University of California
Berkeley
Haas School of Business

Clustering as a Basis of Hedge Fund Manager Selection

March, 2005

Applied Finance Project
Haas MFE Program

Prepared by:

Kan Liang

Acknowledgements

The author thanks Francis Gupta from Dow Jones Indexes for supporting and advising this project with his insights, efforts and time, and for allowing me to select and evolve the topic for the research. Scott Lang provided the data of Hedge Funds and Hamish Seegopaul helped the author improve the way to present the results. The author appreciates John Prestbo and Ridgely Walters from Dow Jones Indexes, and Vassilis Karavas from Lyra Capital with their helpful comments and suggestions. The author also appreciates his faculty advisor, Jonathan Berk, who reviewed his work.

About the Author

Kan Liang is a recent graduate from the Masters Program in Financial Engineering at Haas Business School, UC Berkeley. This research was the subject of Kan Liang's Applied Finance Project and was conducted while he was interning at Dow Jones Indexes from October 2004 through January 2005. Any questions concerning the paper can be directed to Kan Liang at **kliang@haas.berkeley.edu** .

Abstract

This paper focuses on component selection from a universe of hedge fund managers. It uses two clustering methods: K-means and K-medoids to categorize a universe of hedge fund managers on the basis of hedge fund style/strategy. For each one of the methods, this paper presents the test for the robustness of the resulting classes by alternating between using the standard deviations and mean absolute deviations to standardize the data, and Euclidean and Manhattan metric as the measure of distance between any two funds. Two sets of results are presented with respect to the trade-offs that exist between two types of error: Type I error, which leads to some managers that are representative of the strategy being excluded from the possible set of candidates for inclusion, and Type II error, which leads to some managers that are not representative of the strategy being considered for possible inclusion.

I Introduction

Component selection is the key to portfolio construction. A portfolio can be comprised of securities (as is the case for the purposes of investing or the construction of a performance benchmark/index) or it can be made up of a group of portfolio managers (as is the case for purposes of investing in a fund-of-funds or again for the purposes of constructing a performance benchmark/index). But in all cases, the process used to select the components can make a significant difference in the characteristics of the resulting portfolio.

This paper focuses on component selection from a universe of hedge fund managers. Since hedge fund managers can use a variety of hedge fund strategies to generate performance, the goal of this analysis is to use clustering - a statistical technique - to assist in identifying groups of managers that are best representative of their style. Understanding a hedge fund manager's style is essential to ensure that the managers that are selected as components to make up the portfolio deliver the exposure to the portfolio that they are meant to deliver. For example, when constructing a portfolio of long-only mutual fund managers it is very common to perform return-based style analysis to understand the exposures (or factor loadings) of the managers. Having verified a mutual fund manager's style it can then be considered as a potential candidate for a portfolio seeking to gain exposure to that specific style.

In general, the selection of hedge fund managers for inclusion within a fund-of-funds or a portfolio is based on manager characteristics (e.g., tenure and assets under management), quantitative methods (e.g., performance and statistical or factor based style analysis) and qualitative due-diligence (such as organization structure and management objectives/philosophy).

Manager characteristics are meant to serve the purpose of an initial screen and meant to identify managers who would be considered for further quantitative and qualitative analysis.

Quantitative methods refer to methods grounded in statistical analysis, i.e., trying to group or categorize the managers by strategy or style by using a measure of similarity or dissimilarity among their return streams. These include statistics such manager returns, volatilities, correlations and as is the case in this paper distance metrics. Since quantitative methods require hard data to compute the statistics, hedge fund manager groupings based on them are only applicable to the time period the analysis covers and, therefore, are backward looking. But still, the statistics can provide empirical justification for the current strategies/styles employed by a manager.

Qualitative due-diligence on the other hand are more focused on understanding the organization of the money manager, the risk management systems in place and the objective and philosophy of the investment product. Due-diligence assists in gauging the risk factors and the extent of that risk that may be associated with including a manger within a portfolio. This also includes checking the consistency between historical and current manager performance and the performance-generating philosophy. The composition of a manager's hedge fund portfolio together with the manager's buy and sell decisions are used to understand the manager's value proposition in the past and looking forward.

As mentioned, this article uses quantitative methods to categorize a universe of hedge fund managers on the basis of hedge fund style/strategy. Quantitative methods vary significantly with respect to the statistics and criteria used to categorize managers by style, and depending on the statistics/criteria employed can lead to significant differences in the grouping of components. In this paper we focus on using clustering algorithms to group hedge fund managers. We use two clustering methods: K-means and K-medoids and for each one of the methods, we test for the robustness of the resulting classes by alternating between using the standard deviations and mean absolute deviations to standardize the data, and Euclidean and Manhattan metric as the measure of distance between any two funds.

It must also be mentioned that in this paper clustering analysis is implemented to check/verify the style/strategy of managers that fall within the following four strategies:

1. Convertible Arbitrage
2. Distressed Securities
3. Event Driven
4. Merger Arbitrage.

These are four of the six hedge fund strategies that Dow Jones Hedge Fund Indexes computes and publishes benchmarks for, and uses clustering as one of the methods to quantitatively understand and verify hedge fund manager style[1]. Having performed the analysis, we present two sets of results. The first set of results focuses only on the cluster that is most representative of the strategy under consideration and ignores all others. This set of results may lead to some managers that are representative of the strategy being excluded from the possible set of candidates for inclusion. Later on in the article we refer to this as Type I error. Conversely, the second set of results drops only the cluster that is least representative of the strategy. This set of results may lead to some managers that are *not* representative of the strategy being considered for possible inclusion. This will be referred to as Type II error. The challenge while implementing any component selection methodology is to strike a balance between these two types of errors.

Finally, it must also be mentioned that in most applications, the choice of the criteria used to determine the "best" cluster or "rank" the clusters in terms of more or less representative of the strategy is fairly subjective. For purposes of simplicity, in this paper we use the correlations of the cluster members against a broad strategy benchmark (for example the CSFB/Tremont) to determine which cluster is best representative of the strategy. If this criterion were to change, the relative rankings of clusters in terms of more or less representative of the strategy may change. If this were to happen, it means that the components for inclusion and exclusion, respectively, would also change.

This article is structured as follow: a brief survey of literature on clustering follows this introduction. The next section presents the data together with some descriptive statistics for the sample of hedge fund manager covering the four strategies. Then we present the methodology used to perform K-means and K-medoids clustering for a sample of managers. Finally, we discuss the results and conclusions. The Appendix contains all of the statistics computed for the resulting clusters.

II Survey of the literature

The dispersion in hedge fund returns indicates hedge fund performance analysis requires both qualitative analysis and quantitative analysis. In quantitative method, there are two approaches, asset based style analysis and return based style analysis. William F. Sharpe (1992) pioneered his work in the return based investment style analysis using a regression analysis approach.

---

[1] For more information on the Dow Jones Hedge Fund Strategy Benchmarks please visit www.djhedgefundindexes.com.

Fung and Hsieh (1997) extended Sharpe's regression analysis by adding a repressor to proxy the returns of Hedge Funds. To model a manger's returns, they incorporated three factors that are considered determinants of return: the returns from assets in the managers' portfolios, the trading strategies used, and the use of leverage. They then applied factor analysis to extract principal components which corresponded to the most dominant common hedge fund styles to which the manager's returns would be correlated.

Francois-Serge Lhabitant (2001) postulated a multi-factor model to explain hedge fund historical returns. He applied principal component analysis for optimal fitting. Multiple factor models are commonly used to characterize how industry factors and economy wide pervasive factors affect a hedge fund manager's returns.

Clifford De Souza and Suleyman Gokcan (2004) propose the use of the Hurst exponent in conjunction with the D-statistic for manager selection. They utilize a logit model as a formal methodology to isolate the characteristics of a database of liquidated funds and find agreement with an earlier model proposed for liquidation.

Two types of clustering methods have been used to analyze hedge fund manager styles: hierarchical clustering and nonhierarchical clustering. Johnson and Wichern (1998) discussed several hierarchical clustering methods, like the linkage methods and Ward's minimum variance method. Anderberg (1982) described various nonhierarchical clustering methods. One of the widely used nonhierarchical clustering methods is the K-means clustering method, which was originally developed by MacQueen in 1967. Nandita Das (2003) applied the K-means clustering method to classify hedge funds managers based on asset class, size of hedge fund, incentive fee, risk level, and liquidity of the hedge funds.

In cluster analysis, the measure of similarity or dissimilarity is very critical to the composition of the resulting clusters. Romesburg (1988) discussed in detail that various resemblance coefficients that can be calculated for quantitative, qualitative, and mixed attributes.

