Factor based clustering

Fragkiskos, A. and Bauman, E.V.[1]

ABSTRACT

We propose a novel approach to cluster funds based on their factor exposures. The approach uses investment returns as input data and calculates similarity scores across funds, which are then used to form clusters. The derived clusters avoid common pitfalls that correlation based or other cluster methods fall into. They can be used as peer group alternatives to what vendors provide or to further refine existing categories that might be too obscure to make sense of. When tested against long/short equity funds, we find that we can form clusters with relatively high levels of stability across time.

1. INTRODUCTION

Investment professionals need to be able to evaluate and compare funds they have selected or track against funds that share similar traits. There are several reasons for doing so. Investors or advisors that select funds in a portfolio typically have to conform to policy decisions that allocate a certain portion of assets to each asset class, style or risk factor. Having well defined peer groups facilitates the manager selection and asset allocation process, either by replacing an existing fund with a similar one, picking funds that meet certain allocation targets or maximizing diversification by selecting least similar funds to include in a portfolio. Alternatively they might rely on peer group percentile rankings or performance attribution to select such funds. Asset managers that want to promote their internally managed funds need to be able to demonstrate superior performance against competitors, partially because if they fail to do so they stand higher chances of getting fired (Gary et al, 2014). Regulators who want to monitor funds for fraud find it useful to work within well-defined categories in order to detect suspicious patterns, flags or outliers. Index providers like Hedge Fund Research might want to construct investible indices of funds with common objectives.

There are various ways in the industry that peer groups are formed. Morningstar uses portfolio holdings to construct some of its categories. More complex categories, such as multi-alternative funds, are further classified based on their structure or stated strategy, i.e. whether they replicate a hedge fund index or are funds of funds. The process is partially quantitative and qualitative (Morningstar, 2016). Three-year averages are calculated for various analytics every six months, when classification takes place. Analyst reviews may override certain classifications.

[1] Apollon Fragkiskos, MSc, is Director of Research at Markov Processes International. Evgeny Bauman, Phd, is a senior quantitative researcher at Markov Processes International. The views in this paper are the responsibility of the authors, not the institutions they are affiliated with. Email: Afragkiskos@Markovprocesses.com, Ebauman@markovprocesses.com. We are grateful for comments from Michael Markov, Jeffrey Hansen, Jeff Schwarz, Dennis Baldi.

Hedge fund databases typically rely on the manager's self-proclaimed strategy, asset class, regional focus or liquidity to classify funds. Such approaches have been met with skepticism by practitioners. Portfolio holdings for mutual funds are disclosed infrequently and often with a delay of 60 days[2]. More complex strategy holdings that contain derivatives make it very hard or impossible to infer a fund's strategy due to the challenge (leverage, sensitivities, correlations) of associating such positions with more vanilla instruments or understanding their risk or return profile. A manager's self-proclaimed strategy or regional focus may not be accurate, ignoring style shifts that may happen over time as documented in various studies (diBartolomeo & Wikowski (1997), Brown & Goetzmann (1997), Kim, Shukla & Tomas (2000), Cremers and Petajisto (2007), and Mason et al. (2012)).

To overcome such limitations, practitioners often resort to their own qualitative and quantitative research in order to better understand and group funds. Das (2003) uses a nonhierarchical clustering method to classify hedge funds based on their asset class, fund size, incentive fee, risk and liquidity. Gibson et al (2007) use a hard clustering procedure to classify hedge funds based on correlation of their returns. Baghai-Wadji et al (2005) have used Self-Organizing Maps (SOM) to find homogeneous groups of hedge funds based on similar characteristics. They identified nine hedge fund classes based on a 10-year sample of 2,442 dead and active hedge funds. Moreno et al. (2006) have proposed a method to improve the classification of Spanish mutual funds using SOM, k-nearest neighbors and the k-means algorithm. They reported that over 40% of mutual funds were misclassified in the official Spanish mutual funds classification. Alimi et al. (2012) have proposed a method to cluster mutual funds using Ward method and k-means with fuzzy theory based on six characteristics including rate of return, variance, semivariance, turnover rate, Treynor index and Sharpe index for multi-objective portfolio optimization. Mason et Al (2005) use k-means clustering, based on Sharpe's returns-based style analysis and exposures against common style indices, and find that it explains a significant proportion of the cross section of out of sample returns.

In this paper, we propose a quantitative three-step clustering approach that can address some of the limitations that appear with some of the existing approaches. At the heart of this approach is a novel similarity measure that takes into account similarities across different factor exposures. Here is that three-step clustering approach:

a) Derive the funds' main bets via returns-based style analysis
b) Define factor similarities that distinguish factors by asset class, geography, style and other characteristics
c) Cluster funds based on factors, factor exposures and their similarities

This article is organized as follows. The next part demonstrates common pitfalls investment professionals run into when using holdings-based similarities or correlation-based approaches to form peer groups. Following that, we present the methodology behind our approach. We

[2] https://www.sec.gov/rules/final/33-8393.htm#foot79

start with a short description of Sharpe's returns-based style analysis, which is used to generate the data for the formation or peer groups. Next, we introduce and describe our similarity measure. The next part describes the mutual fund data used in our experiments and the features of mutual funds that are formed. There, we demonstrate certain factor characteristics of the peer groups formed against Morningstar's liquid alternative funds and compare against correlation. The last part concludes.

2. CURRENT CHALLENGES

2.1 Similarity pitfalls

Clustering based on features is a well-studied problem. To the best of our knowledge, however, taking into account similarities across different features is a problem that has not been considered before in a portfolio framework. When using holdings-based approaches, for example, to estimate similarity among funds, the inability of available measures to deal with similarities among different holdings may result in funds being placed in the wrong peer groups. Sakakibara et al (2015), for example, perform clustering based on the percent of common stocks that mutual funds invest into, based on their top 10 stock holdings. As Sakakibara puts it:

"A future challenge is a classification that considers the degree of similarity between stocks. Although there are several stocks that are similar (e.g., Toyota and Honda), they are determined to be invested in totally different stocks by the proposed method it may be possible to split mutual funds better if we can ameliorate this problem…. Clustering methods used here to classify the given data always cluster somewhere, but there are instances that are not similar to any others in the mutual funds. Therefore, there is a chance of being wrongly grouped into a cluster."

The same issue arises when trying to calculate similarities among funds based on their factor exposures. To gain some intuition, table 1 shows a hypothetical fund A with certain exposures to a few factors. Underneath that, we show several funds with different exposures. Assuming the betas to each factor as the feature set, the Euclidean distances of each of those funds to fund A are all equal. For the investment professional, however, most of these funds are quite different than fund A, and certain funds would be perceived to be more different than others. The fund called "irrelevant" for example invests solely in commodities. This fund should not be expected to be as similar as other funds that invest in the same asset classes as fund A. While the complimentary fund does not invest in the same exact indices, one would expect the mismatch in maturities not to be as severe, in terms of similarity, as the difference in factors or betas observed in some of the other funds.

Table 1. Hypothetical Fund A and similar funds based on individual factors.

	5YR Corporate	10YR Corporate	5YR HY	10YR HY	10YR Swap	10YR Treasury	Commodity	Euclidean distance
Fund A	15	-	15	-	-	(30)	-	
Complimentary	-	15	-	15	29	-	-	51.4
Neutral	(6)	-	(6)	-	-	12	-	51.4
Concentrated	54	-	-	-	-	-	-	51.4
Diversified	14	14	14	14	14	14	14	51.4
Irrelevant	-	-	-	-	-	-	36	51.4

An approach commonly followed by vendors that produce reports based on factors is to group those into high-level categories, like the ones displayed at the top of Table 1. Doing that in order to estimate similarities, and later clusters, may provide false impressions. Table 2 shows what happens if, instead of working with individual factors, we work with grouped factors.

Table 2. Hypothetical Fund A and similar funds based on grouped factors.

	Corporate	HY	Cash	Commodity	Euclidean distance
Fund A	15	15	(30)	-	
Complimentary	15	15	29	-	59.1
Neutral	(6)	(6)	12	-	51.4
Concentrated	54	-	-	-	51.4
Diversified	27	27	27	27	65.7
Irrelevant	-	-	-	36	51.4

The complimentary fund now shows almost no distinction from fund A in the grouped corporate and high-yield exposures. More strikingly, the irrelevant fund has lower distance to fund A than the diversified fund.

To summarize, when looking at individual factors, traditional clustering techniques assume similar indices to be completely different. When looking at grouped factors, they assume them to be completely identical. We will return to this example after we define our similarity measure.

2.3 Correlation pitfalls

2.3.1 Funds

Some of the methods mentioned in the literature above rely on the correlation of historical returns, or a similar measure based on correlation. We can expect correlations, for the most part, to be able to identify funds with the same strategy, but like any other quantitative measure, correlation has its limitations. To demonstrate that, we examined a pair of funds from the equity long/short mutual fund category of Morningstar that have a high correlation of

79%[3]. We also performed returns-based style analysis on each of these two funds using S&P 500 sectors. Results of this analysis are displayed in table 3.

Table 3. Returns-based style analysis using weekly returns, two years prior to 2015/11/20. Factors are grouped according to their economic properties.

	Energy	Materials	Pro-cyclical				Counter-cyclical				Foreign	
	SP500 Energy	SP500 Materials	SP500 Financials	SP500 Inf Technology	SP500 Industrials	SP500 Consumer Discret	SP500 Telecomm Svcs	SP500 Utilities	SP500 Consumer Staples	SP500 Health Care	MSCI AC Asia Pacific USD	MSCI AC Europe USD
Glenmede Long/Short	-	-	20	-	22	-	30	-	-	-	-	-
James Long Short	-	7	-	-	-	-	15	32	-	13	-	37

What we observe is that despite the high correlation, the two funds seem to invest in completely different sectors. An investor that picks funds based on their equity-sector or macro-economic focus would not have considered these two funds to belong in the same group.

The same exercise over the same time range was performed against the multi-alternative fund category of Morningstar. This time, we picked a pair of funds with low correlation of 80%[4] and examined their factor exposures, which we display in table 4.

Table 4. Returns-based style analysis using weekly returns, two years prior to 2015/11/20.

	MBS	HY	MSCI EU	Large Value	Large Growth	Small Growth	Precious Metals	Energy
Absolute Strategies I	1	12	(1)	(4)	(28)	(7)	8	-
JHancock Alternative Asset Allc A	-	10	-	-	23	-	-	2

This time we observe that both funds are primarily exposed to equities, and in particular, large growth stocks. It would probably make more sense to compare these funds together as opposed to grouping them with fixed income or commodity focused funds. Correlation has no knowledge of the funds' exposures, essentially being fooled by the residual long vs. short equity exposure, and would have indicated they are very different.

2.3.2 Indices

In Table 5 we show the correlation matrix among major market indices, and in chart 1 we display the results of hierarchical clustering based on those correlations.

Table 5. Correlation matrix of major market indices based on monthly returns 10 years prior to 3/31/2016.

