58	3	3	10	4	5	11	26
	(1.58)	(75.42)	(58.07)	(9.18)	(4.09)	(27.85)	(11.71)	(10.94)	(23.70)	(38.19)
Protected	14	1	9	7	2	3	9	4	7	27
	(11.67)	(1.66)	(7.75)	(9.79)	(2.66)	(7.51)	(13.43)	(10.00)	(9.37)	(24.62)
Real Estate	0	14	3	0	0	1	0	0	1	0
	(0.66)	(3.98)	(3.10)	(0.64)	(0.80)	(0.99)	(1.44)	(0.67)	(1.39)	(0.67)
Target Maturity	6	0	4	2	0	1	1	2	1	19
	(4.21)	(0.10)	(5.00)	(2.47)	(0.59)	(3.93)	(2.02)	(5.16)	(2.28)	(10.34)
Unclassified	11	12	17	4	1	5	9	5	6	42
	(8.22)	(7.05)	(15.83)	(6.95)	(2.12)	(7.84)	(17.82)	(8.73)	(11.13)	(19.06)
Total	1,469	733	689	638	263	289	376	315	198	979
	(382.67)	(86.40)	(447.58)	(308.95)	(105.98)	(212.98)	(226.20)	(176.53)	(155.03)	(383.15)

Panel D: 20-group solution (continued)

	Hierarchical K-means group										
	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	IVX	XX	Total
Absolute Return	3	16	5	3	8	8	2	9	9	4	146
	(4.14)	(13.53)	(6.67)	(4.82)	(9.30)	(6.11)	(2.91)	(7.86)	(10.44)	(4.63)	(74.06)
Alternative	0	3	2	1	1	3	1	2	1	1	33
	(0.81)	(2.68)	(2.74)	(1.82)	(2.04)	(4.28)	(1.88)	(2.96)	(2.11)	(1.65)	(28.05)
Bond	66	69	171	72	79	198	51	25	8	11	2,418
	(64.12)	(115.31)	(158.58)	(108.67)	(160.30)	(241.23)	(58.05)	(18.85)	(11.13)	(19.75)	(96.74)
Commodity	0	0	0	1	0	0	1	0	0	1	11
	(2.33)	(0.35)	(0.26)	(5.24)	(0.21)	(0.36)	(4.80)	(0.00)	(0.71)	(3.29)	(12.07)
Equity	162	39	61	185	325	56	88	818	719	520	5,375
	(261.72)	(75.95)	(141.34)	(185.01)	(230.34)	(88.31)	(153.78)	(345.24)	(281.57)	(142.74)	(338.52)
Guaranteed	11	29	9	2	13	7	11	2	3	1	143
	(27.27)	(31.06)	(10.05)	(4.44)	(28.25)	(8.29)	(20.97)	(2.93)	(5.86)	(1.03)	(42.61)
Hedge	0	0	0	0	0	0	0	0	0	0	0
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mixed Asset	14	63	44	26	110	67	30	191	315	63	1,665
	(29.36)	(60.27)	(73.60)	(28.26)	(75.22)	(106.18)	(53.86)	(118.15)	(143.76)	(40.75)	(234.37)
Money Market	3	12	29	5	2	19	4	2	2	1	804
	(10.43)	(23.29)	(40.30)	(15.60)	(4.49)	(22.03)	(9.14)	(1.43)	(3.88)	(0.92)	(46.62)
Protected	5	48	10	4	8	13	10	8	6	2	197
	(10.18)	(51.18)	(16.33)	(8.33)	(12.01)	(12.64)	(13.89)	(11.78)	(4.91)	(2.06)	(51.90)
Real Estate	0	2	1	1	1	3	0	1	1	0	30
	(0.52)	(2.69)	(1.55)	(1.56)	(3.25)	(4.54)	(1.03)	(1.00)	(2.25)	(0.30)	(6.84)
Target Maturity	1	4	2	1	3	6	1	8	30	3	94
	(4.16)	(3.34)	(3.33)	(3.00)	(3.87)	(7.30)	(2.32)	(7.80)	(13.38)	(2.73)	(36.38)
Unclassified	2	7	7	2	9	14	2	11	18	3	188
	(4.42)	(6.20)	(9.83)	(3.21)	(12.53)	(21.24)	(6.10)	(13.48)	(14.69)	(5.35)	(66.13)
Total	270	289	341	302	559	395	202	1,077	1,111	609	11,103
	(254.26)	(146.55)	(253.56)	(171.69)	(230.36)	(268.62)	(188.14)	(486.55)	(441.25)	(181.07)	(845.59)

Table 4
Return characteristics of hierarchical K-means groups

This table presents the four central moments of returns for the hierarchical K-means group indices for the period of January 2000 to January 2014. The return time series of each group is calculated as the cross-sectional average of the funds belonging to the respective group.