III Data

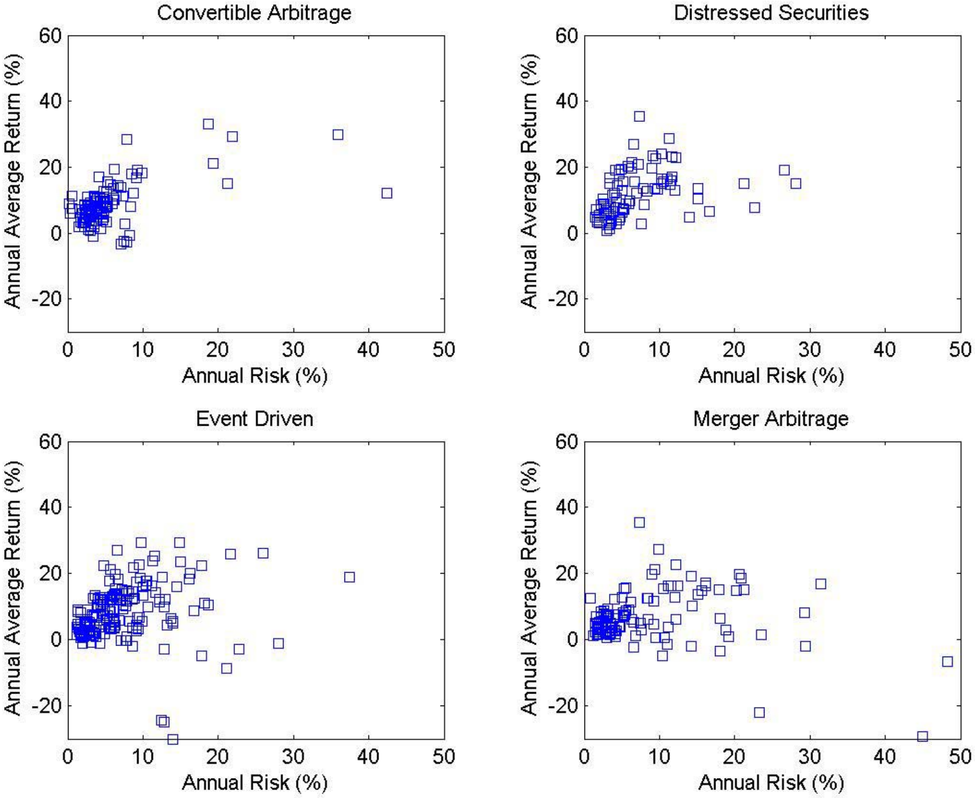
Figure 1. Data Providers Used to Construct the Dow Jones Hedge Fund Manager Universe

Database Provider
TASS
HFR
AltVest (Investor Force)
Hedgefund.net (Tuna)
CISDM (CISDMHedge, CISDMCTA)
Cogent

Monthly return data from July 31, 2001 to June 30, 2004 for Convertible Arbitrage, Distressed Securities, Event Driven and Merge Arbitrage were obtained from Dow Jones Hedge Fund Indexes ("DJHFI"). As of June 30, 2004, the DJHFI hedge fund universe was made up of 7135 unique hedge fund managers from six publicly available hedge fund manger databases listed in Figure 1. This unique universe did not include any duplicates, dead funds, or funds that had not reported since December, 2003.

The analysis presented in this paper covers the following four distinct hedge fund strategies: Convertible Arbitrage, Merger Arbitrage, Distressed Securities and Event Driven. From the set of all hedge fund strategies these strategies were selected for two reasons: (i) Each one of these strategies has a set of factors that can explain its performance, and (ii) a significant proportion of these strategies returns can be explained by those underlying factors. In contrast, the performance of other hedge fund strategies is more dependent on the manger's ability to market time (directional) and other trading-specific manager skills.

Figure 1: Risk vs Return Scatter Plots



Sources: Manager data from TASS, HFR, Alvest, Hedgefund.net, CISDM and Cogent.

Figure 1: presents the risk -return scatter plots for the managers within each strategy. Figure 2 presents the number of managers within each strategy and mean, median, maximum and minimum for the annualized average monthly returns over the 36 monthly returns that are used for the analysis. For the 36 month period ending June 2004, Convertible Arbitrage is represented by 112 funds, Distressed Securities by 88 funds, Event Driven by 181 funds and Merger Arbitrage is represented by 116 funds. The mean of the annualized average monthly return across the funds over this time period of 11.96% is the largest for the Distressed Securities. The average over funds for the other three strategies was 7.32% for Merger Arbitrage, 8.44% for Event Driven and 8.51% for Convertible Arbitrage.

The range of average returns (maximum – minimum) is the greatest for Event Driven funds. Where as the maximum annualized return amongst the funds was 61.59% (this was also the highest maximum amongst the four strategies), the minimum annualized return was -30.08% (this was also the lowest minimum amongst the four strategies) – implying a range of 61.59% - (-30.08%) = 91.61%. With a range of 58.33% - (-29.13%) = 87.47%, Merger Arbitrage funds were also not far behind the Event Driven funds. Convertible Arbitrage funds had a range of 36.36%, while Distressed Securities funds had a smallest range of 34.62%. For all of the strategies, the median average return lay below the mean average return, implying that the average return distribution had a few managers with exceptionally high returns that were driving the means above the medians.

In terms of the standard deviations of the funds, Convertible Arbitrage funds exhibited the smallest mean standard deviation, followed by Distressed Securities funds, which in turn was followed by Event Driven funds and finally by Merger Arbitrage funds. The range of standard deviation outcomes of 22.99% is the smallest for Distressed Securities, followed by 38.40% for Convertible Arbitrage funds, followed by 42.13% for Event Driven and 45.38% for Merger Arbitrage funds.

Figure 2: Data Statistics

Strategy	**Total Funds**	**Mean**	**Max**	**Median**	**Min**
		Annualized Average Manager Return (%)			
Convertible Arbitrage	112	8.51	33.12	7.54	-3.24
Distressed Securities	88	11.96	35.36	10.43	0.74
Event Driven	181	8.44	61.59	7.11	-30.08
Merger Arbitrage	116	7.32	58.33	6.28	-29.13
		Standard Deviation (%)			
Convertible Arbitrage	112	5.37	42.38	3.89	0.23
Distressed Securities	88	6.99	28.09	5.10	1.44
Event Driven	181	7.48	48.03	5.90	1.19
Merger Arbitrage	116	8.85	50.59	5.21	0.87

Sources: Manager data from TASS, HFR, Alvest, Hedgefund.net, CISDM and Cogent.

Figure 3A presents the distribution of annualized average monthly returns for the set of managers within each strategy. Consistent with the scatter plots presented in Figure 1, highlight the fact that most of the managers within each strategy had a significantly positive average monthly return over the three year time period under consideration. Within Distressed Securities, none of the managers had a negative average monthly return.

Figure 3B presents the distribution of managers by the number of months they had negative returns. With an average of 7.8864 months, managers within the Distressed Securities strategy had the fewest negative return months. The mangers with the most negative months were within the Merger Arbitrage strategy with an average of 10.9310 months. (Event Driven 9.8508 months, Convertible Arbitrage 8.6339 months)

Figure 3A. Distribution of Annualized Average Monthly Return
36 Months Ending June 2004