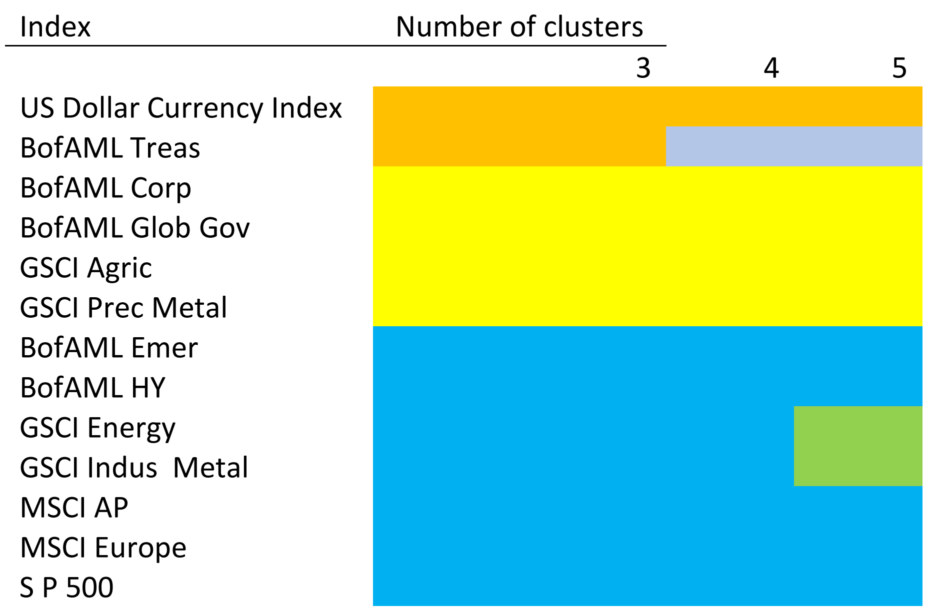
---

[3] In terms of how this ranks against all other fund pairs of the same category, it is 78% rank (100% being the highest).

[4] In terms of how this ranks against all other fund pairs of the same category, it is 0% rank (nearly lowest ranked pair across all others).

	DXY	BofAML Treas	BofAML Corp	BofAML HY	BofAML Glob Gov	BofAML Emer	S P 500	MSCI AP	MSCI Europe	GSCI Indus Metal	GSCI Prec Metal	GSCI Energy	GSCI Agric
DXY	100%	-1%	-36%	-47%	-79%	-56%	-51%	-61%	-70%	-58%	-46%	-56%	-54%
BofAML Treas		100%	41%	-25%	47%	5%	-31%	-22%	-27%	-24%	23%	-35%	2%
BofAML Corp			100%	66%	50%	70%	37%	53%	45%	26%	33%	14%	32%
BofAML HY				100%	26%	77%	74%	77%	76%	51%	21%	46%	37%
BofAML Glob Gov					100%	41%	21%	35%	38%	35%	49%	20%	44%
BofAML Emer						100%	61%	71%	68%	48%	36%	39%	37%
S P 500							100%	83%	89%	53%	10%	48%	34%
MSCI AP								100%	90%	58%	24%	52%	39%
MSCI Europe									100%	61%	22%	55%	41%
GSCI Indus Metal										100%	39%	54%	40%
GSCI Prec Metal											100%	32%	36%
GSCI Energy												100%	36%
GSCI Agric													100%

Chart 1. Hierarchical cluster membership of major indices based on correlation of monthly returns 10 years prior to 3/31/2016.



Even by using long-term correlations, we observe that different asset classes are mixed together in the same cluster groups. For the investment professional, bucketing together agricultural commodities and global government bonds, two entirely different asset classes with different risk and return drivers and trading characteristics, can be counter-intuitive.

2.4 Stability

Depending on the use case, clusters may need to have a low turnover over a semi-annual, annual or longer time frame. For most practical applications mentioned in the introduction, high stability is desirable. Commonly used approaches offer various degrees of stability which may or may not be adequate.

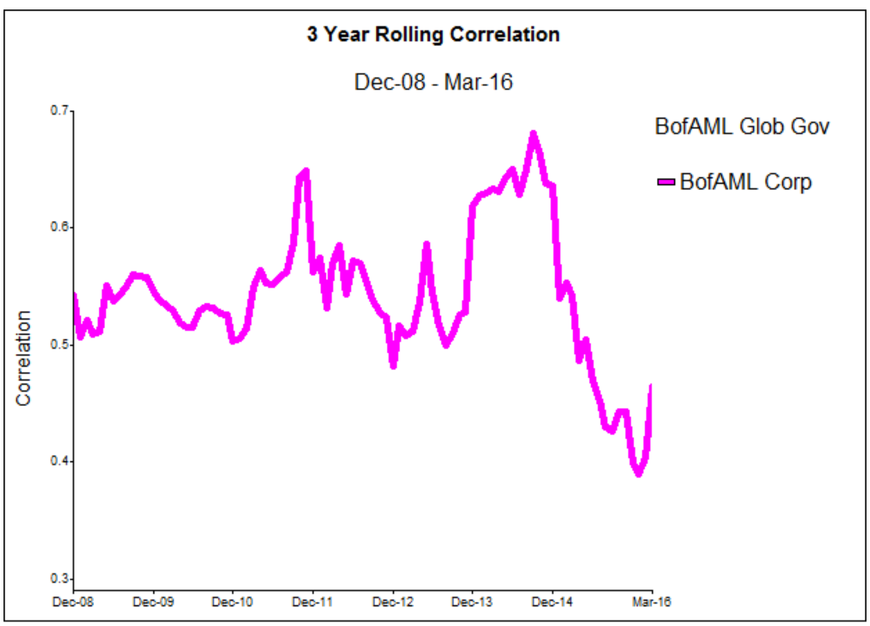
2.4.1 Holdings

The Royce Funds[5] compiled historical fund category data, based on year-end snapshots from 2003, 2008 and 2010, of all the funds within the small and large value, growth and blend Morningstar categories, including funds that had been merged or liquidated. They then compared each snapshot to the peer group as of end of March 2014. They found that 17% of the funds left the small value category since 2010, 27% since 2008 and 32% since 2003. The other two categories were more stable. Jeffrey Ptak from Morningstar[6] calculated the median rolling five-year percentage of funds that either stayed within the same category, left it or died. The calculation was performed over the period of June 1996 to May 2011 for the combination of small, mid, large and growth, value, blend categories. He found that the percentage of funds leaving their category ranges from as low as 10% for the large growth category, to as high as 50% for the mid value category.

2.4.1 Correlation

Using correlations as a measure can be very unstable. Even with a three-year window of monthly correlations among indices of the same asset class, the values may be widely different among periods, as shown in chart 2. It is unlikely that the perception of similarity among these two indices by an investor with a long-term view would vary as much over time.

Chart 2. Rolling window correlation between Global government bonds and US Corporate bonds over a period that spans the last 10 years prior to 3/31/2016.



Created with MPI Analytics

3. Similarity

[5] https://www.advisorperspectives.com/articles/2014/06/24/how-morningstar-category-flux-impacts-peer-group-analysis
[6] http://www.morningstar.com/advisor/t/45786651/in-practice-pitfalls-of-peer-groups.htm

In this section, we present our framework for calculating similarity among funds. We start by describing the building blocks of the whole process. The first step is returns-based style analysis, which we use to provide transparency into each fund. We proceed by describing the factor similarities and, right after that, we introduce a new measure to calculate fund similarities that takes into account similarities among different betas and factors.

3.1 Style analysis

The first step in the process enables us to overcome the limitations associated with using position or aggregate fund data. For that purpose, we can use returns-based style analysis (originated by Sharpe in 1992), a technique that aims to replicate the performance of a fund over a time period via a set of factors, typically market indices, and forms a factor mimicking portfolio. The extent to which the factor exposures obtained are similar among funds will determine their fund similarities. The key advantage of style-analysis over holdings approaches is that returns are generally more up to date than holdings, and the factors used offer good, easily digestible insights into a fund's main bets, which often overcome limitations associated with complex or numerous holdings.

Sharpe's approach assumes betas are constant over the window that regression takes place. This is generally not true, particularly for certain categories of funds like the long/short equity funds examined later in the paper. A way to get around this is to perform rolling regressions over shorter windows within the entire analysis period. Each window would estimate a different beta, and by compiling betas across all windows, we could retrieve, to a certain extent, a fund's dynamic exposures. Note that each window's beta gets assigned to the window's last date. That beta, however, represents the average beta over the window. So, the dynamic beta pattern observed over the entire period is a collection of moving average betas. This is not the same as obtaining the beta that corresponds to each point in time.

To be able to model funds with dynamic long or short exposures, we can use a model that does not make any assumptions on constant betas or long only constraints, like the dynamic style analysis (DSA) model developed by Markov et al (2004). See appendix C for more details.

When analyzing funds, one may have a narrow set of factors in mind. Alternatively, one may not be sure which factors should be in the final model, in which case he could employ model selection techniques against a universe of factors to determine the best model, such as stepwise regression. When calculating similarities across different types of funds, our measure does not require funds to use the same factors; each fund can use its own unique factors. The similarity calculations are designed to take into account all factors found across all funds, assuming zero betas on factors that are not part of each fund's model. Although this introduces a different similarity scale vs. only working with a fund's final model, it doesn't change the similarity rank between funds that employ a common set of factors or between funds that employ one set of factors and funds that employ a different set. Of course, the validity of the calculated similarities and clusters depends upon using proper factors. If one ends up using fixed-income factors on equity funds or vice versa, the final similarities will not make much sense, so care has to be taken to define at least a reasonable starting universe. On the flipside, if one equity factor is included in the final model in favor of a different but similar equity factor, the fund similarities and clusters, as will be shown later on, will not be significantly affected.

3.2 Factor similarity

Our approach enables the use of similarity among factors based on a single or combination of measures that the investor has in mind when drafting his investment policies, which may or may not include correlation. Alternatively, the similarity may be based on intuition, prior experience or expectations regarding how similarly various indices will behave. We do not attempt to quantify such similarity, rather we propose a framework that can accommodate a set of similarity scores set by the user[7]. This step essentially requires the user to supply a two-dimensional matrix of similarities across all factors, similar to how a correlation matrix is defined.

$$S_X\left(x_i, x_j\right) = \text{externally supplied}$$

where x_i and x_j represent the time series of returns of factors i and j. The factor similarity can take values between 0 and 1, with 0 indicating absolutely no similarity and 1 indicating perfect similarity.

If the regression is not based on the budget constraint then, purely for similarity calculation purposes, we have to assume an artificial cash factor with beta equal to one minus the sum of betas across all other factors. This cash factor should typically have high similarity with risk free government bonds and low with stocks.

3.3 Beta similarity

We define the similarity among two betas to be linearly related to the distance between their values[8].

$$S_B\left(\beta_i^A, \beta_j^B\right) = 1 - \left| \frac{\left|\beta_i^A\right|}{\sum_i \left|\beta_i^A\right|} - \frac{\left|\beta_i^B\right|}{\sum_i \left|\beta_i^B\right|} \right| \quad (3)$$

where i and j represent factors and A, B represent funds A and B. Beta similarity can take values between 0 and 1, with 0 indicating absolutely no similarity and 1 indicating perfect similarity.

As evident from (3), the sign of betas is ignored. That is because for most applications, like when creating peer groups, we try to find funds that invest in similar factors. Whether a

[7] This is similar practice to users providing market forecasts to be used in a Black-Litterman portfolio construction framework.

[8] Our similarity measure focuses on comparing beta values that either belong in the same time t, typically the most current betas, or values that represent the average over a time period. In order to calculate a similarity score that takes into account betas (and factors) at each point in time, we can calculate the similarity scores at each point in time and then average those scores. This is explained further in the experiment section.

manager is long one factor and another manager is short the same factor, it still means that both managers trade the same factor, which will push them toward the same peer group.

3.4 Similarity among betas *and* factors

Based on the betas to a given set of factors (3.3) and the similarities among factors (3.2), we can now proceed to calculate similarities among funds.

$$S_{BX}\left(\beta_i^A x_i, \beta_j^B x_j\right) = S_B\left(\beta_i^A, \beta_j^B\right) * S_X\left(x_i, x_j\right) \quad (4)$$

The similarity between the two terms above is decomposed to similarities among terms of the same nature, as the similarity between betas and factors is not defined. The reason the two terms on the right hand side of (4) are multiplied is that, for example, in the extreme case that either of those terms has zero similarity, it should force the total similarity on the left hand side to be zero. For example, if a fund has very small beta to factor i, then, even if factor i has high similarity to factor j, the two funds should not be considered to be similar. In that case, similarity will be close to zero but not negative as similarity is defined in the range [0,1].