Panel A: two-group solution

	Hierarchical K-means group	
	I	II
Mean (%)	0.23	0.29
Stdev. (%)	3.60	0.55
Skewness	-0.78	-0.01
Kurtosis	4.15	3.34

Panel B: six-group solution

	Hierarchical K-means group					
	I	II	III	IV	V	VI
Mean (%)	0.13	0.25	0.18	0.20	0.32	0.32
Stdev. (%)	3.62	3.21	4.13	0.36	0.72	0.86
Skewness	-0.37	-0.73	-0.74	-0.37	-0.14	0.33
Kurtosis	5.26	3.73	3.94	6.82	3.09	2.95

Panel C: 15-group solution

	Hierarchical K-means group									
	I	II	III	IV	V	VI	VII	VIII	IX	X
Mean (%)	0.09	0.01	0.20	0.33	0.16	0.23	0.77	0.16	0.35	0.01
Stdev. (%)	4.64	3.59	0.36	0.86	4.06	4.35	4.61	2.54	3.50	5.02
Skewness	-0.50	-0.78	-0.37	0.31	0.05	-0.29	-0.14	-1.57	-0.42	-0.67
Kurtosis	4.92	3.59	6.82	2.92	3.84	4.10	4.61	10.05	5.20	4.98

	Hierarchical K-means group				
	XI	XII	XIII	XIV	XV
Mean (%)	0.32	0.07	0.13	0.35	0.38
Stdev. (%)	0.72	4.40	4.68	2.56	4.33
Skewness	-0.14	-0.75	-1.19	-0.12	-0.81
Kurtosis	3.09	6.84	6.63	5.76	4.18

Panel D: 20-group solution

	Hierarchical K-means group									
	I	II	III	IV	V	VI	VII	VIII	IX	X
Mean (%)	0.22	0.20	0.45	0.46	-0.01	0.16	0.68	0.01	0.42	0.32
Stdev. (%)	4.56	0.36	3.49	3.64	4.03	2.22	3.60	3.00	2.71	0.76
Skewness	-0.69	-0.37	-0.36	-0.85	0.07	-0.34	-0.71	-0.82	-0.42	-0.04
Kurtosis	4.72	6.82	6.42	4.60	4.08	7.46	5.90	7.83	6.12	3.33

	Hierarchical K-means group									
	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
Mean (%)	0.69	0.65	0.45	0.43	0.30	0.17	0.22	-0.02	0.05	0.20
Stdev. (%)	3.41	2.02	2.42	4.08	2.81	3.11	4.12	5.03	3.75	5.19
Skewness	-0.14	0.58	-0.35	-1.02	-1.04	-0.43	-0.35	-0.66	-0.69	-0.64
Kurtosis	8.71	6.31	5.98	9.14	6.80	5.14	5.17	4.93	3.91	5.73

Table 5
Average monthly transition probabilities

This table presents the average monthly transition probabilities (%) of the hierarchical K-means classification solutions.

Panel A: two-group solution

from\to	I	II
Group I	**99.7**	0.3
Group II	0.8	**99.2**

Panel B: six-group solution

from\to	I	II	III	IV	V	VI
Group I	**89.9**	3.8	5.4	0.1	0.2	0.5
Group II	3.4	**92.1**	4.0	0.1	0.1	0.4
Group III	2.1	2.5	**95.3**	0.0	0.0	0.0
Group IV	0.2	0.2	0.0	**97.6**	0.8	1.2
Group V	0.3	0.1	0.0	0.3	**94.7**	4.5
Group VI	1.5	0.8	0.0	1.2	5.7	**90.8**

Panel C: 15-group solution

from\to	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV
Group I	**85.9**	1.8	0.0	0.0	0.1	0.3	0.2	0.4	0.1	6.5	0.0	0.4	1.8	0.3	2.1
Group II	1.4	**84.2**	0.0	0.0	0.3	2.2	0.8	0.4	2.8	2.3	0.0	3.0	0.7	1.1	0.7
Group III	0.0	0.0	**97.6**	1.0	0.1	0.0	0.0	0.2	0.1	0.0	0.8	0.0	0.0	0.3	0.0
Group IV	0.0	0.0	0.8	**90.6**	0.3	0.3	0.3	1.5	0.3	0.0	5.0	0.1	0.0	0.5	0.1
Group V	0.3	2.1	1.3	1.3	**75.8**	0.5	1.8	3.5	4.7	2.7	0.1	0.9	0.5	3.0	1.5
Group VI	0.5	4.6	0.1	1.6	0.5	**78.3**	0.8	0.8	3.7	0.7	1.0	3.5	1.3	1.4	1.2
Group VII	0.8	1.9	0.0	1.1	2.6	0.4	**81.0**	1.8	1.1	2.5	0.9	1.5	1.1	1.2	1.9
Group VIII	2.0	1.6	0.9	1.6	3.2	1.0	0.7	**76.7**	2.7	1.1	0.7	1.3	0.6	3.8	2.2
Group IX	0.3	6.2	0.0	1.0	2.6	2.1	1.3	2.3	**72.1**	2.1	0.3	2.4	0.6	5.6	1.1
Group X	7.9	3.9	0.0	0.0	1.0	0.7	1.3	0.8	0.8	**77.6**	0.9	1.1	2.7	0.5	0.8
Group XI	0.0	0.0	0.3	4.3	0.0	0.1	0.1	0.2	0.1	0.0	**94.5**	0.0	0.0	0.1	0.2
Group XII	1.1	5.0	0.1	1.0	0.5	2.7	0.4	0.6	1.1	1.7	1.0	**81.3**	1.5	0.6	1.4
Group XIII	3.3	2.0	0.0	0.0	0.6	0.3	0.4	0.6	1.0	1.4	1.0	1.9	**85.3**	0.1	2.1
Group XIV	0.5	3.6	0.4	2.0	3.1	1.8	0.6	4.2	6.2	1.1	1.6	1.7	0.1	**71.7**	1.2
Group XV	3.6	1.8	0.0	0.3	1.5	2.5	1.2	1.9	3.2	1.1	1.7	0.3	1.4	2.4	**77.0**