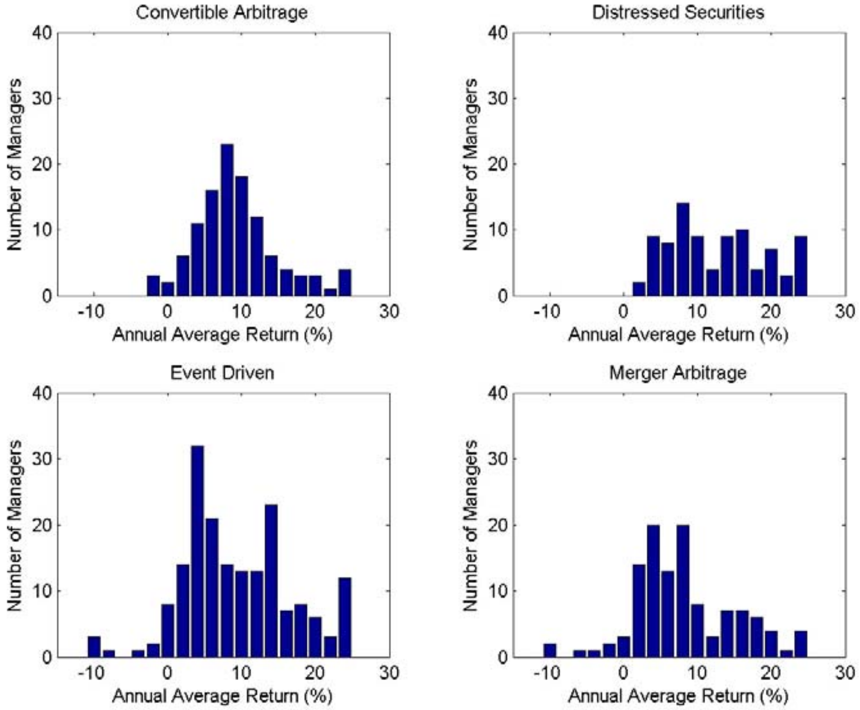
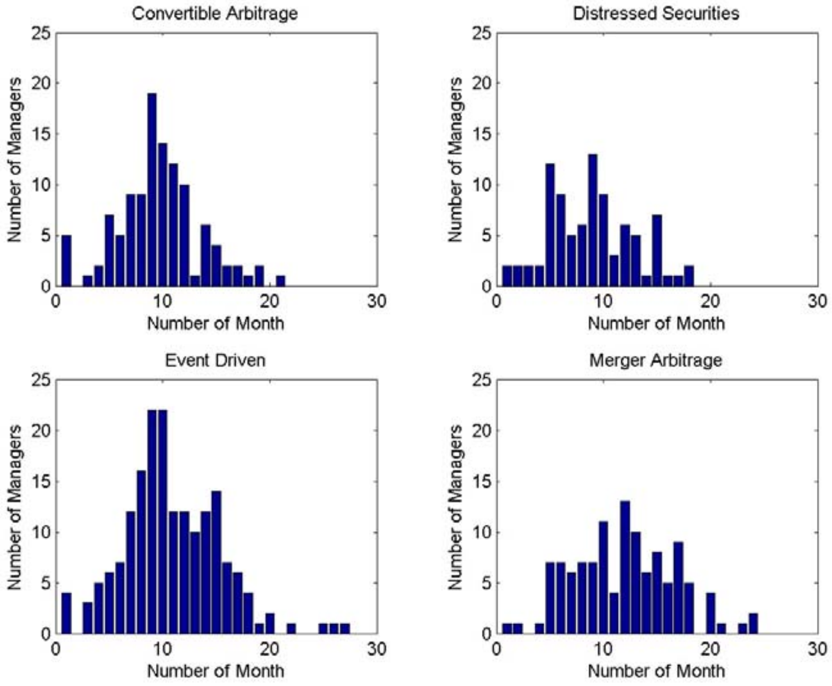


Figure 3B. Distribution of Managers by Number of Months with Negative Returns
Ending June 2004



Sources: Manager data from TASS, HFR, Alvest, Hedgefund.net, CISDM and Cogent.

IV Methodology

Cluster analysis is a statistical method of representative identification. Manager selection is concerned with the identification of the purest manager in discrete clusters with the best representation of the underlying manager style.

The first step is to specify models that would be adequately cluster the managers. For the purposes of this study the K-means and K-medics clustering algorithms are adopted. K-means and K-medoids algorithms aim at maximizing the distance between the centroids of clusters (the intra-distance) and minimizing the distance of each member with a cluster to its centroid (the inter-distance).

To test for the sensitivity of the cluster outcomes, the methodology employs two distance measures: Euclidean distance and Manhattan distance; and two methods in the denominator to standardize manager returns: standard deviation and mean absolute deviation. The following criteria are used to determine the efficacy of the four scenarios resulting from the combination of the distance metrics and standardization metrics:

- The correlations between cluster members' and cluster centroids to minimize the inter-cluster difference
- The correlations between cluster centroids to maximize the intra-cluster difference
- The correlations between cluster centroids and broad benchmarks to determine style purity of clusters

A. Distance Measure

To group the hedge fund managers, the distance metric between two monthly return points has to be chosen prior to the clustering. There are two simple distance measures, Euclidean distance and Manhattan distance.

- Euclidean Distance

$$D = \sqrt{\sum_{j=1}^{k}\sum_{i=1}^{n}\left\| r_i^{(j)} - r_c^{(j)} \right\|^2}$$

where $\left\| r_i^{(j)} - r_c^{(j)} \right\|^2$ is a chosen distance measure between a monthly return $r_i^{(j)}$ of a manager in the j-th cluster and the cluster's centre $r_c^{(j)}$.

- Manhattan Distance

$$D = \sum_{j=1}^{k}\sum_{i=1}^{n}\left\| r_i^{(j)} - r_c^{(j)} \right\|$$

where $\left\| r_i^{(j)} - r_c^{(j)} \right\|$ is a chosen distance measure between a monthly return $r_i^{(j)}$ of a manager in the j-th cluster and the cluster's centre $r_c^{(j)}$.

B. Standardization Procedure

To generate meaningful results, the returns of the relevant funds are standardized prior to the use of the cluster analysis to smooth and make it more homogenous. This is so that funds within the same style, but that may use different levels of leverage, are grouped together. The returns are demeaned first by subtracting the mean. Then the mean deviations are divided by the standard deviation (STD) or mean absolute deviation (MAD).

C. Clustering Technique – K-means and K-medoids

K-means and K-medoids clustering algorithm are applied to monthly return for each manager style to find the style purity. With a given measure of standardized distance between funds The algorithm is composed of the following steps:

1. Place K points into the monthly return space represented by the Hedge Funds that are being clustered. These points represent the initial group centroids.
2. Assign each fund to the group that has the closest centroid.
3. When all funds have been assigned, recalculate the positions of the K centroids.
4. Repeat steps 2 and 3 until the centroids no longer move. This produces a separation of the funds into clusters from which the metric to be minimized can be calculated.

D. Global Optimal

There is a local minima problem with K-means and K-medoids because the initial partition is set arbitrarily. To avoid the local minima when minimizing the within cluster Euclidean/Manhattan distance, the randomization process is replicated hundreds times.

E. Choice of Number of Clusters

The number of clusters must be specified before the clustering. The number of clusters is not known *a priori* for optimal manager selection. So the numbers 2, 3, and 4 of clusters is applied for clustering. The cluster results are compared and the manager transitions are tracked when the number of clusters increasing.

F. Measure of Representation

The Euclidean/Manhattan distance is a measure of clustering. However the representative cluster is selected on the basis of the correlation between the centroid and the corresponding strategy's CFSB benchmark.

Results

The resulting clusters, when the number of clusters is constrained to three, and using the four different combinations discussed above, are presented in the figures 4A through 4D. The managers are presented in the risk-return space by strategy.

Figure 5 presents the optimal cluster identification via K-means and K-medoids methods to minimize Type I error, while figure 6 presents the cluster selection under the two methods when minimizing Type II error.

Figure 7 and figure 8 summarize the size of the optimal cluster for the different methodologies and the objectives of minimizing Type 1 and Type II error, respectively.