3.5 Similarity among betas *and* factors and idiosyncratic

$$S_{U,BX}\left(u^A, \beta^B x\right) \equiv 0 \quad (5)$$

Where u represents the time series of the idiosyncratic returns, $S_{U,BX}\left(u^A, \beta x\right)$ represents the similarity between fund's A idiosyncratic returns and a beta exposure βx of fund B and any factor.

3.6 Idiosyncratic similarity

$$S_U\left(u^A, u^B\right) \quad (6)$$

The estimated idiosyncratic nature of a fund will typically have little to no similarity to the idiosyncratic nature of any other fund. This assumes that the models used on each fund do not leave out any significant missing factors that present similarities with factors present in other funds. Assuming care has been taken to select a proper starting universe or set of factors, such assumption should generally hold. This does not imply, and it's certainly not required, that the model employed perfectly explains a fund. To take this into account, we are adjusting similarities by the model's explanatory power in step 3.9. So, typically we can assume that (6) equals to zero. Alternatively, we can estimate it based on the idiosyncratic correlations among two funds. In practice, such correlations are most of the time very small.

3.7 Systematic similarity

$$\text{Systematic similarity: } S_{SYST}^{A,B}\left(\sum_i b_i^A x_i^A, \sum_j b_i^B x_i^B\right) = \sum_i \sum_j S_B\left(b_i^A, b_j^B\right) * S_X\left(x_i^A, x_j^B\right) \qquad (7)$$

This is the similarity among two funds calculated by excluding the idiosyncratic term. The measure goes through all the combinations of beta and factor exposures between the two funds and sums those up.

3.8 Rescaled systematic similarity

$$\text{Rescaled Systematic similarity: } S_{SYST}^{A,B\,*} = \frac{S_{SYST}^{A,B}}{\sqrt{S_{SYST}^{A,A} * S_{SYST}^{B,B}}} \qquad (8)$$

The reason that we rescale the systematic similarity by dividing with the above denominator is to achieve the same scale across funds. This is analogous to how covariance is rescaled by the individual standard deviations (square roots of variance) to arrive at the correlation measure.

3.9 Fund similarity

$$\text{Fund similarity: } S^{A,B} = R^{2,A} * R^{2,B} * S_{SYST}^{A,B\,*} + \left(1 - R^{2,A}\right) * \left(1 - R^{2,B}\right) * S_U\left(u^A, u^B\right) \qquad (9)$$

This is the rescaled systematic similarity adjusted by the explanatory power of each fund's model. It is the similarity measure we use for clustering purposes.

In a regression framework, we cannot expect a model to perfectly explain what a fund does; the more complex the fund, the lower the explanatory power will be. The above equation adjusts the systematic similarity by R^2, a measure of how good the fit is. The reason we multiply the two R^2 is that, if a fund's model has zero explanatory power, we also want the similarity among funds to be zero, because in that case, a fund's factors and exposures cannot be trusted to measure the similarity among the two funds. This effectively balances the total similarity between the systematic and idiosyncratic portion[9].

In a clustering framework, what matters most is how the explanatory power of each model ranks against other models, not how good the model is. This enables us to properly cluster funds even if they belong in a complex category with moderate explanatory power (for example market neutral hedge funds). Of course, if such explanatory power is very low among most funds, then the clusters formed should be used with caution.

4. Clustering

Now that we have defined a similarity score, we can proceed to use it for clustering purposes in the same way we could use other similarity measures to perform clustering, such as

[9] A point to be clarified here is that a model may be very well constructed such that it only contains statistically significant exposures. This will typically, but not necessarily, coincide with a high R^2. The opposite may also be true; statistically insignificant coefficients may be associated with a high R^2. We believe R^2 to be a more generic measure that represents how well a fund is explained in our framework than t-stats or p-values.

correlation or Euclidean distance. The main clarification that needs to be made here concerns the time varying aspect of betas.

In an evolutionary clustering framework, we are interested in producing a sequence of clusters, one for each point in time. For the particular application of peer grouping, such sequence can be formed with a low frequency, such as twice per year. There are two criteria we need to keep in mind when forming such clusters. First, we require the cluster groups to be as consistent as possible, having a low turnover, or history cost, between subsequent clustering events. Second, we want the updated clusters to be as close to the current dataset as possible, i.e. having a high snapshot quality. The history cost is adversely affected by the fact that funds that had been part of a category are no longer part of it and new funds have been added that didn't exist the last time clustering took place. The snapshot quality will present variation as funds temporarily adjust their risk appetite or shift assets according to market conditions. In practice we seek a balance between the two.

In our framework, snapshot quality is determined by the most recent betas (t=T), which indicate the current allocations of a fund to factors. Those betas produce the current similarities among funds. In order to take into account the historical evolution of each fund's factors and exposures, we can apply a windowing technique that produces a moving average or exponentially moving average measure, as in Chakrabati et al.

$$\text{Moving average: } {}^{mavg}S_{A,B,t} = \sum_{i=T-p}^{T} S_{A,B,i}$$

$$\text{Exponentially weighted moving average: } {}^{EWMA}S_{A,B,t} = (1-\beta)S_{A,B,t} + \beta\, {}^{EWMA}S_{A,B,t}$$

To compare clusters over time, we need to ensure that the numbers of each cluster are consistently marked so that we maximize overlap of funds among two consecutive clusters. For that, we follow the approach described in Bauman (1982).

5. Case studies revisited

In this section, we revisit the examples presented in the introduction of the paper to show how the new measure can help provide intuitive solutions.

5.1 Similarities

We return to the example of section 2, in order to calculate similarities among fund A and a given set of funds. Following the steps outlined in section 3, we need to define a similarity matrix among the factors. An example of such matric is shown in table 6. It is based on intuition, like the fact that nearby term structure points are more similar than ones that are further away. We also assume that risk free asset classes are more similar to each other than risk free vs. risky ones. This matrix is a reasonable suggestion and different users may have different views of what the similarities are among indices. They may even measure those using a combination of quantitative measures. Our framework allows for all sorts of user inputs.

Table 6. Example similarity matrix across some indices.

	5YR Corporate	10YR Corporate	5YR HY	10YR HY	10YR Swap	10YR Treasury	Commodity
5YR Corporate	1	0.9	0.8	0.6	0.5	0.5	0
10YR Corporate		1	0.6	0.8	0.5	0.5	0
5YR HY			1	0.9	0.25	0.25	0
10YR HY				1	0.25	0.25	0
10YR Swap					1	0.9	0
10YR Treasury						1	0
US EQ							1

Given we know what the betas are, we can proceed in calculating similarities among the funds. Table 5 displays the systematic similarities where for intuition we have omitted results that contain R^2.

Table 7. Rank of similarity between a group of funds and fund A.

	5YR Corporate	10YR Corporate	5YR HY	10YR HY	10YR Swap	10YR Treasury	Commodity	Similarity rank
Fund A	15	-	15	-	-	(30)	-	
Complimentary	-	15	-	15	29	-	-	1
Neutral	(6)	-	(6)	-	-	12	-	2
Concentrated	54	-	-	-	-	-	-	3
Diversified	14	14	14	14	14	14	14	4
Irrelevant	-	-	-	-	-	-	36	5

Our measure identifies the complimentary fund to be the one most similar to fund A, as one would intuitively expect. The fund that ranks second is the neutral fund, with exposures in a different sign than fund A, but similar in magnitude, while investing in the same factors as fund A does. The concentrated and diversified funds are next, with ranks determined by the measure. The irrelevant fund scores the lowest as it invests in an entirely different asset class. We believe these results to be closer to what one would expect the similarity of the funds to be, given prior factor similarities, than to the ones produced by available feature clustering procedures.

5.2 Global indices

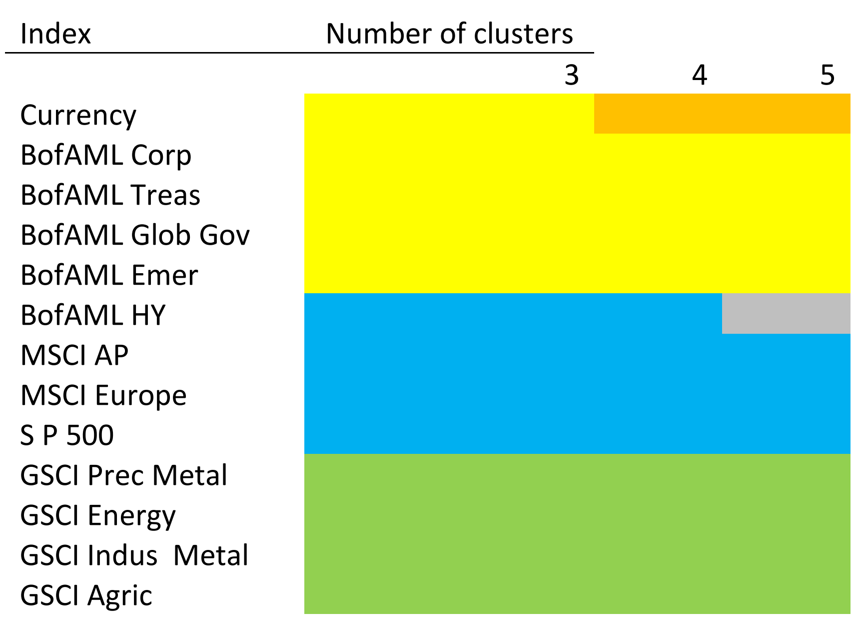
Starting at the index level, table 2 displays an example similarity matrix across major indices, where for simplicity, we have assumed no similarity among indices of different asset classes. The similarity of the indices within each asset class is driven by credit quality, in the case of fixed income, region, in the case of equity, and type of commodity in the case of commodity.

Table 8. Example similarity matrix across major global indices.

	BofAML Treas	BofAML Corp	BofAML HY	BofAML Glob Gov	BofAML Emer	S P 500	MSCI AP	MSCI Europe	GSCI Indus Metal	GSCI Prec Metal	GSCI Energy	GSCI Agric
Currency	-	-	-	-	-	-	-	-	-	-	-	-
BofAML Treas		50%	50%	80%	50%	-	-	-	-	-	-	-
BofAML Corp			80%	50%	50%	-	-	-	-	-	-	-
BofAML HY				25%	25%	50%	50%	50%	-	-	-	-
BofAML Glob Gov					50%	-	-	-	-	-	-	-
BofAML Emer						-	-	-	-	-	-	-
S P 500							80%	80%	-	-	-	-
MSCI AP								80%	-	-	-	-
MSCI Europe									-	-	-	-
GSCI Indus Metal										80%	80%	80%
GSCI Prec Metal											80%	80%
GSCI Energy												80%

Given those similarities, chart 3 displays what clustering could look like.

Chart 3. Cluster membership of major indices based on similarity



The cluster memberships are now closer to intuition, with indices of the same asset class being consistently together, even when we have only chosen three clusters.

6. Experiments

6.1 Data

We now apply our approach to the Equity long/short category of funds in Morningstar. This is the largest category, in terms of AUM, among liquid alternative strategies. It is of particular interest as the funds within this category display dynamic beta properties and trading patterns that do not immediately lead to distinct groups of funds. The first step is to gather the funds that will be analyzed. These were collected using the criteria in table 9.

Table 9. Criteria for fund inclusion.

At least 36 months of historical return data
Live as of 11/20/2015
Primary share class

The universe of factors used during style analysis is shown in table 10.

Table 10. Universe of factors used to regress equity long/short funds.

Index	**Source**
Cash – ICE BofAML US 3-month Treasury Bill Index	ICE BofAML
SP500 Energy	S&P
SP500 Materials	S&P
SP500 Industrials	S&P
SP500 Consumer Discretionary	S&P
SP500 Consumer Staples	S&P
SP500 Health Care	S&P
SP500 Financials	S&P
SP500 Information Technology	S&P
SP500 Telecomm Services	S&P
SP500 Utilities	S&P
MSCI AC Asia Pacific USD	MSCI
MSCI AC Europe USD	MSCI

We used weekly returns from both data sets that span the period 1/4/2011 – 4/29/2016.