Panel D: 20-group solution

from\to	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX
Group I	**87.6**	0.0	0.4	0.3	0.0	0.0	0.3	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.4	0.1	1.1	6.5	1.0	1.8
Group II	0.0	**97.6**	0.3	0.1	0.0	0.1	0.0	0.1	0.1	0.4	0.1	0.3	0.2	0.1	0.1	0.4	0.1	0.0	0.0	0.0
Group III	0.9	0.4	**79.1**	0.3	0.5	0.6	1.0	1.5	0.7	2.2	0.6	1.8	1.5	2.8	0.3	2.4	0.4	1.0	1.9	0.3
Group IV	1.2	1.3	0.4	**74.5**	0.7	1.4	1.1	2.6	0.4	2.7	1.1	1.1	3.8	1.1	0.6	3.2	1.1	0.6	0.4	0.8
Group V	0.0	0.1	0.3	1.3	**89.3**	0.4	1.2	0.6	0.1	0.9	0.3	0.2	0.2	0.6	0.7	0.2	0.1	0.5	2.2	0.7
Group VI	0.1	0.1	2.1	2.0	0.4	**75.8**	1.3	0.9	1.2	2.6	0.0	0.8	1.6	1.5	0.6	1.5	1.0	1.0	5.4	0.1
Group VII	1.3	0.1	1.8	1.4	0.5	1.1	**69.8**	0.6	0.2	4.9	1.1	0.7	1.2	2.4	1.3	2.9	2.2	3.4	1.8	1.3
Group VIII	0.2	0.1	2.7	1.1	0.2	2.9	0.2	**60.8**	2.1	9.2	2.7	2.8	1.3	2.8	0.1	4.6	1.0	1.1	3.8	0.4
Group IX	0.0	0.8	2.9	0.7	0.2	2.1	0.2	3.2	**61.1**	12.0	1.7	1.0	3.5	2.0	1.7	4.4	1.3	0.3	0.1	0.8
Group X	0.0	0.2	0.5	0.3	0.1	0.4	0.6	1.8	1.8	**87.1**	1.1	0.7	2.0	0.2	0.5	1.5	1.2	0.0	0.0	0.0
Group XI	0.1	0.2	2.2	0.7	0.0	0.1	1.5	0.5	1.3	7.7	**76.0**	0.9	0.5	1.4	2.0	1.4	1.5	0.9	0.8	0.2
Group XII	0.4	0.5	2.2	0.3	0.4	0.4	1.3	2.2	2.5	3.0	4.3	**73.2**	0.4	2.5	2.4	0.7	0.4	0.2	2.6	0.0
Group XIII	0.3	1.4	3.5	1.0	0.0	0.6	0.3	0.4	0.7	10.7	0.8	1.1	**69.5**	0.3	2.3	2.4	1.0	0.9	2.6	0.4
Group XIV	0.2	0.1	2.5	0.6	1.2	3.4	1.2	1.5	0.4	2.9	0.4	1.5	0.2	**76.7**	0.8	0.8	0.2	0.3	4.3	0.8
Group XV	0.8	0.1	0.4	1.6	1.0	1.4	0.5	0.0	0.8	2.5	0.7	1.4	0.4	1.3	**81.2**	0.4	0.3	1.8	2.5	0.5
Group XVI	0.1	0.7	0.9	3.1	0.1	1.1	1.4	3.0	1.4	5.4	0.4	0.4	2.1	1.0	1.3	**71.0**	1.5	1.0	3.5	0.5
Group XVII	3.8	0.5	2.0	1.3	0.1	1.4	2.0	1.1	2.1	9.8	1.0	0.1	2.3	0.1	2.0	4.7	**61.1**	0.9	2.8	1.0
Group XVIII	8.0	0.0	0.4	0.6	0.1	0.5	1.5	0.2	0.4	0.0	0.9	0.1	0.3	0.1	1.4	0.4	0.3	**78.3**	3.9	2.5
Group XIX	1.2	0.0	1.5	0.1	0.1	2.7	0.6	1.4	0.0	0.0	0.3	0.6	0.7	1.4	1.6	0.7	0.9	2.3	**83.5**	0.4
Group XX	3.7	0.0	0.2	0.7	0.4	0.1	0.5	0.5	0.1	0.0	0.3	0.1	0.3	0.6	0.5	0.1	0.6	1.4	1.2	**88.8**

Note: The sums of the rows do not equal one due to rounding.

Figure 4
Monthly maximum fund migrations

This figure shows the monthly maximum percentage of funds that migrate to a different group.