Figure 4A. Convertible Arbitrage Risk vs. Return Scatter Plot

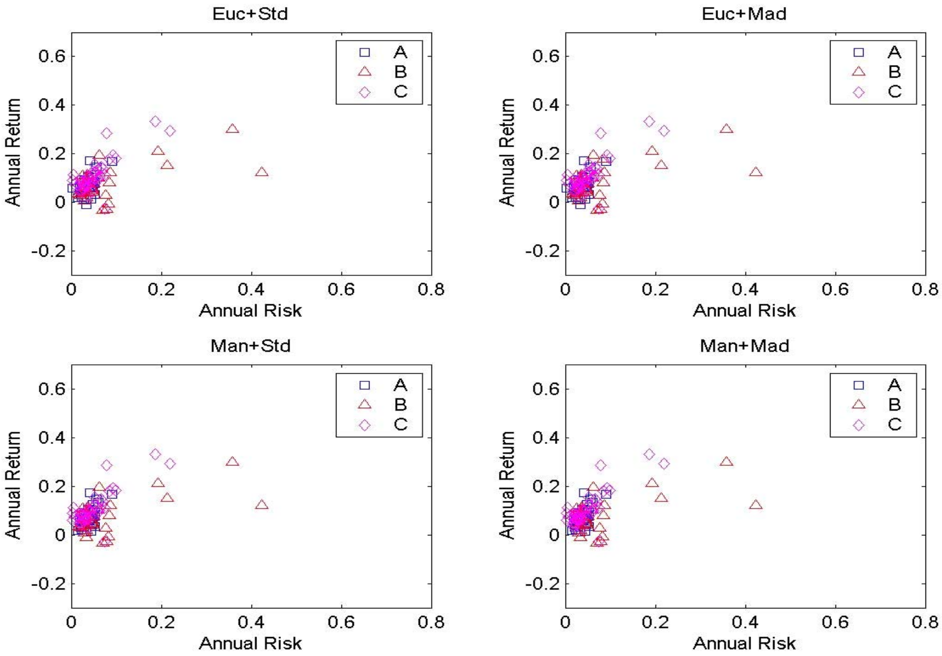


Figure 4B. Distressed Securities Risk vs. Return Scatter Plot

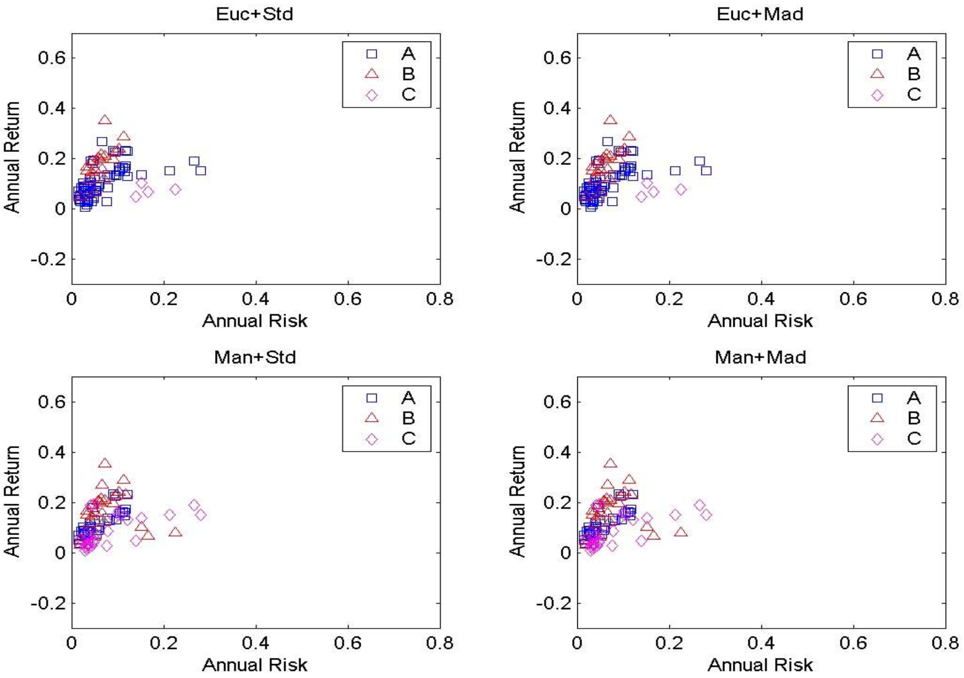


Figure 4C. Event Driven Risk vs. Return Scatter Plot

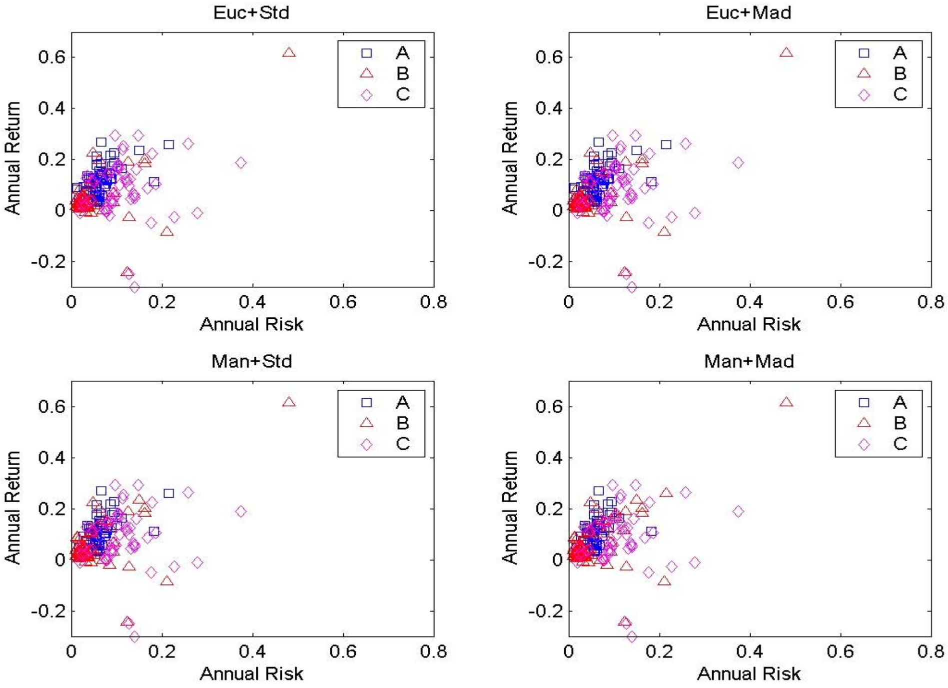
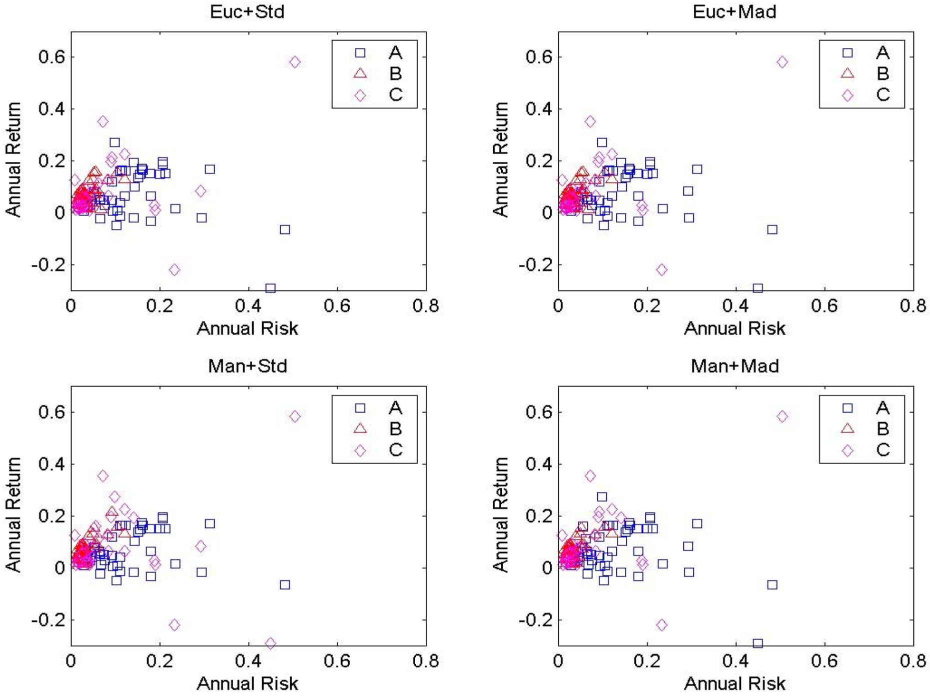


Figure 4D. Merger Arbitrage Risk vs. Return Scatter Plot



Sources: Manager data from TASS, HFR, Alvest, Hedgefund.net, CISDM and Cogent.

Figure 5 Cluster Results For Manager Selection (Minimize **Type I Error**)

Kmeans Clustering Method

	Convertible Arbitrage			Distressed Securities			Event Driven			Merger Arbitrage		
	2	3	4	2	3	4	2	3	4	2	3	4
Euclidean & Std	A	C	A	A	A	A	A	A	A	A	A	A
Euclidean & Mad	A	C	A	A	A	A	A	A	A	A	A	A
Manhattan & Std	A	C	A	A	A	A	A	A	A	A	A	A
Manhattan & Mad	A	C	A	A	A	A	A	A	A	A	A	A

Kmedoids Clustering Method

	Convertible Arbitrage			Distressed Securities			Event Driven			Merger Arbitrage		
	2	3	4	2	3	4	2	3	4	2	3	4
Euclidean & Std	A	C	C	B	B	B	A	A	A	B	B	A
Euclidean & Mad	A	A	A	A	A	A	A	A	D	B	B	A
Manhattan & Std	A	C	C	B	B	B	A	A	A	B	B	A
Manhattan & Mad	A	A	A	A	A	A	A	A	D	B	B	A

Figure 6 Cluster Results For Manager Selection (Minimize **Type II Error**)

Kmeans Clustering Method

	Convertible Arbitrage			Distressed Securities			Event Driven			Merger Arbitrage		
	2	3	4	2	3	4	2	3	4	2	3	4
Euclidean & Std	A	AC	ABC	A	AC	AC	A	AB	ABC	A	AB	ABC
Euclidean & Mad	A	AC	ABC	A	AC	AC	A	AB	ABC	A	AB	ABC
Manhattan & Std	A	AC	ABC	A	AC	AC	A	AB	ABC	A	AB	ABC
Manhattan & Mad	A	AC	ABC	A	AC	AC	A	AB	ABC	A	AB	ABC

Kmedoids Clustering Method

	Convertible Arbitrage			Distressed Securities			Event Driven			Merger Arbitrage		
	2	3	4	2	3	4	2	3	4	2	3	4
Euclidean & Std	A	AC	AC	B	AB	ABD	A	AB	ACD	B	AB	ABC
Euclidean & Mad	A	AC	ACD	A	AC	ABC	A	AB	ABD	B	AB	ABC
Manhattan & Std	A	AC	AC	B	AB	ABD	A	AB	ACD	B	AB	ABC
Manhattan & Mad	A	AC	ACD	A	AC	ABC	A	AB	ABD	B	AB	ABC

Figure 7. Cluster Results For Manager Selection (Minimize **Type I Error**)