6.2 Style analysis

The next step is to perform regressions in order to produce the betas. These regressions were run weekly, spanning the period 1/4/2013 – 4/29/2016, using dynamic style analysis. Each week, the regressions used weekly fund and factor returns over a two-year rolling window, performing factor selection in order to isolate the most relevant factors for each fund. We applied the budget constraint so that all betas sum up to one. The results from the regression that we used during the beta similarity calculations were the most current betas.

6.3 Factor similarity

The similarity among factors takes into account the asset class, sector category and region where they belong. The sector category that each index represents is defined in terms of cyclical or counter-cyclical characteristics. Below we show such sector categories and associated similarities.

Table 11: Factor similarities.

	Pro-Cyclical				Counter-cyclical				Energy & Materials		Foreign	
	SP500 Financials	SP500 Inf Technology	SP500 Consumer Discret	SP500 Industrials	SP500 Utilities	SP500 Health Care	SP500 Consumer Staples	SP500 Telecomm Svcs	SP500 Energy	SP500 Materials	MSCI AC Asia Pacific USD	MSCI AC Europe USD
SP500 Financials	1	0.7	0.7	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.1	0.4
SP500 Inf Technology		1	0.7	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.1	0.4
SP500 Consumer Discret			1	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.1	0.4
SP500 Industrials				1	0.3	0.3	0.3	0.3	0.3	0.3	0.1	0.4
SP500 Utilities					1	0.7	0.7	0.7	0.3	0.3	0.1	0.4
SP500 Health Care						1	0.7	0.7	0.3	0.3	0.1	0.4
SP500 Consumer Staples							1	0.7	0.3	0.3	0.1	0.4
SP500 Telecomm Svcs								1	0.3	0.3	0.1	0.4
SP500 Energy									1	0.7	0.1	0.4
SP500 Materials										1	0.1	0.4
MSCI AC Asia Pacific USD											1	0.1
MSCI AC Europe USD												1

6.4 Number of clusters

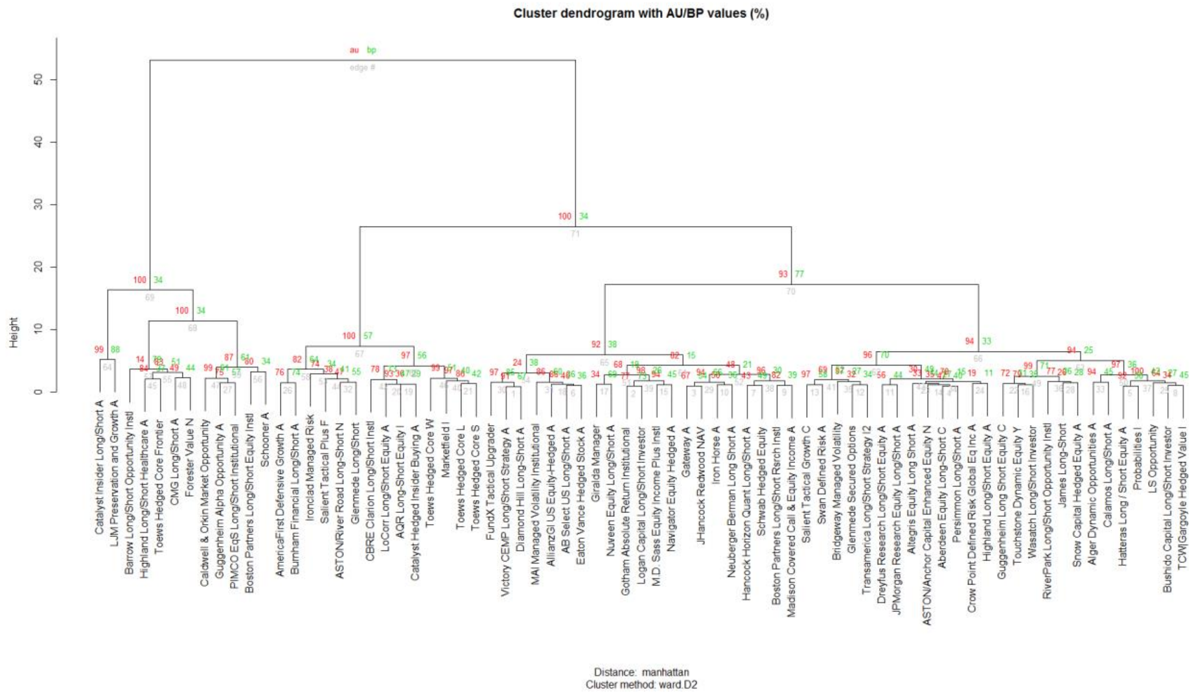
The steps that lead to the calculation of fund similarity are described in the methodology section. We use such similarities to form clusters. We include the calculation of moving average similarity in our experiments as described in section 4. For distinction purposes, we will describe the similarity that does not take into account past similarities as a current similarity. We also produced such clusters based on the correlation among funds in order to compare and contrast results. Each week, we calculated a correlation matrix among funds, using the same weekly fund returns over a two-year rolling window. Based on these correlations, we formed clusters of funds within each category and each week. Further results will be presented side-by-side for each approach.

We select to keep three clusters as the optimal number, taking into account the number of funds within each universe as well as stability considerations—the higher the number of clusters, the less stable we anticipate those clusters to be. Appendix D has some more details on how this gets estimated.

6.5 Cluster significance

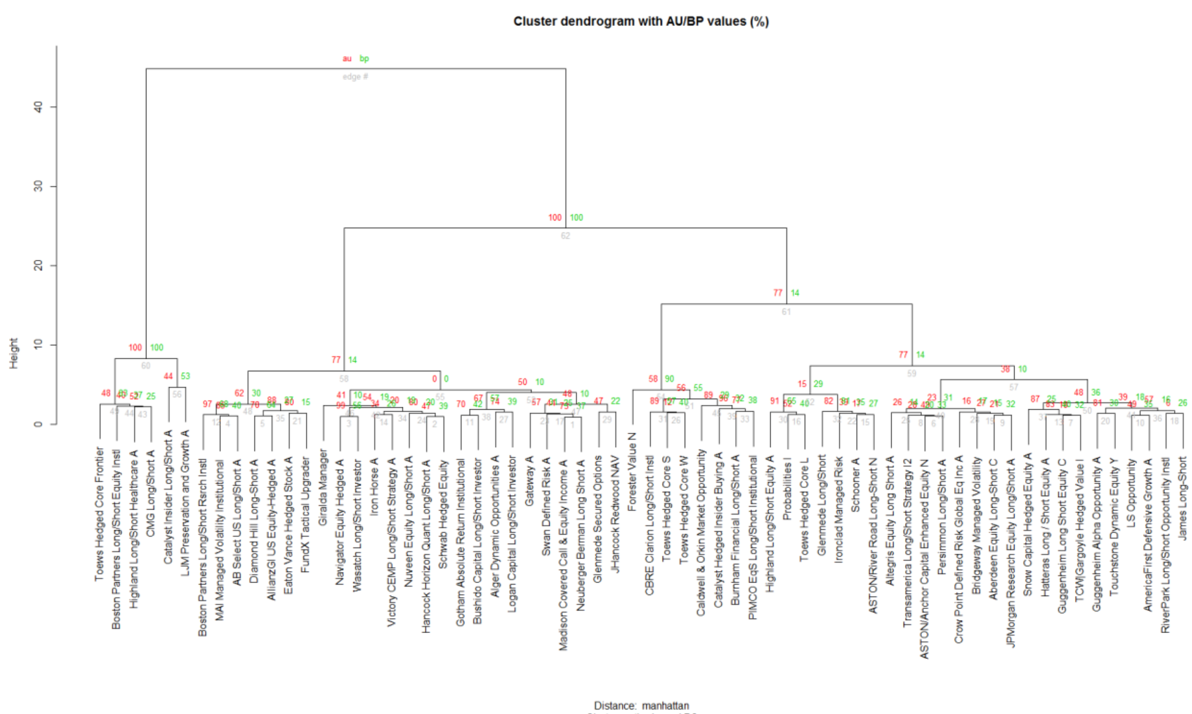
Taking the most recent date of our simulation as an example (4/29/2016), we plot the cluster dendrogram with computed p-values that indicate how strongly the data support the clusters obtained. We observe that in the case of three clusters, the approximately unbiased (AU) p-values, calculated via multiscale bootstrap resampling, are very high, indicating a strong structure.

Chart 4. Cluster dendrogram and p-values on 4/29/2016 based on current similarity.



We show the same for clusters based on moving average similarity calculated using current similarities based on past 52 weeks.

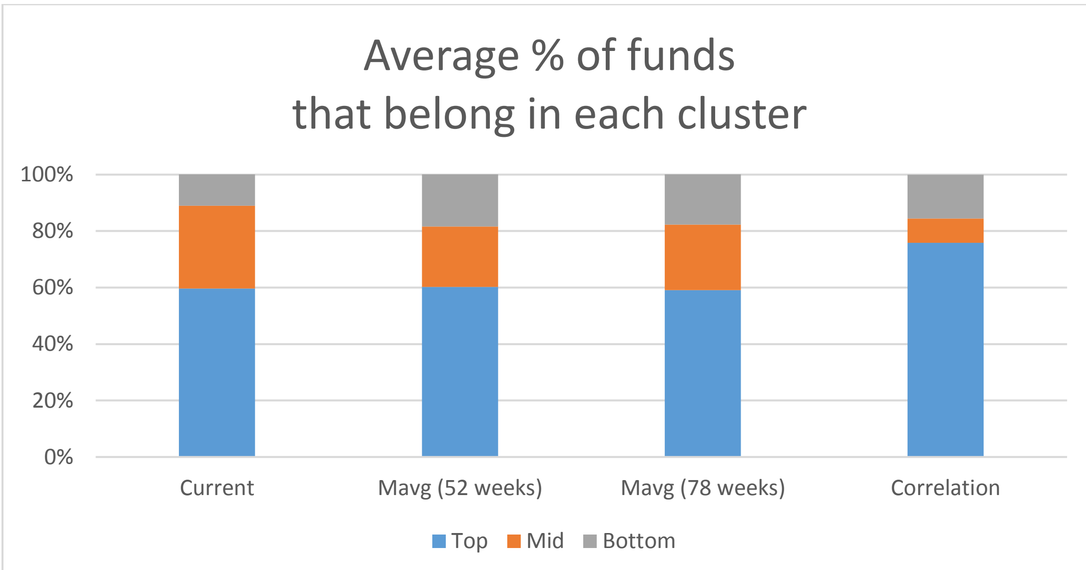
Chart 5. Cluster dendrogram and p-values on 4/29/2016, based on a 52-week moving average similarity.



6.6 Cluster membership

It is of practical importance that the funds analyzed are not heavily concentrated in one cluster. We calculate and display the average allocation of funds that belong in each cluster over the entire period. The results are being displayed for the current clustering approach, moving average and correlation based. Compared to correlation, we see that our approach results in more balanced clusters, in terms of the number of funds found in each. With a 76% average concentration of funds in the first cluster, the correlation based clustering does not seem very useful in terms of differentiating the way funds invest.

Chart 6. Fund allocation across clusters.



To get a more detailed look of how many funds are allocated across clusters, we display the time-varying allocation, based on the moving average and correlation approaches. We observe that with our measure cluster membership is more diversified than it used to be. Under the correlation measure, we note that there are many periods when the second cluster is consisted of just one fund.

Chart 7. Average percent allocation of funds into clusters calculated over the entire period, based on a 52-week moving average similarity.