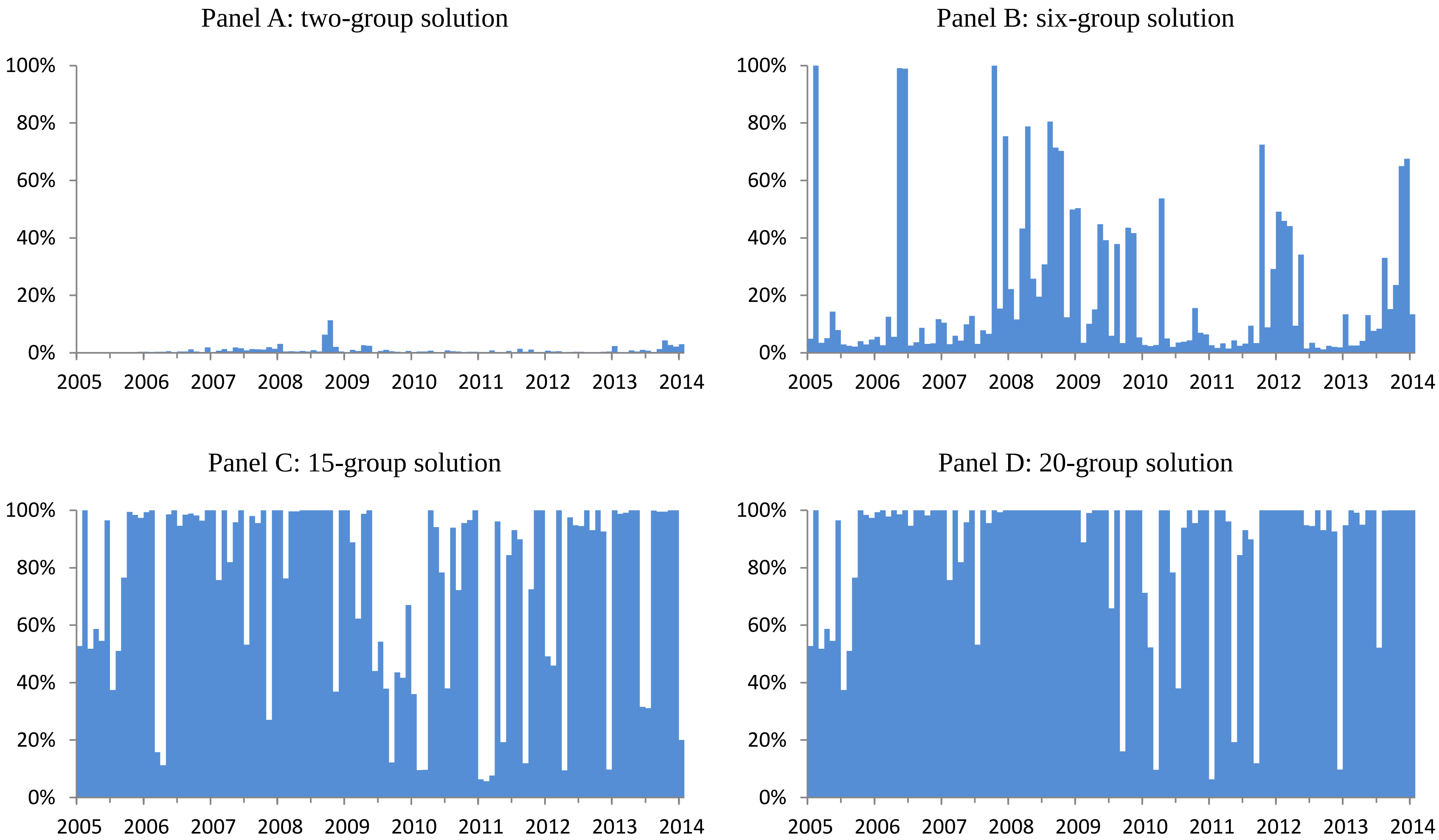
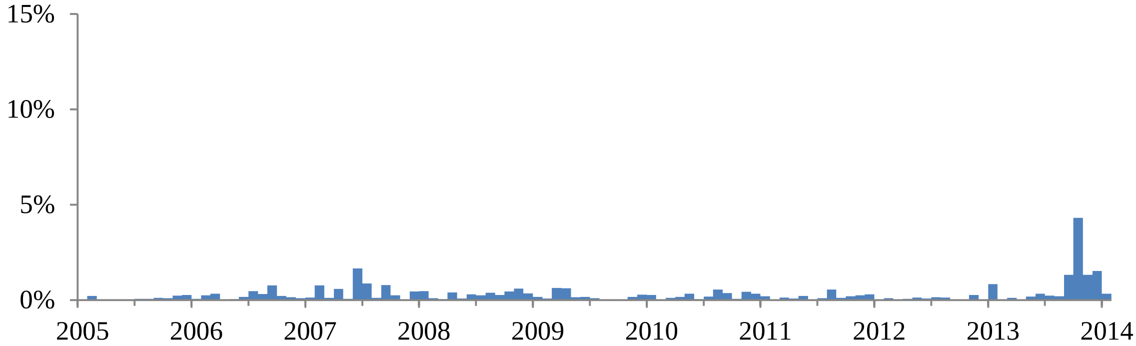