Kmeans Clustering Method

	Convertible Arbitrage (112)			Distressed Securities (88)			Event Driven (181)			Merger Arbitrage (116)		
	2	3	4	2	3	4	2	3	4	2	3	4
Euclidean & Std	75	42	33	67	65	47	121	72	64	59	44	45
Euclidean & Mad	75	41	32	66	65	48	116	71	74	59	45	45
Manhattan & Std	75	38	29	63	36	37	102	65	65	59	49	43
Manhattan & Mad	75	43	40	63	36	37	93	63	60	59	50	41

Kmedoids Clustering Method

	Convertible Arbitrage (112)			Distressed Securities (88)			Event Driven (181)			Merger Arbitrage (116)		
	2	3	4	2	3	4	2	3	4	2	3	4
Euclidean & Std	63	31	33	33	50	25	106	67	60	52	48	43
Euclidean & Mad	81	41	33	38	27	23	105	73	33	55	49	38
Manhattan & Std	57	28	25	43	36	26	117	84	52	52	49	33
Manhattan & Mad	78	47	37	56	28	25	98	64	27	57	44	27

Figure 8. Cluster Results For Manager Selection (Minimize **Type II Error**)

K-means Clustering Method

	Convertible Arbitrage (112)			Distressed Securities (88)			Event Driven (181)			Merger Arbitrage (116)		
	2	3	4	2	3	4	2	3	4	2	3	4
Euclidean & Std	75	80	96	67	69	51	121	128	150	59	89	109
Euclidean & Mad	75	79	95	66	69	52	116	128	151	59	90	101
Manhattan & Std	75	79	95	63	61	62	102	130	141	59	85	99
Manhattan & Mad	75	83	84	63	61	62	93	132	146	59	87	95

K-medoids Clustering Method

	Convertible Arbitrage (112)			Distressed Securities (88)			Event Driven (181)			Merger Arbitrage (116)		
	2	3	4	2	3	4	2	3	4	2	3	4
Euclidean & Std	63	71	68	33	78	82	106	110	142	52	95	102
Euclidean & Mad	81	81	86	38	56	74	105	114	132	55	85	103
Manhattan & Std	57	67	56	43	82	63	117	130	133	52	95	95
Manhattan & Mad	78	78	87	56	58	76	98	122	139	57	84	91

Figure 9 presents the impact on the clustering results when the clustering attributes varies, where Y denotes there is an impact and N denotes no impact. Regarding the clustering method, distressed securities shows more divergence between K-means method and K-medoids method while the other three strategies have common results when cluster number less than 4. When the cluster number increases, the divergence increases. The clustering results also show that different distance measure gives different manager selection pool when the rest attributes are fixed. It implies the similarity measure plays critical role in clustering. However the standardization has no impact on the clustering results. It is obvious that the cluster number may change the manager selection. Larger cluster numbers can reduce the size of the pool, which is used for manager selection.

Figure 9. Impact of Clustering Attributes by Strategy and Number of Clusters

	CA			**DS**			**ED**			**MA**		
Attribute	**2**	**3**	**4**	**2**	**3**	**4**	**2**	**3**	**4**	**2**	**3**	**4**
Clustering Method	N	Y	Y	Y	Y	Y	N	N	Y	N	N	Y
Distance Measure	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Standardization	N	N	N	N	N	N	N	N	N	N	N	N

CA = Convertible Arbitrage; DS = Distressed Securities; ED = Event Driven; MA = Merger Arbitrage

VI Conclusions

The clustering exercise results in some interesting findings, some of which may require further investigation and result in further research. The following is a summary of some of the more relevant findings for this group of managers and for the period under consideration:

(i) In general, irrespective of the number of clusters, clusters differ by K-means and K-medoids for all strategies. When the number of clusters is constrained to four, the make up of the clusters that are most representative of the strategy tend to converge using the two methods. As one would expect, the differences in cluster composition are less significant when the goal is to exclude funds that are the least representative of the strategy. But differences still persist for the Convertible Arbitrage and Distressed Securities.

(ii) Across all strategies, clusters obtained through the use of a K-means algorithm are less sensitive than the clusters obtained through the use of a K-medoids algorithm to the choice of the standardization factor and the choice of the distance metric.

(iii) For all of the strategies, excluding merger arbitrage, the K-medoid clusters are more sensitive to the choice of the standardization factor, i.e., mean absolute deviation versus standard deviation, than to the choice the distance metric, i.e., Euclidean versus Manhattan. Merger arbitrage managers are not sensitive to either.

(iv) If the goal is to include only the cluster that is most representative of the strategy, then the K-medoids methodology produce clusters that are more “strategy-pure” than the K-means methodology.[2] The criteria used here to determine the more strategy-pure cluster is intra-cluster correlations between the members together with inter-cluster correlations between cluster centroids.

---

[2] If qualitative due-diligence is an expensive and time-consuming process, then this goal may be more appropriate than the goal of excluding “strategy-impure” managers.

References

Michael R. Anderberg, “Cluster Analysis for Applications,” Academic Press. 1982.

Blatt, M., S. Wiseman and E. Domany, “Data Clustering Using Model Granular Magnet,” Neural Computation, 9, 1805-1842. 1997. (http://arxiv.org/PS_cache/cond-mat/pdf/9702/9702072.pdf)

Nandita Das, “Hedge Fund Classification using K-means Clustering Method,” 9th International Conference on Computing in Economics and Finance, University of Washington, Seattle. July, 2003.

De Souza, Clifford and Suleyman Gokcan, “Hedge Fund Investing: A Quantitative Approach to Hedge Fund Manager Selection and De-Selection,” Journal of Wealth Management, Volume 6, Number 4, pp. 52-73. 2004. (http://www.iijournals.com/JPPM/default.asp?Page=5&ISS=9676)

Dow Jones Hedge Fund Strategy Benchmarks: Index Calculation Methodology, 2005. www.djhedgefundindexes.com .

Fung, W. and D. A. Hsieh, “Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds”, Review of Financial Studies, 10, 275-302. 1997.

Johnson, Richard A. and Wichern, Dean W., “Applied Multivariate Statistical Analysis,” Prentice Hall. 1998.

Kaufman, L. and P.J. Rousseeuw, “Finding Groups in Data: An Introduction to Cluster Analysis”, Wiley. 1990.

Francois-Serge Lhabitant, “Hedge Fund Investing: A Quantitative Look Inside the Black Box.” 2001. (http://www.capco.com/journal.aspx?id=266)

J.B. McQueen, “Some Methods for classification and Analysis of Multivariate Observations,” Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability,” Berkeley, University of California Press, 1:281-297. 1967.

Charles H. Romesburg, “Clustering Analysis for Researchers,” Lifetime Learning Publications. 1988.

William F. Sharpe, “Asset Allocation: Management Style and Performance Measurement,” Journal of Portfolio Management, 18, 7-19, 1992. (http://www.stanford.edu/~wfsharpe/art/sa/sa.htm)

John A. Tucker, “ETFs: A Closer Look at Style Indices.” http://www.ssga.com/library/esps/jtuckeretfsstyleindices20031001/page.html

Appendix – Convertible Arbitrage

Figure 1: Cluster Size and Fund Transitions Across Clusters as a Function of the Number of Clusters
(Totals in Bold)

K-means

	Clusters								
	Two		**Three**			**Four**			
	A	B	A	B	C	A	B	C	D
Euc & Std	**75**	**37**	38A = **38**	32B = **32**	37A+5B = **42**	28A+5C = **33**	29B = **29**	34C = **34**	10A+3B+3C = **16**
Euc & Mad	**75**	**37**	38A = **38**	33B = **33**	37A+4B = **41**	27A+5C = **32**	29B = **29**	33C+1B = **34**	11A+3B+3C = **17**
Man & Std	**75**	**37**	41A = **41**	33B = **33**	34A+4B = **38**	26A+3C = **29**	31B = **31**	35C = **35**	15A+2B = **17**
Man & Mad	**75**	**37**	40A = **40**	29B = **29**	35A+8B = **43**	38A+2C = **40**	19B = **19**	25C = **25**	2A+10B+16C = **28**

K-medoids

	Clusters								
	Two		**Three**			**Four**			
	A	B	A	B	C	A	B	C	D
Euc & Std	**63**	**49**	40A = **40**	41B = **41**	23A+8B = **31**	32A+3C = **35**	24B+4C = **28**	8A+1B+24C = **33**	16B = **16**
Euc & Mad	**81**	**31**	39A+2B = **41**	2A+29B = **31**	40A = **40**	22A+2B+9C = **33**	26B = **26**	2A+30C = **32**	17A+3B+1C = **21**
Man & Std	**57**	**55**	39A = **39**	45B = **45**	18A+10B = **28**	31A = **31**	43B = **43**	25C = **25**	8A+2B+3C = **13**
Man & Mad	**78**	**34**	47A = **47**	34B = **34**	31A = **31**	35A+2B = **37**	1A+24B = **25**	3B+31C = **34**	11A+5B = **16**