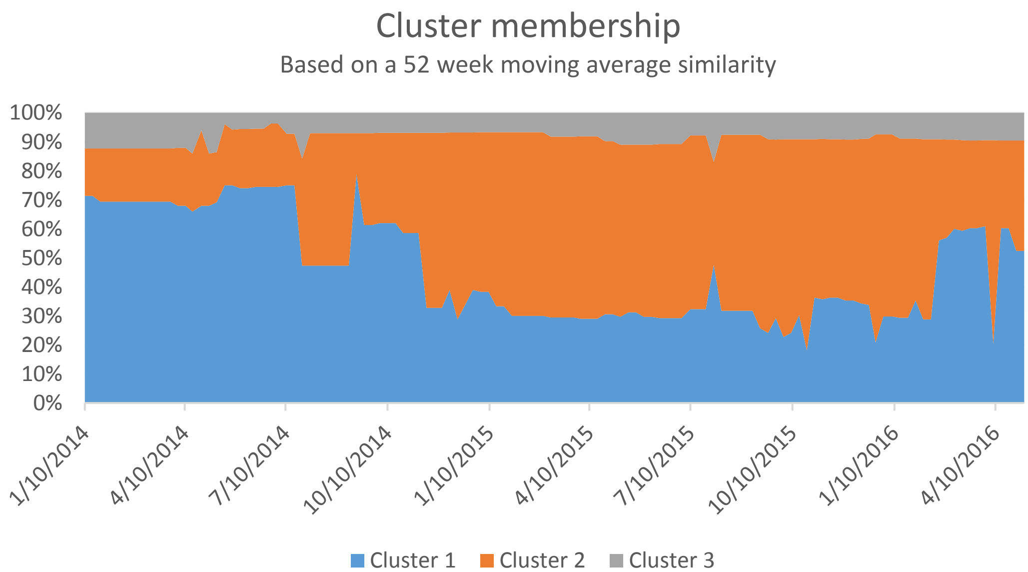
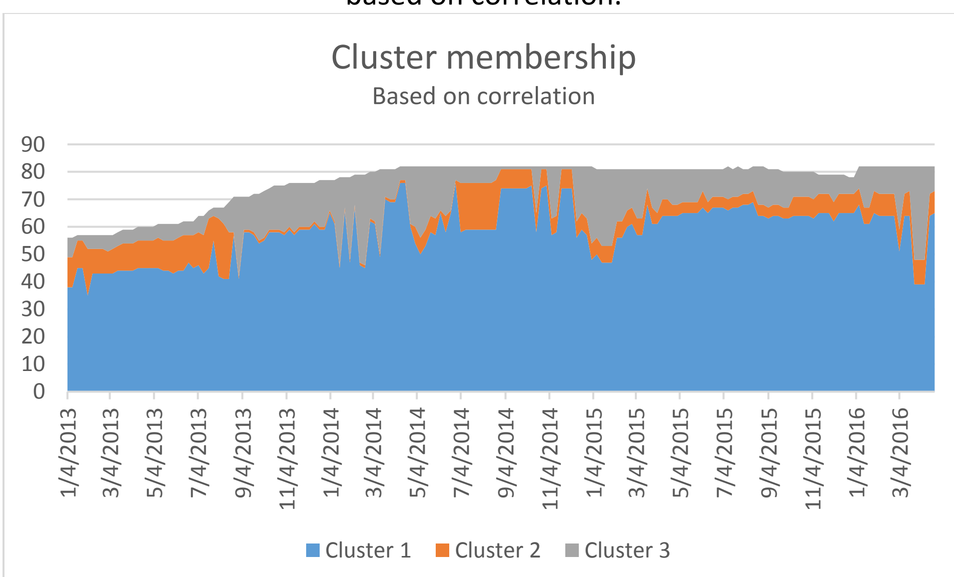


Chart 8. Average percent allocation of funds into clusters calculated over the entire period, based on correlation.



6.7 Cluster stability

In chart 9, we present the stability of clusters across time for different parameterizations. Stability here is defined as what percentage of funds remain in the same cluster between consecutive periods. The first number in the legend indicates the window size, or the number of weeks used to calculate moving average. The second number indicates the number of weeks between two points in time for which we calculate stability. We present the same results for correlation in chart 10. When comparing the stability of the two approaches, it is important to take into account the allocation of funds across clusters presented in chart 7.

Given that correlation produces more concentrated clusters, it is expected that it will be more stable. Charts 11 and 12 display the same results for each period in time in order to show trends or periods when stability was higher or lower than the average.

Chart 9. Average percent of funds that belong in the same clusters calculated over the entire period based on similarity.

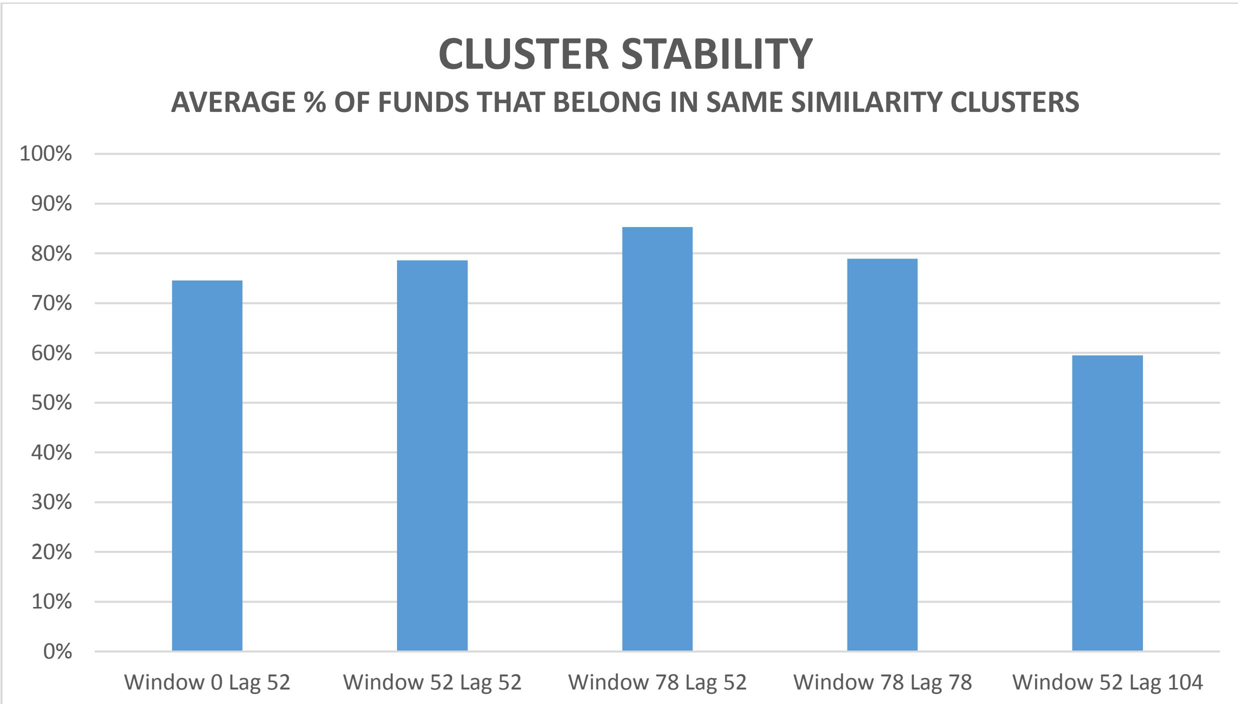


Chart 10. Average percent of funds that belong in the same clusters calculated over the entire period based on correlation.

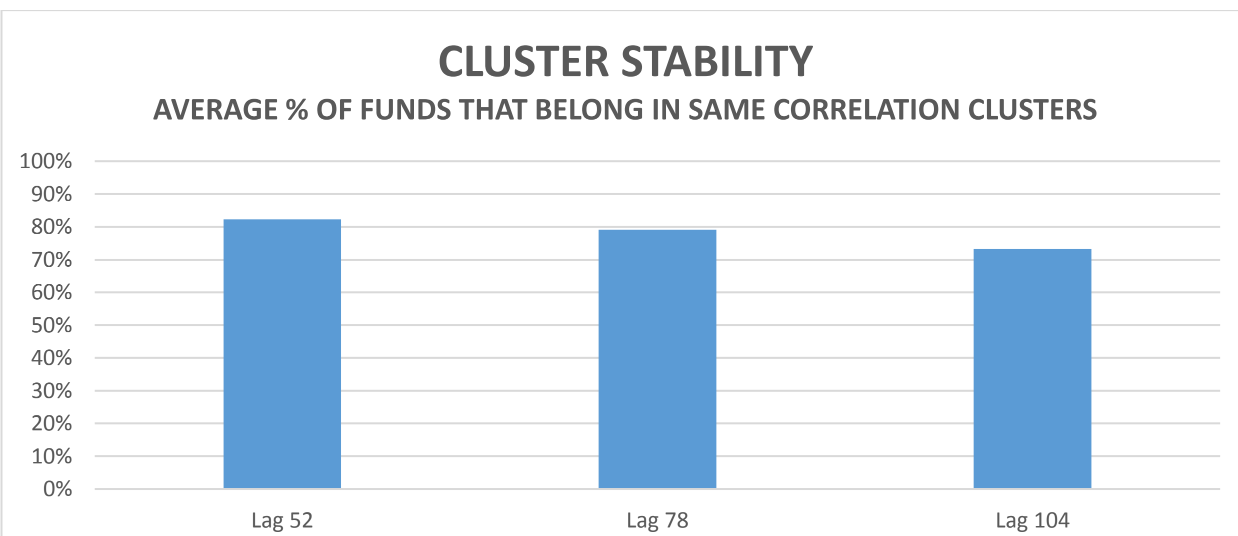


Chart 11. Percent of funds that belong in the same clusters for each period based on similarity.