Figure 5
Pairwise fund migrations in the two-group solution

This figure shows the pairwise monthly fund migrations for the two-group solution, resulting from the hierarchical K-means classification. The labels of the groups that are given by the authors are based on the investment objectives and the return characteristics of the funds assigned to the groups

Panel A: from group I, “Variable Returns”, to group II, “Stable Returns”



Panel B: from group II, “Stable Returns”, to group I, “Variable Returns”

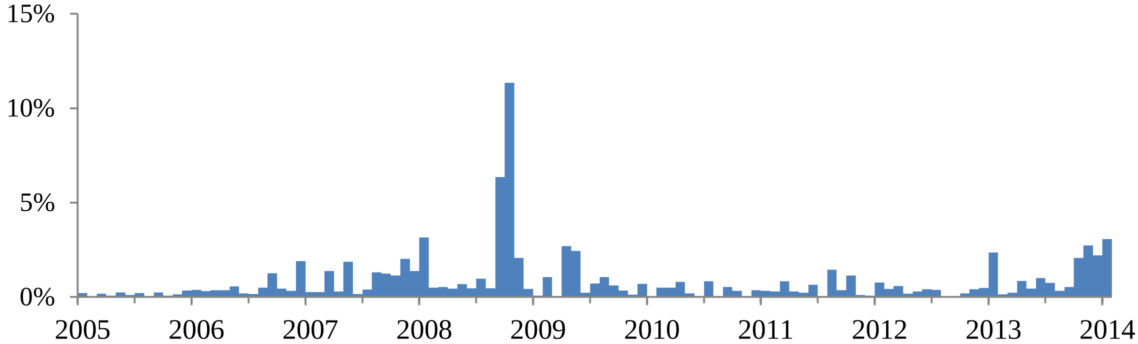
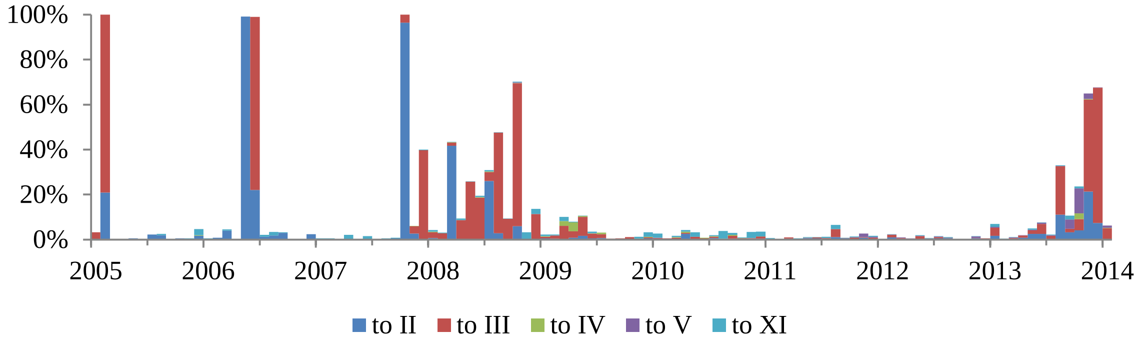


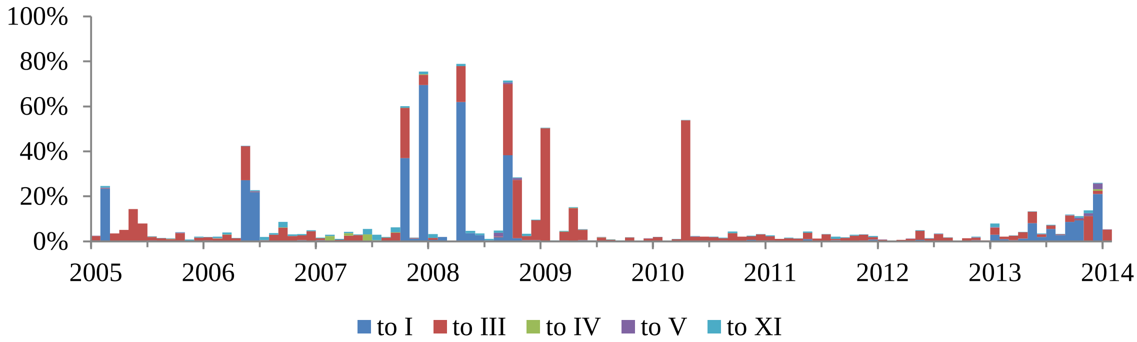
Figure 6
Pairwise fund migrations in the six-group solution

This figure shows the pairwise monthly fund migrations for the six-group solution, resulting from the hierarchical K-means classification. The labels of the groups are given by the authors based on the investment objectives of the funds assigned to the groups.