Key

Distance Metric	**Standardization**
Euc = Euclidean	Std = Standard Deviation
Man = Manhattan	Mad = Mean Absolute Deviation

Appendix – Convertible Arbitrage

Figure 2: Correlation Across Clusters (Centroids)

K-means

	Clusters									
	Two	**Three**			**Four**					
	A&B	A&B	A&C	B&C	A&B	A&C	A&D	B&C	B&D	C&D
Euc & Std	0.45	0.17	0.77	0.59	0.20	0.78	0.71	0.62	0.32	0.57
Euc & Mad	0.45	0.17	0.77	0.60	0.20	0.78	0.71	0.62	0.29	0.57
Man & Std	0.42	0.18	0.79	0.60	0.21	0.82	0.82	0.58	0.10	0.63
Man & Mad	0.42	0.12	0.76	0.57	0.08	0.78	0.59	0.43	0.68	0.80

K-medoids

	Clusters									
	Two	**Three**			**Four**					
	A&B	A&B	A&C	B&C	A&B	A&C	A&D	B&C	B&D	C&D
Euc & Std	0.32	0.32	0.18	-0.01	0.32	0.30	0.07	-0.05	0.21	0.28
Euc & Mad	-0.16	-0.02	0.18	-0.18	-0.16	0.30	0.05	-0.18	0.44	0.35
Man & Std	0.32	0.32	0.64	0.26	0.32	0.64	0.23	0.26	0.57	0.34
Man & Mad	0.07	0.07	0.30	0.34	-0.16	0.30	-0.05	-0.18	0.43	0.26

Key

Distance Metric	**Standardization**
Euc = Euclidean Man = Manhattan	Std = Standard Deviation Mad = Mean Absolute Deviation

Appendix – Convertible Arbitrage

Figure 3: Correlation Ranges Within Clusters (Among Managers)

K-means

		Clusters								
		Two		**Three**			**Four**			
		A	B	A	B	C	A	B	C	D
Euc & Std	Max	0.94	0.90	0.95	0.90	0.94	0.96	0.90	0.94	0.78
	Med	**0.74**	**0.73**	**0.77**	**0.77**	**0.78**	**0.82**	**0.78**	**0.78**	**0.51**
	Min	0.16	0.12	0.24	0.14	0.26	0.65	0.54	0.28	0.13
Euc & Mad	Max	0.94	0.91	0.95	0.90	0.94	0.96	0.89	0.94	0.82
	Med	**0.74**	**0.73**	**0.77**	**0.75**	**0.78**	**0.83**	**0.78**	**0.78**	**0.52**
	Min	0.17	0.13	0.25	0.13	0.26	0.66	0.54	0.28	0.08
Man & Std	Max	0.95	0.91	0.97	0.90	0.95	0.95	0.91	0.95	0.87
	Med	**0.72**	**0.69**	**0.78**	**0.76**	**0.77**	**0.82**	**0.77**	**0.77**	**0.71**
	Min	0.15	0.09	0.20	0.10	0.28	0.56	0.13	0.28	0.07
Man & Mad	Max	0.95	0.91	0.97	0.92	0.94	0.97	0.91	0.95	0.93
	Med	**0.72**	**0.68**	**0.78**	**0.70**	**0.77**	**0.78**	**0.85**	**0.80**	**0.69**
	Min	0.16	0.09	0.20	0.09	0.26	0.26	0.13	0.27	0.13

K-medoids

		Clusters								
		Two		**Three**			**Four**			
		A	B	A	B	C	A	B	C	D
Euc &Std	Max	1.00	1.00	1.00	1.00	0.48	1.00	1.00	0.42	0.53
	Med	**0.51**	**0.57**	**0.52**	**0.55**	**0.15**	**0.54**	**0.64**	**0.16**	**0.15**
	Min	0.04	-0.11	0.04	-0.11	-0.01	0.07	0.07	-0.26	-0.12
Euc & Mad	Max	1.00	1.00	0.48	1.00	1.00	0.42	1.00	1.00	1.00
	Med	**0.21**	**0.40**	**0.12**	**0.39**	**0.52**	**0.09**	**0.39**	**0.55**	**0.75**
	Min	-0.26	0.06	-0.28	0.06	0.04	-0.26	0.06	0.07	0.51
Man & Std	Max	1.00	1.00	1.00	1.00	0.64	1.00	1.00	0.61	0.76
	Med	**0.53**	**0.57**	**0.54**	**0.55**	**0.41**	**0.54**	**0.55**	**0.41**	**0.49**
	Min	0.04	-0.11	0.07	-0.11	0.17	0.07	-0.11	0.17	0.07
Man & Mad	Max	1.00	0.82	0.42	0.82	1.00	0.42	0.68	1.00	0.80
	Med	**0.22**	**0.49**	**0.10**	**0.49**	**0.58**	**0.09**	**0.40**	**0.57**	**0.65**
	Min	-0.26	-0.09	-0.26	-0.09	0.18	-0.26	0.06	0.11	0.43

Key

Distance Metric	**Standardization**
Euc = Euclidean Man = Manhattan	Std = Standard Deviation Mad = Mean Absolute Deviation

Appendix – Convertible Arbitrage

Figure 4: Common Funds Across Clusters Irrespective of the Distance Metric

K-means

Standardization = Standard Deviation

Standardization = Mean Absolute Deviation

Two Clusters

Standardization = Standard Deviation

		Manhattan		
		A	**B**	**Total**
Euclidean	**A**	72	3	**75**
	B	3	34	**37**
	Total	**75**	**37**	**112**

Standardization = Mean Absolute Deviation

	Manhattan		
	A	**B**	**Total**
A	72	3	**75**
B	3	34	**37**
Total	**75**	**37**	**112**

Three Clusters

Standardization = Standard Deviation

		Manhattan			
		A	**B**	**C**	**Total**
	A	36	1	1	**38**
Euclidean	**B**	0	32	0	**32**
	C	5	0	37	**42**
	Total	**41**	**33**	**38**	**112**

Standardization = Mean Absolute Deviation

	Manhattan			
	A	**B**	**C**	**Total**
A	35	1	2	**38**
B	0	28	5	**33**
C	5	0	36	**41**
Total	**40**	**29**	**43**	**112**

Four Clusters

Standardization = Standard Deviation

		Manhattan				
		A	**B**	**C**	**D**	**Total**
	A	25	0	2	6	**33**
Euclidean	**B**	0	29	0	0	**29**
	C	4	0	30	0	**34**
	D	0	2	3	11	**16**
	Total	**29**	**31**	**35**	**17**	**112**

Standardization = Mean Absolute Deviation

	Manhattan				
	A	**B**	**C**	**D**	**Total**
A	30	0	2	0	**32**
B	0	18	0	11	**29**
C	3	0	23	8	**34**
D	7	1	0	9	**17**
Total	**40**	**19**	**25**	**28**	**112**

Appendix – Convertible Arbitrage

K-medoids

Standardization = Standard Deviation | Standardization = Mean Absolute Deviation

Two Clusters

		Manhattan					Manhattan		
		A	**B**	**Total**			**A**	**B**	**Total**
Euclidean	**A**	57	6	**63**		**A**	76	5	**81**
	B	0	49	**49**		**B**	2	29	**31**
	Total	**57**	**55**	**112**		**Total**	**78**	**34**	**112**

Three Clusters

		Manhattan						Manhattan			
		A	**B**	**C**	**Total**			**A**	**B**	**C**	**Total**
Euclidean	**A**	32	3	5	**40**		**A**	34	4	3	**41**
	B	0	41	0	**41**		**B**	4	27	0	**31**
	C	7	1	23	**31**		**C**	9	3	28	**40**
	Total	**39**	**45**	**28**	**112**		**Total**	**47**	**34**	**31**	**112**

Four Clusters

		Manhattan							Manhattan				
		A	**B**	**C**	**D**	**Total**			**A**	**B**	**C**	**D**	**Total**
Euclidean	**A**	27	1	5	2	**35**		**A**	28	0	4	1	**33**
	B	0	24	4	0	**28**		**B**	1	24	0	1	**26**
	C	4	2	16	11	**33**		**C**	2	0	30	0	**32**
	D	0	16	0	0	**16**		**D**	6	1	0	14	**21**
	Total	**31**	**43**	**25**	**13**	**112**		**Total**	**37**	**25**	**34**	**16**	**112**

Appendix – Convertible Arbitrage

Figure 5 Median Correlations of Common Funds in Figure 4 with Broad Benchmark

K-means

Standardization = Standard Deviation

Standardization = Mean Absolute Deviation

Two Clusters

Standardization = Standard Deviation

		Manhattan		
		A	**B**	**Total**
Euclidean	**A**	0.65	0.60	**0.65**
	B	0.72	0.38	**0.40**
	Total	**0.65**	**0.39**	**0.59**