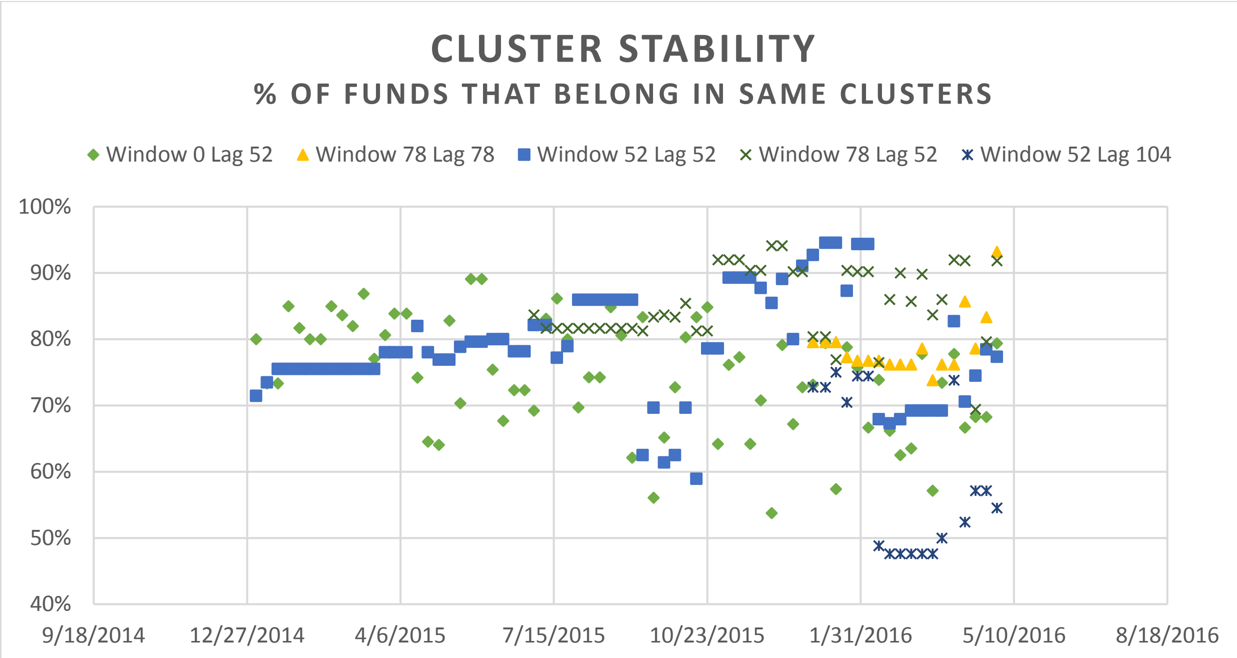
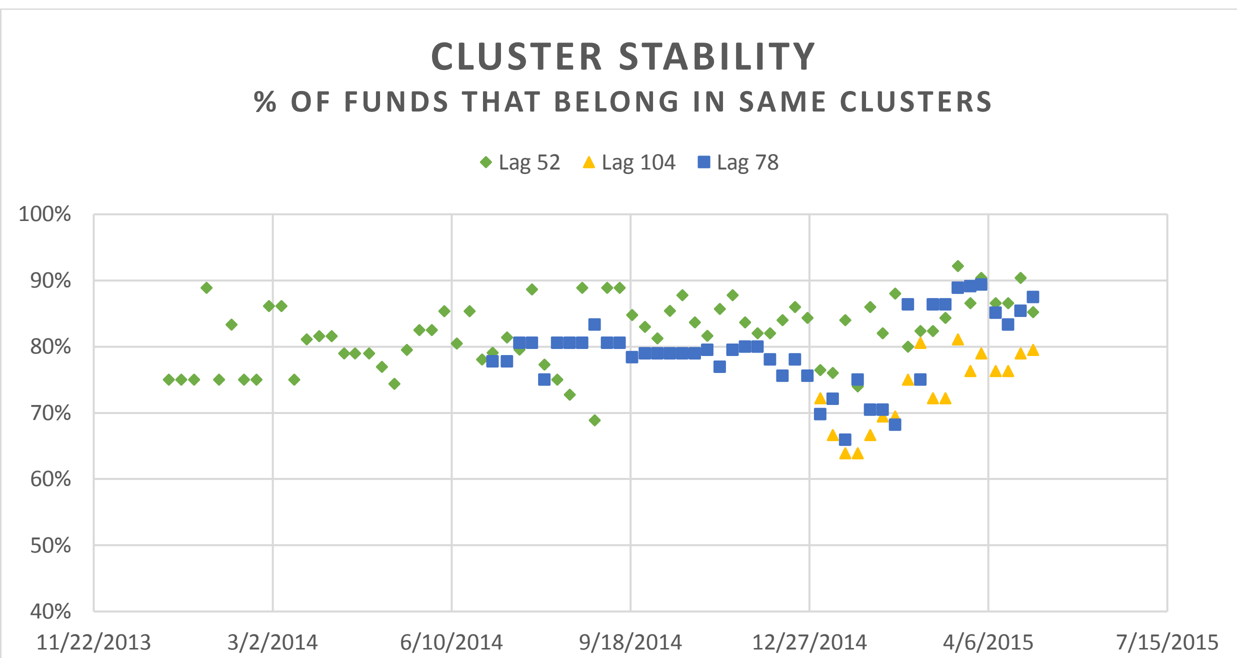


Chart 12. Percent of funds that belong in the same clusters for each period based on correlation.

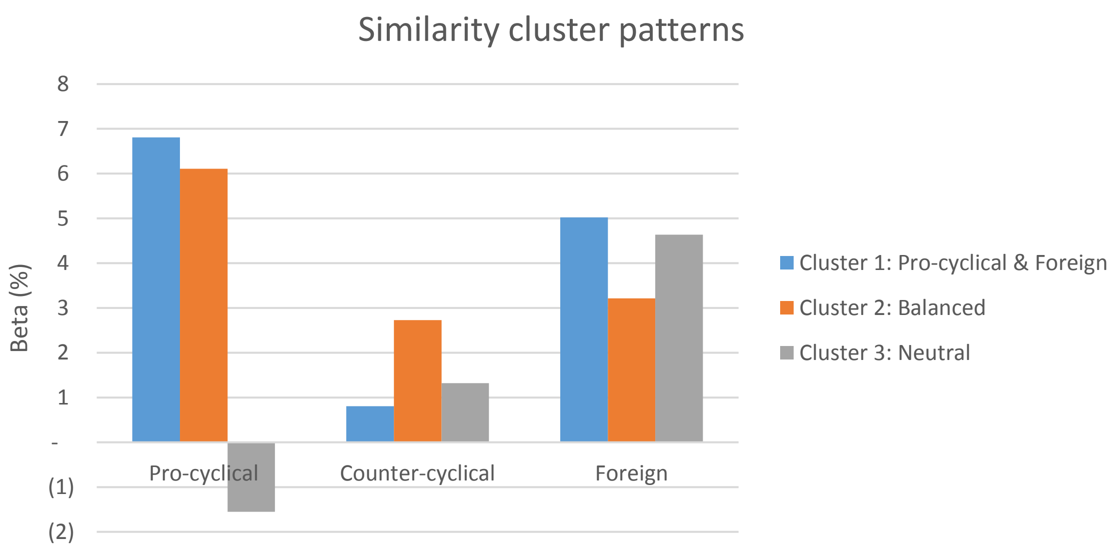


We observe that even the current clustering approach produces stable results with over 70% of the funds belonging in the same clusters a year later. Correlation-based clustering appears to be more stable than the current approach, but as previously mentioned, a big reason for this has to do with its highly concentrated top cluster. Following Morningstar's three-year window approach to classify funds, we observe that increasing the window with moving average allows us to achieve very high stability levels in the case of using similarity. Unfortunately, we didn't have enough data in this rather recent universe to test longer horizons. Based on chart 9, however, we can detect an almost linear stability increase when we increase the number of moving average periods. Had we been able to use three years like Morningstar, we would expect to reach stability levels above 90% at the one year horizon. These levels of stability are comparable to the ones found among holdings-based approaches.

6.8 Cluster themes

The hardest part in a clustering procedure is to try to make sense of the resulting themes that appear. Although not a direct objective, it is possible to extract informative conclusions based on how the data dictate the different clusters. To give some insight on the clustered funds, we display in chart 13 the average factor exposures in each cluster as of 11/20/2015, based on our current similarity measure. There are some subtle and some more obvious differences among the clusters, which lead us to suggest the cluster names shown in the legend.

Chart 13. Cluster attributes and names.



6.9 Performance

Although not a direct objective, it is generally desirable that clusters among funds will result in different performance behaviors. In the next chart, we calculate three out of sample portfolios, where each portfolio invests in each cluster, based on current clusters (no moving average) and equal rebalancing among the funds of each cluster on a weekly basis, based on each week's cluster members. Although it's not practical to rebalance each week across multiple funds, this exercise can provide valuable insight with respect to the performance diversity of the three clusters over time.

We observe that the correlation approach provides only limited diversity, whereas the similarity approach provides more distinguished clusters. Perhaps other mutual or hedge fund categories will provide more performance differentiation, which will be the subject of future research.

Chart 14. Out of sample portfolios that invest in each of three similarity clusters.

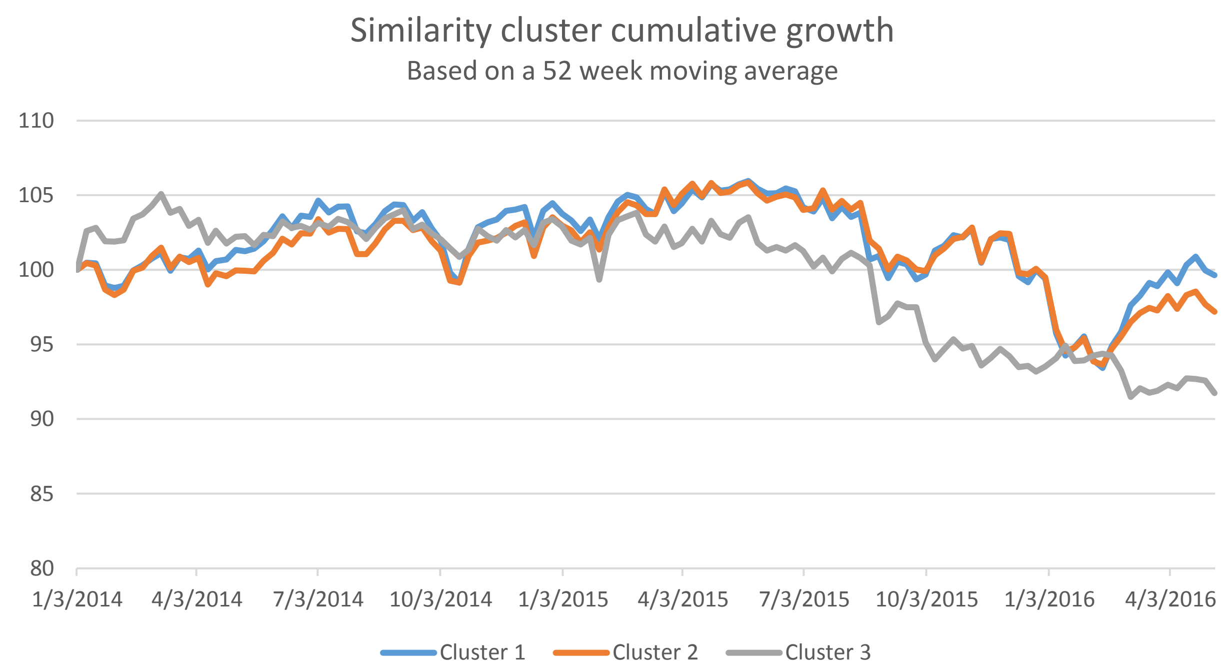
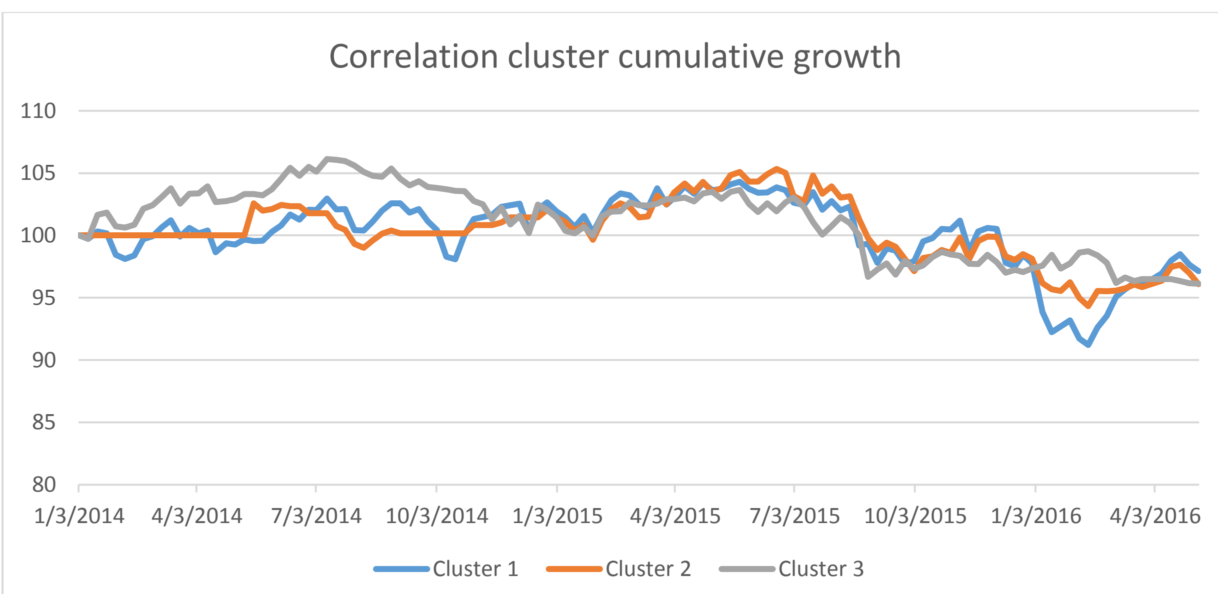


Chart 15. Out of sample portfolios that invest in each of three correlation clusters.