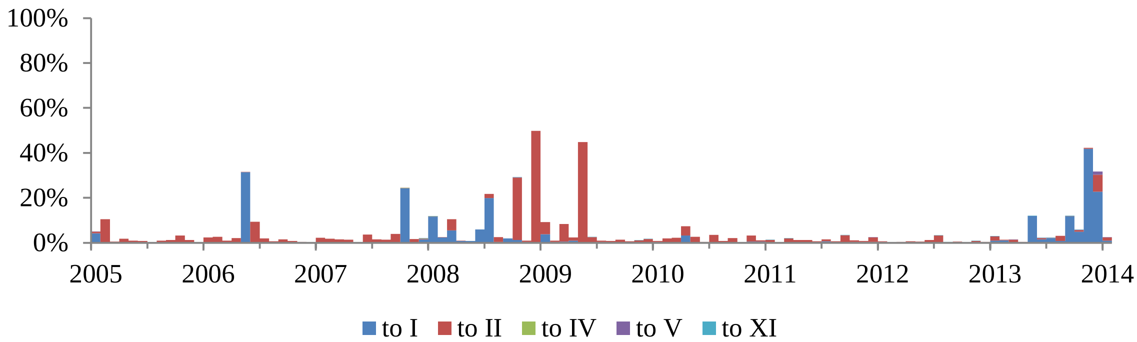
Panel A: from group I, “Bond High Yield/Equity Global”



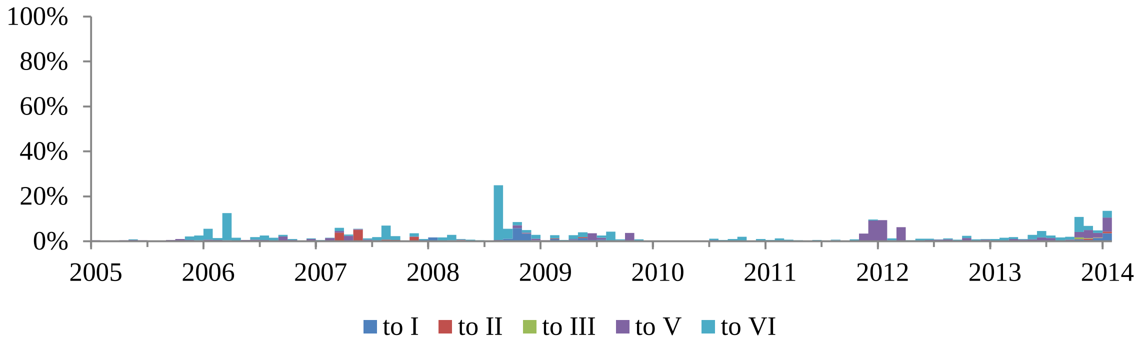
Panel B: from group II, “Equity Global”



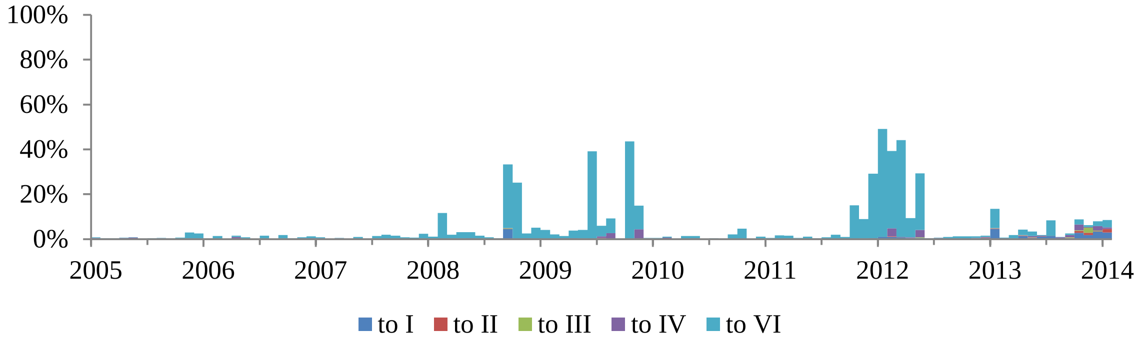
Panel C: from group III, “Equity Europe”



Panel D: from group IV, “Money Market”



Panel E: from group V, “Bond Eurozone”



Panel F: from group VI, “Bond Global”