Standardization = Mean Absolute Deviation

	Manhattan		
	A	**B**	**Total**
A	0.65	0.60	**0.65**
B	0.72	0.38	**0.40**
Total	**0.65**	**0.39**	**0.59**

Three Clusters

Standardization = Standard Deviation

		Manhattan			
		A	**B**	**C**	**Total**
Euclidean	**A**	0.62	0.22	0.53	**0.60**
	B	-	0.36	-	**0.36**
	C	0.65	-	0.66	**0.66**
	Total	**0.64**	**0.35**	**0.66**	**0.59**

Standardization = Mean Absolute Deviation

	Manhattan			
	A	**B**	**C**	**Total**
A	0.64	0.33	0.45	**0.60**
B	-	0.32	0.64	**0.37**
C	0.65	-	0.66	**0.66**
Total	**0.64**	**0.31**	**0.65**	**0.59**

Four Clusters

Standardization = Standard Deviation

		Manhattan				
		A	**B**	**C**	**D**	**Total**
Euclidean	**A**	0.70	-	0.68	0.56	**0.67**
	B	-	0.37	-	-	**0.37**
	C	-	0.10	0.35	0.22	**0.34**
	D	-	0.10	0.35	0.22	**0.34**
	Total	**0.70**	**0.37**	**0.66**	**0.41**	**0.59**

Standardization = Mean Absolute Deviation

	Manhattan				
	A	**B**	**C**	**D**	**Total**
A	0.68	-	0.68	-	**0.68**
B	-	0.32	-	0.50	**0.37**
C	0.48	-	0.66	0.70	**0.66**
D	0.44	-0.21	-	0.35	**0.35**
Total	**0.64**	**0.31**	**0.66**	**0.51**	**0.59**

Appendix – Convertible Arbitrage

K-medoids

Standardization = Standard Deviation

Standardization = Mean Absolute Deviation

Two Clusters

		Manhattan		Total
		A	**B**	
Euclidean	**A**	0.64	0.50	**0.63**
	B	-	0.42	**0.42**
	Total	**0.64**	**0.44**	**0.59**

	Manhattan		Total
	A	**B**	
A	0.65	0.58	**0.65**
B	0.47	0.34	**0.35**
Total	**0.65**	**0.36**	**0.59**

Three Clusters

		Manhattan			Total
		A	**B**	**C**	
Euclidean	**A**	0.62	0.45	0.42	**0.59**
	B	-	0.40	-	**0.40**
	C	0.72	0.72	0.67	**0.68**
	Total	**0.64**	**0.40**	**0.66**	**0.59**

	Manhattan			Total
	A	**B**	**C**	
A	0.68	0.65	0.72	**0.68**
B	0.43	0.31	-	**0.34**
C	0.59	0.42	0.60	**0.59**
Total	**0.66**	**0.36**	**0.61**	**0.59**

Four Clusters

		Manhattan				Total
		A	**B**	**C**	**D**	
Euclidean	**A**	0.61	0.37	0.48	0.60	**0.59**
	B	-	0.47	0.67	-	**0.51**
	C	0.68	0.66	0.68	0.70	**0.68**
	D	-	0.29	-	-	**0.29**
	Total	**0.63**	**0.40**	**0.66**	**0.65**	**0.59**

	Manhattan				Total
	A	**B**	**C**	**D**	
A	0.70	-	0.65	0.60	**0.68**
B	0.54	0.30	-	0.30	**0.31**
C	0.35	-	0.59	-	**0.57**
D	0.63	0.45	-	0.66	**0.65**
Total	**0.66**	**0.31**	**0.59**	**0.65**	**0.59**

Appendix – Convertible Arbitrage

Figure 6: Common Funds Across Clusters Irrespective of the Clustering Method

Euclidean Distance
Standard Deviation

Euclidean Distance
Mean Absolute Deviation

Manhattan Distance
Standard Deviation

Manhattan Distance
Mean Absolute Deviation

Two Clusters

K-means

K-medoids		A	B	Total		A	B	Total		A	B	Total		A	B	Total
	A	63	0	**63**	**A**	75	6	**81**	**A**	57	0	**57**	**A**	71	7	**78**
	B	12	37	**49**	**B**	0	31	**31**	**B**	18	37	**55**	**B**	4	30	**34**
	Total	**75**	**37**	**112**	**Total**	**75**	**37**	**112**	**Total**	**75**	**37**	**112**	**Total**	**75**	**37**	**112**

Three Clusters

K-means

K-medoids		A	B	C	Total		A	B	C	Total		A	B	C	Total		A	B	C	Total
	A	30	0	10	**40**	**A**	7	3	31	**41**	**A**	37	0	2	**39**	**A**	10	2	35	**47**
	B	1	32	8	**41**	**B**	1	30	0	**31**	**B**	2	33	10	**45**	**B**	2	27	5	**34**
	C	7	0	24	**31**	**C**	30	0	10	**40**	**C**	2	0	26	**28**	**C**	28	0	3	**31**
	Total	**38**	**32**	**42**	**112**	**Total**	**38**	**33**	**41**	**112**	**Total**	**41**	**33**	**38**	**112**	**Total**	**40**	**29**	**43**	**112**

Four Clusters

K-means

K-medoids		A	B	C	D	Total		A	B	C	D	Total		A	B	C	D	Total		A	B	C	D	Total
	A	24	0	3	8	**35**	**A**	11	0	16	6	**33**	**A**	19	0	2	10	**31**	**A**	9	0	13	15	**37**
	B	0	14	11	3	**28**	**B**	0	24	0	2	**26**	**B**	0	31	9	3	**43**	**B**	0	19	0	6	**25**
	C	9	0	20	4	**33**	**C**	21	0	2	9	**32**	**C**	3	0	21	1	**25**	**C**	31	0	2	1	**34**
	D	0	15	0	1	**16**	**D**	0	5	16	0	**21**	**D**	7	0	3	3	**13**	**D**	0	0	10	6	**16**
	Total	**33**	**29**	**34**	**16**	**112**	**Total**	**32**	**29**	**34**	**17**	**112**	**Total**	**29**	**31**	**35**	**17**	**112**	**Total**	**40**	**19**	**25**	**28**	**112**

Appendix – Distressed Securities

Figure 1: Cluster Size and Fund Transitions Across Clusters as a Function of the Number of Clusters
(Totals in Bold)

K-means

	Clusters								
	Two		**Three**			**Four**			
	A	B	A	B	C	A	B	C	D
Euc & Std	**67**	**21**	65A = **65**	2A+17B = **19**	4B = **4**	46A+1B = **47**	18B = **18**	4C = **4**	19A = **19**
Euc & Mad	**66**	**22**	65A = **65**	1A+18B = **19**	4B = **4**	46A+2B = **48**	17B = **17**	4C = **4**	19A = **19**
Man & Std	**63**	**25**	35A+1B = **36**	1A+24B = **25**	27A = **27**	36A+1C = **37**	22B+1C = **23**	25C = **25**	3B = **3**
Man & Mad	**63**	**25**	36A = **36**	25B = **25**	27A = **27**	36A+1C = **37**	22B+1C = **23**	25C = **25**	3B = **3**

K-medoids

	Clusters								
	Two		**Three**			**Four**			
	A	B	A	B	C	A	B	C	D
Euc & Std	**55**	**33**	24A+4B = **28**	23A+27B = **50**	8A+2B = **10**	22A+5C = **27**	24B+1C = **25**	1A+2B+3C = **6**	5A+24B+1C = **30**
Euc & Mad	**38**	**50**	21A+6B = **27**	3A+29B = **32**	14A+15B = **29**	23A = **23**	28B = **28**	23C = **23**	4A+4B+6C = **14**
Man & Std	**45**	**43**	45A+1B = **46**	36B = **36**	6B = **6**	17A+4B+2C = **23**	1A+22B+3C = **26**	23A+2B = **25**	5A+8B+1C = **14**
Man & Mad	**56**	**32**	28A = **28**	30B = **30**	28A+2B = **30**	25A = **25**	29B = **29**	22C = **22**	3A+1B+8C = **12**

Key

Distance Metric	**Standardization**
Euc = Euclidean Man = Manhattan	Std = Standard Deviation Mad = Mean Absolute Deviation

Appendix – Distressed Securities

Figure 2: Correlation Across Clusters (Centroids)

K-means

	Clusters									
	Two	**Three**			**Four**					
	A&B	A&B	A&C	B&C	A&B	A&C	A&D	B&C	B&D	C&D
Euc & Std	0.53	0.60	0.06	0.05	0.63	0.09	0.82	0.03	-0.01	0.53
Euc & Mad	0.52	0.59	0.05	0.05	0.60	0.07	0.82	0.04	-0.03	0.52
Man & Std	0.66	0.64	0.83	0.58	0.70	0.82	0.02	0.62	0.01	0.66
Man & Mad	0.66	0.65	0.83	0.59	0.70	0.82	0.01	0.61	0.02	0.66