7. Robustness

The above experiments provide a sense of cluster stability in relation to time varying betas. We have kept the similarities constant as we do not anticipate investors to change their mind about those often. Even if this is the case, however, there may be different perceptions among individuals of the same firm of how similar factors are, or uncertainty for each individual around their values. To gain a better understanding on how similarity uncertainty affects cluster stability, we generated 1000 scenarios and formed clusters for each. In each scenario, the shocked similarities were produced by drawing from a multivariate normal distribution, with mean equal to the original similarities used and standard deviation equal to 10%. We also repeated the experiments with a higher standard deviation equal to 30%. Based on the shocked similarities and the betas of a given date[10], we calculated fund similarities and then clusters among the funds. The number of clusters was kept to three, as per the previous experiments.

Similar to section 6.7, to calculate cluster stability, we measured the average percent of funds that did not display any cluster deviation across all simulations, i.e. funds that are always assigned to the same cluster. For a 10% similarity standard deviation, we found that 69% of the funds did not change clusters, whereas for a 30% deviation, this number went down to 63%. This measure, however, doesn't tell us if funds that changed clusters were continuously changing clusters or only changed clusters once. Furthermore, this measure depends on the number of simulations; the higher that number, the higher the chances that any given fund will change clusters. To gain further insight, we calculated the average percent of times across all funds that a fund changes clusters. In the 10% deviation case, this number was found to be 0.2%, whereas in the 30% deviation case, it was found to be 4%. So, even though several funds did change clusters, this was a rather rare phenomenon, indicating high levels of cluster stability in relation to similarity uncertainty.

8. Conclusion

We presented a framework that allows one to cluster funds that have limited or no transparency. It can also be used in place of holdings-based approaches that may be difficult to make sense of or produce data for. Our framework is based on a novel similarity measure that takes into account how similar are the factors that different funds invest in, how similar are the ways – represented by betas – that they invest in those factors and how idiosyncratic the funds are. Our measure is able to overcome common pitfalls encountered in correlation or feature-based clustering, making it more applicable to the investment professional. We showed a case study of this framework, producing clusters against the equity long/short mutual fund category of Morningstar and found our clusters to be more diversified and stable than ones produced based on correlation. The stability observed was at par with stability based on holdings-based

[10] We chose 11/20/2015 to be that date. There is no particular reason for it, and we could have worked with betas as of any other date.

approaches[11]. Clusters produced with our approach were also found to be robust against alternative specifications of factor similarity values. At this stage, we have found little differentiation from a performance standpoint. It will be interesting to apply this framework to other categories and investigate the derived clusters and performance characteristics.

Appendix

A. Alternative similarity population

The similarity among factors defined in this paper does not need to be based on a direct comparison of factors against each other. Hereby we show an alternative and more generic framework, where we look at each factor's attributes and populate similarities based on those attributes. Such a framework facilitates the population of similarities in the case one has to consider multiple factors across multiple funds.

The first attribute we look at is asset class. Given all those factors belong in the same asset class, their similarities equal 1.

Table 14. Similarities by asset class

	SP500 Financials	SP500 Inf Technology	SP500 Consumer Discret	SP500 Industrials	SP500 Utilities	SP500 Health Care	SP500 Consumer Staples	SP500 Telecomm Svcs	SP500 Energy	SP500 Materials	MSCI AC Asia Pacific USD	MSCI AC Europe USD
SP500 Financials	1	1	1	1	1	1	1	1	1	1	1	1
SP500 Inf Technology		1	1	1	1	1	1	1	1	1	1	1
SP500 Consumer Discret			1	1	1	1	1	1	1	1	1	1
SP500 Industrials				1	1	1	1	1	1	1	1	1
SP500 Utilities					1	1	1	1	1	1	1	1
SP500 Health Care						1	1	1	1	1	1	1
SP500 Consumer Staples							1	1	1	1	1	1
SP500 Telecomm Svcs								1	1	1	1	1
SP500 Energy									1	1	1	1
SP500 Materials										1	1	1
MSCI AC Asia Pacific USD											1	1
MSCI AC Europe USD												1

The next attribute we look at is the sector category that each index represents in terms of cyclical or counter-cyclical characteristics. Below we show such sector categories and associated similarities.

Table 15: Categories and similarities by sector category

[11] Keeping in mind that we did not examine the same universe like the one we are comparing against. The long/short equity fund universe is more obscure than growth or value categories we compared against, providing even more merit to our approach.

	Pro-Cyclical				Counter-cyclical				Energy & Materials		Foreign	
	SP500 Financials	SP500 Inf Technology	SP500 Consumer Discret	SP500 Industrials	SP500 Utilities	SP500 Health Care	SP500 Consumer Staples	SP500 Telecomm Svcs	SP500 Energy	SP500 Materials	MSCI AC Asia Pacific USD	MSCI AC Europe USD
SP500 Financials	1	0.7	0.7	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.1	0.4
SP500 Inf Technology		1	0.7	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.1	0.4
SP500 Consumer Discret			1	0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.1	0.4
SP500 Industrials				1	0.3	0.3	0.3	0.3	0.3	0.3	0.1	0.4
SP500 Utilities					1	0.7	0.7	0.7	0.3	0.3	0.1	0.4
SP500 Health Care						1	0.7	0.7	0.3	0.3	0.1	0.4
SP500 Consumer Staples							1	0.7	0.3	0.3	0.1	0.4
SP500 Telecomm Svcs								1	0.3	0.3	0.1	0.4
SP500 Energy									1	0.7	0.1	0.4
SP500 Materials										1	0.1	0.4
MSCI AC Asia Pacific USD											1	0.1
MSCI AC Europe USD												1

We do not attempt to go more granular in terms of how similar different sectors are to each other. Next, we quote similarities by region. Below we show the similarity based on the region each index represents, noting that MSCI Europe has constituents from both developed and emerging market countries.

Table 16: Categories and similarities by region

	Developed										Emerging Markets	Global
	SP500 Fina	SP500 Inf T	SP500 Con	SP500 Indu	SP500 Util	SP500 Hea	SP500 Con	SP500 Tele	SP500 Ene	SP500 Mat	MSCI AC Asia Pacific USD	MSCI AC Europe USD
SP500 Financials	1	1	1	1	1	1	1	1	1	1	0.1	0.5
SP500 Inf Technology		1	1	1	1	1	1	1	1	1	0.1	0.5
SP500 Consumer Discret			1	1	1	1	1	1	1	1	0.1	0.5
SP500 Industrials				1	1	1	1	1	1	1	0.1	0.5
SP500 Utilities					1	1	1	1	1	1	0.1	0.5
SP500 Health Care						1	1	1	1	1	0.1	0.5
SP500 Consumer Staples							1	1	1	1	0.1	0.5
SP500 Telecomm Svcs								1	1	1	0.1	0.5
SP500 Energy									1	1	0.1	0.5
SP500 Materials										1	0.1	0.5
MSCI AC Asia Pacific USD											1	0.5
MSCI AC Europe USD												1

Finally, we take an extra step of defining similarities by region. In table 13, even though on an aggregate level indices from the same region are considered equivalent, we have to consider how likely it is for two funds with exposure to the same index to have the same behavior. Given the US sector indices are more concentrated than the foreign indices, we deem it more likely that similar factor exposures by two funds to a US sector index signify similar behavior more so than similar factor exposures to a foreign index. The similarities based on this are displayed in table 14.

Table 17: Categories and similarities within region

	SP500 Financials	SP500 Inf Technology	SP500 Consumer Discret	SP500 Industrials	SP500 Utilities	SP500 Health Care	SP500 Consumer Staples	SP500 Telecomm Svcs	SP500 Energy	SP500 Materials	MSCI AC Asia Pacific USD	MSCI AC Europe USD
SP500 Financials	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.1	0.5
SP500 Inf Technology		0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.1	0.5
SP500 Consumer Discret			0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.1	0.5
SP500 Industrials				0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.1	0.5
SP500 Utilities					0.8	0.8	0.8	0.8	0.8	0.8	0.1	0.5
SP500 Health Care						0.8	0.8	0.8	0.8	0.8	0.1	0.5
SP500 Consumer Staples							0.8	0.8	0.8	0.8	0.1	0.5
SP500 Telecomm Svcs								0.8	0.8	0.8	0.1	0.5
SP500 Energy									0.8	0.8	0.1	0.5
SP500 Materials										0.8	0.1	0.5
MSCI AC Asia Pacific USD											0.8	0.5
MSCI AC Europe USD												0.8

The steps that lead to the calculation of systematic similarity are the same as the ones described in the methodology section. For the systematic similarity calculation, we apply an extra loop in order to take into account the similarity calculation due to the four different similarity scores in 6.3. Formula 6 now becomes:

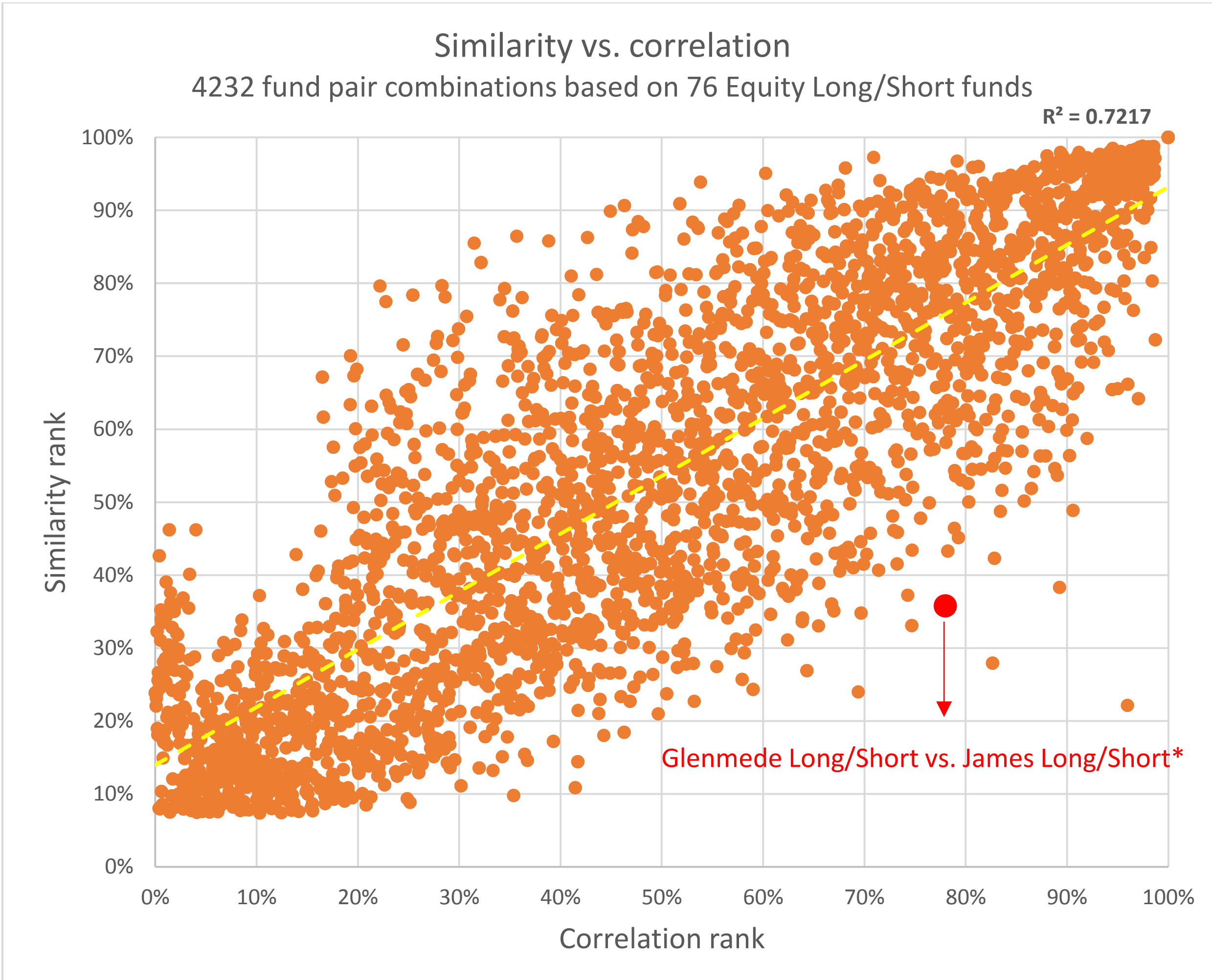
$$S_{SYST}^{A,B\,*} = \sum_k w_k S_{SYST}^{A,B}\left(\sum_i b_i^A x_i^A, \sum_j b_i^B x_i^B\right) = \sum_k w_k \sum_i \sum_j S_B\left(b_i^A, b_j^B\right) * S_X\left(x_i^A, x_j^B\right) \qquad (6)$$

where k ranges from 1 to 4 and w_k is a weight associated with each similarity score in 6.3. Such a weight may equally weight each factor similarity score, or the user may want to place more emphasis on a particular attribute. Formulas 7-8 are then using $S_{SYST}^{A,B\,*}$ instead of $S_{SYST}^{A,B}$.