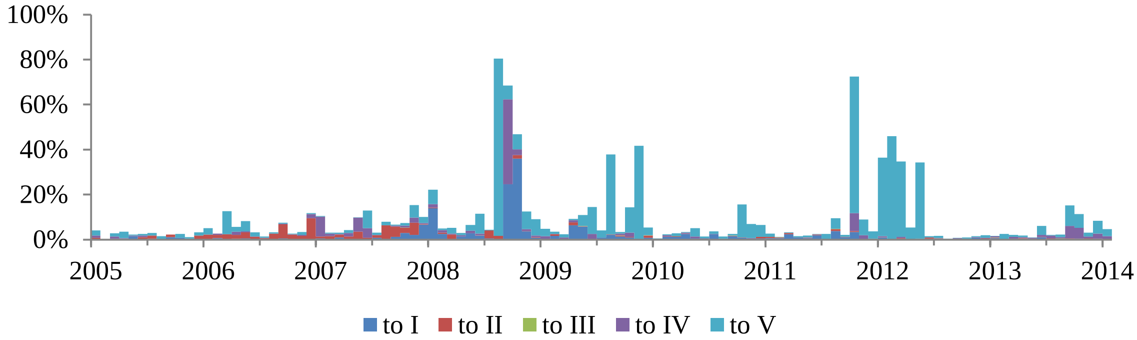
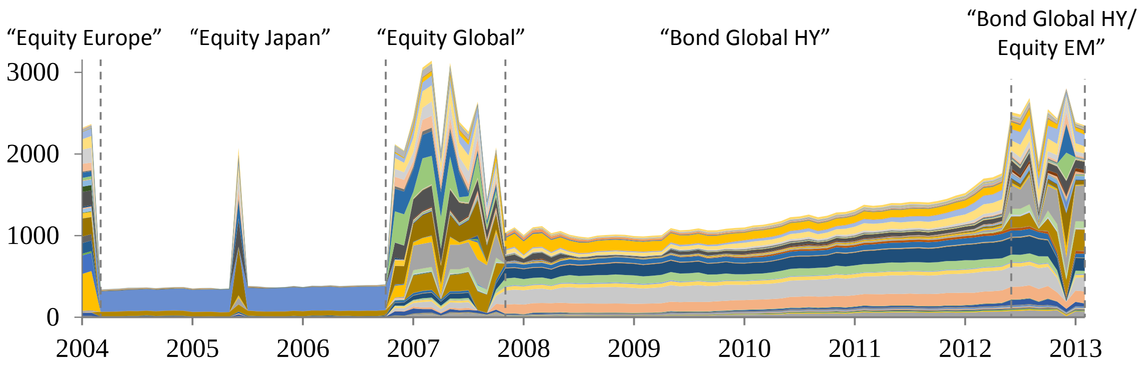


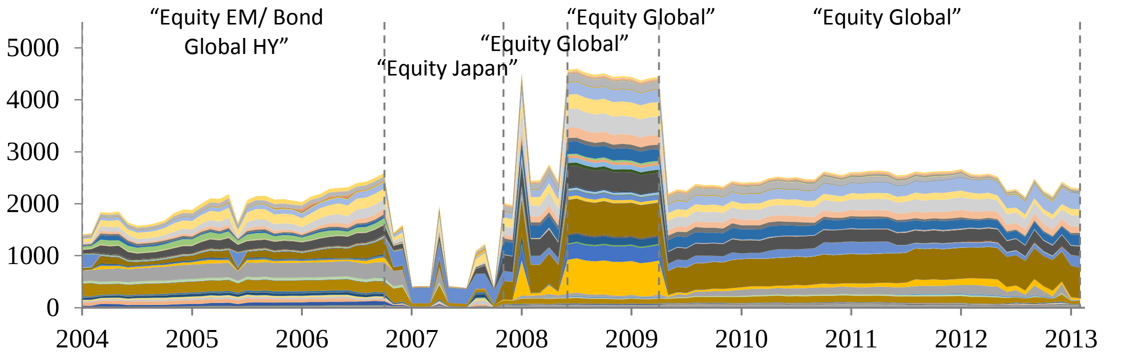
Figure 7
Investment objectives in the six-group solution over time

This figure shows the investment objectives comprising each group in each month for the six-group solution resulting from the hierarchical K-means classification. The labels of the groups are given by the authors based on the average investment objectives of the funds assigned to the groups.

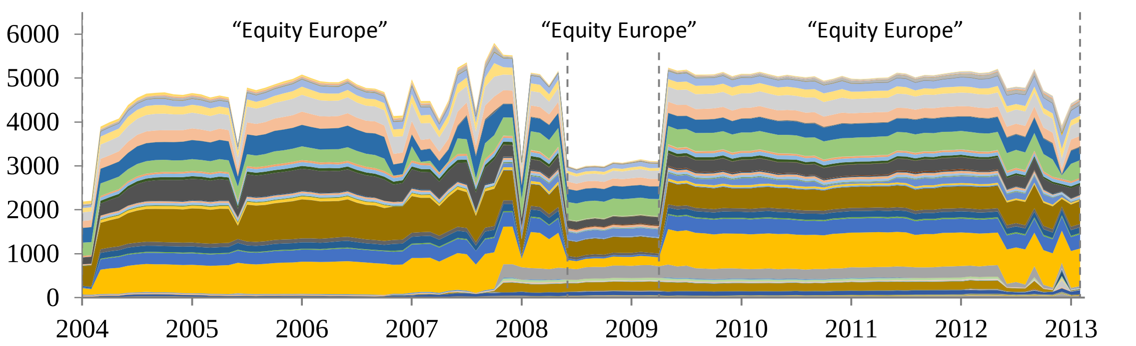
Panel A: Group I, "Bond High Yield/Equity Global"