B. Correlation pitfalls

The below is a scatter chart of our calculated similarity measure plotted against correlation in terms of ranks, in the case of Equity long/short funds. We observe that, for the most part, the two measures relate to each other. There are many cases, however, where the two deviate, one of them being discussed in section 2 of the paper.

Chart 16. Similarity vs. correlation rank for the Equity long/short category



C. Dynamic style analysis (DSA)

The model solves for the betas at each point in time via the below dynamic quadratic programming:

$$\min_{\boldsymbol{\beta}} \sum_{t=1}^{T}(r_t - \mathbf{x}_t'\boldsymbol{\beta}_t)^2 + \lambda\sum_{t=2}^{T}(\boldsymbol{\beta}_t - \mathbf{B}_t\boldsymbol{\beta}_{t-1})'\mathbf{D}_t(\boldsymbol{\beta}_t - \mathbf{B}_t\boldsymbol{\beta}_{t-1}) \qquad (2)$$

$$\lambda > 0$$

$$\sum_{i=1}^{n}\beta_t^i = 1$$

$$\text{n} = \text{number of factors}$$

$$\text{B}_\text{t} = Turnover\_matrix$$

$$\text{D}_\text{t} = \text{scaling matrix}$$

$$\text{t} = 1,\ldots,T$$

The DSA model above is consisted of two terms. The first term tries to solve for the beta at each point in time by minimizing the sum of squared residuals, similar to an OLS approach. This, however, points to an infinite number of solutions. To limit this number and arrive at an optimal solution, the model imposes the second term as a penalty, which calculates how much turnover there is between successive points in time. λ is a tolerance parameter. The larger the value of the parameter λ, the more weight is attached to the penalty term that penalizes high turnover, and the smoother the betas become. As λ increases, the betas in (2) become very smooth and resemble the betas that would be obtained from a least squares solution over the entire range of observations. At $\lambda \to \infty$, the betas are the same as the ones obtained with the Sharpe model, in the presence of long-only constraints, or OLS, in the unconstrained case. As λ decreases, the betas become very volatile. The value of λ is calibrated based on a cross validation statistic called predicted R-squared[12]. B is a matrix that controls if what gets smoothed between consecutive time steps is the quantity or weight of each asset. D is a matrix that contains individual scaling parameters. Its purpose is to control the smoothness of individual betas as well as provide the appropriate scaling for the combined objective[13]. This is the model that we use for the experiments of this paper. It is a similar model to flexible least squares (FLS) [Kalaba, Tesfatsion 1989,1990] and [Lutkepohl 1996] with a couple of key differences. FLS doesn't incorporate beta constraints and does not solve for the value of λ.

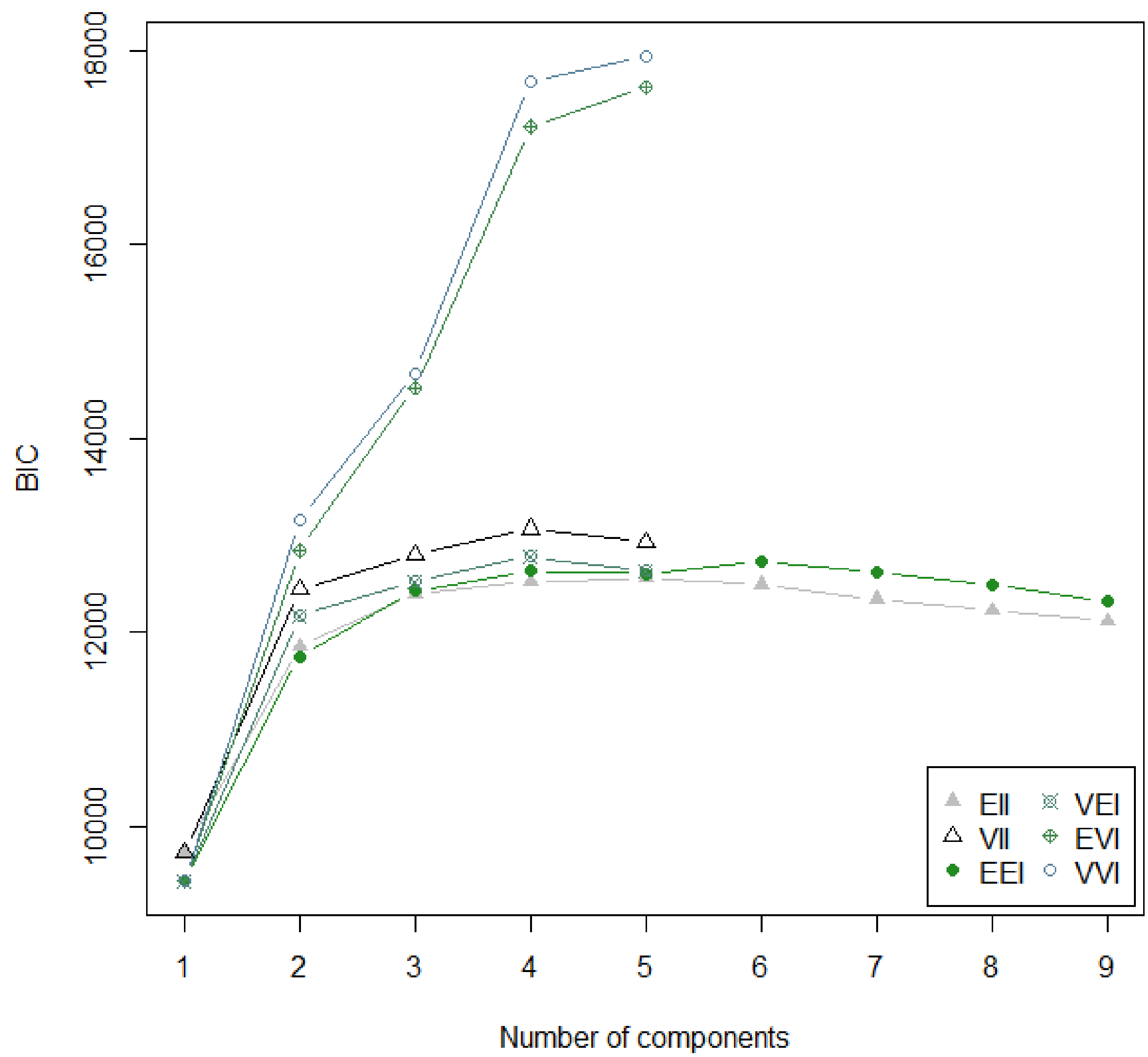
D. Optimal number of clusters

We first display the outcome of a test for the optimal number of clusters performed on just one of the analysis periods involved. Given the arbitrary nature of the optimal number of clusters, we feel that such at test does not need to take place very frequently.

[12] To calculate predicted R-squared, we re-estimate the model for a given fund while taking out one of the 12 annual return observations and then assessing the difference between the removed annual return observation and the estimated value for that observation made by the remaining observations. In this case, we re-estimate 12 different times taking out each observation in turn. We combine all 12 of the differences between estimated and actual returns to calculate the predicted R-squared.

[13] For more details see Markov et Al.

Chart 4. Optimal number of clusters, showing BIC for 6 model parameterizations and up to 9 clusters.



We follow a model-based clustering process based on Dean and Raftery (2006). The similarities x are viewed as coming from a mixture density

$$f(x) = \sum_{k=1}^{G} \tau_k f_k(x)$$

where f_k is the probability density function of the observations in cluster k, and τ_k is the probability that an observation comes from the kth mixture component - $\tau_k \in (0,1)$ and $\sum_{k=1}^{G} \tau_k = 1$. Each component is then modeled by a Gaussian distribution, with points near the mean of clusters having higher density than points away from it. The multivariate covariance matrix among the clusters can be parametrized in different ways, some of which are displayed in chart 5. The higher the BIC value, the better each parametrization will fit the data. Generally, we observe that some of the models indicate 4 as the optimal number of clusters, where the top two models indicate 5.

Literature

Alimi, M. Zandieh, and M. Amiri: “Multi-objective portfolio optimization of mutual funds under downside risk measure using fuzzy theory,” International Journal of Industrial Engineering Computations, 3(5):859–872 (2012)

Bauman, E.V., Structuring Nominal Attributes in the Problem of Expert Examination, in *Ekspertnye otsenki v zadachakh upravleniya* (Expert Estimates in Control), Moscow: Inst. Probl. Upravlen., 1982.

Brown, Stephen J., and William N. Goetzmann. "Mutual fund styles." *Journal of financial Economics* 43, no. 3 (1997): 373-399.

Chakrabarti, Deepayan, Ravi Kumar, and Andrew Tomkins. "Evolutionary clustering." *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006.

Cremers, KJ Martijn, and Antti Petajisto. "How active is your fund manager? A new measure that predicts performance." *Review of Financial Studies* 22, no. 9 (2009): 3329-3365.

David Moreno, Paulina Marco, and Ignacio Olmeda: “Self-organizing maps could improve the classification of Spanish mutual funds,” European Journal of Operational Research 174(2):1039–1054 (2006)

Dean, Nema, and Adrian E. Raftery. "Variable Selection for Model-Based Clustering." *Journal of the American Statistical Association* 101.473 (2006): 168-178.

DiBartolomeo, Dan, and Erik Witkowski. "Mutual fund misclassification: Evidence based on style analysis." *Financial Analysts Journal* 53, no. 5 (1997): 32-43.

Kim, Moon, Ravi Shukla, and Michael Tomas. "Mutual fund objective misclassification." *Journal of Economics and Business* 52, no. 4 (2000): 309-323.

Markov, Michael, Vadim Mottl, and Ilya Muchnik. "Principles of nonstationary regression estimation: A new approach to dynamic multi-factor models in finance." *DIMACS, Rutgers Univ., Piscataway, NJ, Tech. Rep* 47 (2004).

Morningstar. *The Morningstar category classifications*. Working paper, 2016.

Ramin Baghai-Wadji, Rami El-Berry, Stefan Klocker, and Markus Schwaiger: “The Consistency of Self-Declared Hedge Fund Styles? A Return-Based Analysis with Self-Organizing Maps,” Financial Stability Report 9:64–76, Oesterreichische Nationalbank (2005)

Sakakibara, Takumasa, et al. "Clustering Mutual Funds Based on Investment Similarity." *Procedia Computer Science* 60 (2015): 881-890.

Sharpe, William F. "Asset allocation: Management style and performance measurement." *The Journal of Portfolio Management* 18.2 (1992): 7-19.