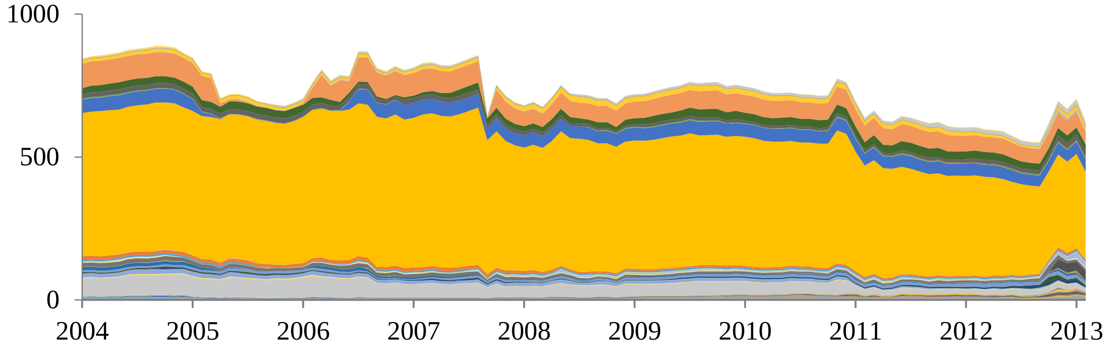
Panel B: Group II, "Equity Global"



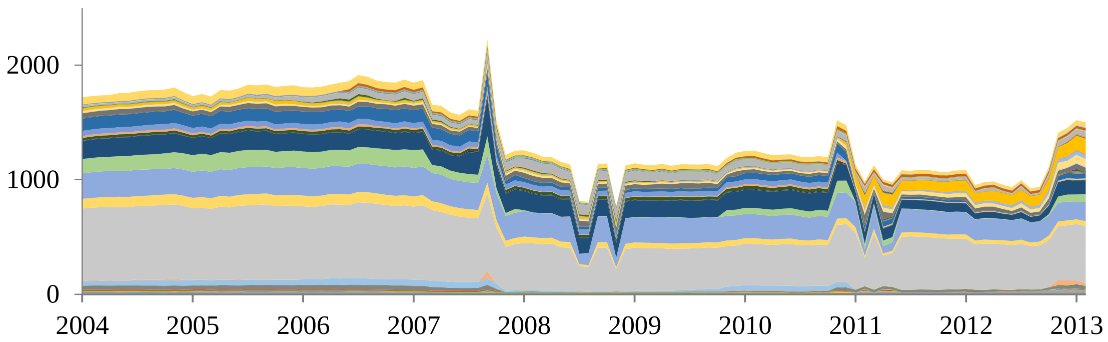
Panel C: Group III, "Equity Europe"



Panel D: Group IV, “Money Market”



Panel E: Group V, “Bond Eurozone”



Panel E: Group VI, “Bond Global”

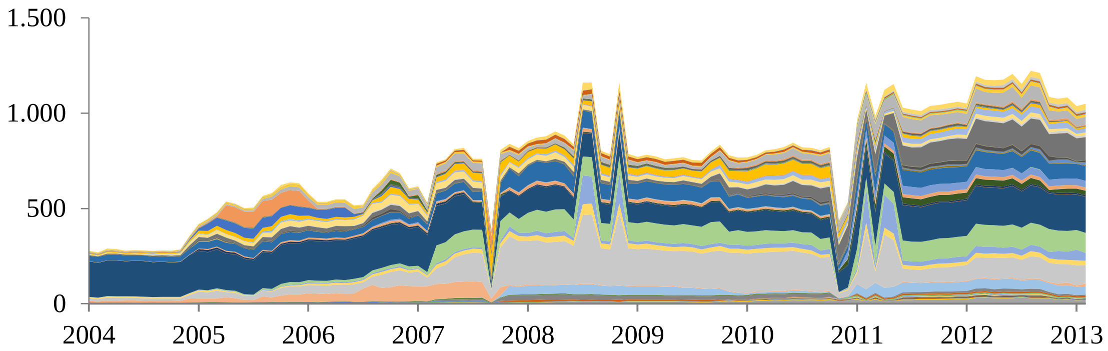


Table 6
Explanatory power of classification schemes

This table presents the average adjusted R^2 (standard deviation in parentheses) from regressing the cross-sectional fund returns against the hierarchical K-means (HKM) and the Lipper classification scheme respectively. The HKM classifies the funds based on their historical returns, while Lipper classifies the funds based on their self-declared investment objectives.

Groups	Adjusted R^2	
Panel A: HKM Classification		
	In-sample	Out-of-sample
2	0.18	0.18
	(0.03)	(0.15)
6	0.29	0.27
	(0.03)	(0.16)
15	0.45	0.41
	(0.03)	(0.13)
20	0.46	0.42
	(0.13)	(0.13)
Panel B: Lipper Classification		
	In-sample	
13	0.25	
	(0.20)	
134	0.51	
	(0.14)	
286	0.63	
	(0.11)	

Table 7
Peer group benchmarking adequacy of classification schemes

This table presents the average adjusted R^2 (standard deviation in parentheses) from regressing the return time series of each fund against the respective group's return index. The hierarchical K-means (HKM) classifies the funds based on their historical returns, while Lipper classifies the funds based on their self-declared investment objectives.

HKM Classification		*Lipper Classification*	
Groups	Adjusted R^2	Groups	Adjusted R^2
2	0.60	13	0.59
	(0.28)		(0.28)
6	0.69	134	0.68
	(0.22)		(0.28)
15	0.73	286	0.74
	(0.22)		(0.26)
20	0.76		
	(0.21)		