K-medoids

	Clusters									
	Two	**Three**			**Four**					
	A&B	A&B	A&C	B&C	A&B	A&C	A&D	B&C	B&D	C&D
Euc & Std	0.52	0.04	0.08	0.29	0.04	0.10	0.29	-0.12	0.80	-0.07
Euc & Mad	0.72	0.29	0.88	0.26	0.29	0.88	0.20	0.26	-0.01	0.06
Man & Std	0.69	0.52	0.24	-0.12	0.29	0.29	0.36	0.92	0.63	0.73
Man & Mad	0.29	0.29	0.88	0.26	0.29	0.88	0.73	0.26	0.36	0.58

Key

Distance Metric	**Standardization**
Euc = Euclidean Man = Manhattan	Std = Standard Deviation Mad = Mean Absolute Deviation

Appendix – Distressed Securities

Figure 3: Correlation Ranges Within Clusters (Among Managers)

K-means

		Clusters								
		Two		**Three**			**Four**			
		A	B	A	B	C	A	B	C	D
Euc & Std	Max	0.95	0.74	0.95	0.74	0.98	0.97	0.73	0.98	0.93
	Med	**0.79**	**0.45**	**0.79**	**0.54**	**0.97**	**0.82**	**0.51**	**0.97**	**0.85**
	Min	0.52	-0.20	0.53	-0.03	0.73	0.64	-0.02	0.73	0.61
Euc & Mad	Max	0.95	0.73	0.95	0.73	0.98	0.97	0.74	0.98	0.93
	Med	**0.79**	**0.45**	**0.79**	**0.54**	**0.97**	**0.82**	**0.50**	**0.97**	**0.85**
	Min	0.52	-0.24	0.53	-0.05	0.72	0.61	-0.05	0.72	0.61
Man & Std	Max	0.95	0.73	0.98	0.73	0.92	0.98	0.74	0.93	1.00
	Med	**0.79**	**0.47**	**0.84**	**0.42**	**0.83**	**0.83**	**0.55**	**0.84**	**0.99**
	Min	0.24	0.01	0.64	-0.01	0.35	0.65	-0.01	0.33	0.99
Man & Mad	Max	0.95	0.73	0.98	0.73	0.92	0.98	0.74	0.93	1.00
	Med	**0.79**	**0.42**	**0.84**	**0.42**	**0.83**	**0.82**	**0.54**	**0.84**	**0.99**
	Min	0.39	0.00	0.65	0.00	0.35	0.65	-0.02	0.33	0.99

K-medoids

		Clusters								
		Two		**Three**			**Four**			
		A	B	A	B	C	A	B	C	D
Euc &Std	Max	0.63	0.77	0.36	0.80	1.00	0.28	0.77	1.00	1.00
	Med	**0.47**	**0.58**	**0.07**	**0.59**	**0.41**	**0.07**	**0.58**	**0.76**	**0.73**
	Min	-0.10	-0.12	-0.22	-0.12	0.02	-0.22	-0.08	0.21	0.21
Euc & Mad	Max	1.00	1.00	1.00	0.28	1.00	1.00	0.28	1.00	0.84
	Med	**0.79**	**0.56**	**0.81**	**0.07**	**0.68**	**0.84**	**0.07**	**0.75**	**0.07**
	Min	-0.07	-0.07	0.16	-0.22	-0.03	0.57	-0.22	0.15	-0.03
Man & Std	Max	0.63	1.00	0.63	0.77	1.00	0.26	1.00	1.00	0.60
	Med	**0.47**	**0.69**	**0.47**	**0.58**	**0.76**	**0.07**	**0.72**	**0.71**	**0.46**
	Min	-0.02	-0.09	-0.02	-0.08	-0.01	-0.22	0.01	0.16	-0.10
Man & Mad	Max	1.00	0.26	1.00	0.26	1.00	1.00	0.26	1.00	0.60
	Med	**0.72**	**0.07**	**0.76**	**0.07**	**0.67**	**0.73**	**0.07**	**0.73**	**0.46**
	Min	0.16	-0.22	0.16	-0.22	0.15	0.16	-0.22	0.20	-0.10

Key

Distance Metric	**Standardization**
Euc = Euclidean	Std = Standard Deviation
Man = Manhattan	Mad = Mean Absolute Deviation

Appendix – Distressed Securities

Figure 4: Common Funds Across Clusters Irrespective of the Distance Metric

K-means

Standardization = Standard Deviation

Standardization = Mean Absolute Deviation

Two Clusters

Standardization = Standard Deviation

		Manhattan		
		A	**B**	**Total**
Euclidean	**A**	61	6	**67**
	B	2	19	**21**
	Total	**63**	**25**	**88**

Standardization = Mean Absolute Deviation

	Manhattan		
	A	**B**	**Total**
A	62	4	**66**
B	1	21	**22**
Total	**63**	**25**	**88**

Three Clusters

Standardization = Standard Deviation

		Manhattan			
		A	**B**	**C**	**Total**
Euclidean	**A**	36	3	26	**65**
	B	0	19	0	**19**
	C	0	3	1	**4**
	Total	**36**	**25**	**27**	**88**

Standardization = Mean Absolute Deviation

	Manhattan			
	A	**B**	**C**	**Total**
A	36	3	26	**65**
B	0	19	0	**19**
C	0	3	1	**4**
Total	36	25	27	**88**

Four Clusters

Standardization = Standard Deviation

		Manhattan				
		A	**B**	**C**	**D**	**Total**
Euclidean	**A**	37	4	6	0	**47**
	B	0	18	0	0	**18**
	C	0	0	1	3	**4**
	D	0	1	18	0	**19**
	Total	**37**	**23**	**25**	**3**	**88**

Standardization = Mean Absolute Deviation

	Manhattan				
	A	**B**	**C**	**D**	**Total**
A	37	5	6	0	**48**
B	0	17	0	0	**17**
C	0	0	1	3	**4**
D	0	1	18	0	**19**
Total	**37**	**23**	**25**	**3**	**88**

Appendix – Distressed Securities

K-medoids

Standardization = Standard Deviation | Standardization = Mean Absolute Deviation

Two Clusters

		Manhattan					Manhattan		
		A	**B**	**Total**			**A**	**B**	**Total**
Euclidean	**A**	42	13	**55**		**A**	33	5	**38**
	B	3	30	**33**		**B**	23	27	**50**
	Total	**45**	**43**	**88**		**Total**	**56**	**32**	**88**

Three Clusters

		Manhattan						Manhattan			
		A	**B**	**C**	**Total**			**A**	**B**	**C**	**Total**
Euclidean	**A**	20	7	1	**28**		**A**	23	1	3	**27**
	B	20	28	2	**50**		**B**	2	27	3	**32**
	C	6	1	3	**10**		**C**	3	2	24	**29**
	Total	**46**	**36**	**6**	**88**		**Total**	**28**	**30**	**30**	**88**

Four Clusters

		Manhattan							Manhattan				
		A	**B**	**C**	**D**	**Total**			**A**	**B**	**C**	**D**	**Total**
Euclidean	**A**	19	0	3	5	**27**		**A**	17	1	1	4	**23**
	B	1	19	1	4	**25**		**B**	2	22	0	4	**28**
	C	2	3	1	0	**6**		**C**	3	0	17	3	**23**
	D	1	4	20	5	**30**		**D**	3	6	4	1	**14**
	Total	**23**	**26**	**25**	**14**	**88**		**Total**	**25**	**29**	**22**	**12**	**88**

Appendix – Distressed Securities

Figure 5 Median Correlations of Common Funds in Figure 4 with Broad Benchmark

K-means

Two Clusters

		Manhattan		
		A	**B**	**Total**
Euclidean	**A**	0.74	0.57	**0.72**
	B	0.25	0.18	**0.18**
	Total	**0.74**	**0.25**	**0.66**

	Manhattan		
	A	**B**	**Total**
A	0.74	0.56	**0.72**
B	0.38	0.18	**0.21**
Total	**0.74**	**0.25**	**0.66**

Three Clusters

		Manhattan			
		A	**B**	**C**	**Total**
Euclidean	**A**	0.83	0.55	0.67	**0.72**
	B	-	0.25	-	**0.25**
	C	-	0.02	0.38	**0.03**
	Total	**0.83**	**0.25**	**0.67**	**0.66**

	Manhattan			
	A	**B**	**C**	**Total**
A	0.83	0.55	0.67	**0.72**
B	-	0.25	-	**0.25**
C	-	0.02	0.38	**0.03**
Total	**0.83**	**0.25**	**0.67**	**0.66**

Four Clusters

		Manhattan				
		A	**B**	**C**	**D**	**Total**
Euclidean	**A**	0.83	0.59	0.69	-	**0.76**
	B	-	0.24	-	-	**0.24**
	C	-	-	0.38	0.02	**0.03**
	D	-	0.42	0.66	-	**0.66**
	Total	**0.83**	**0.26**	**0.67**	**0.02**	**0.66**

	Manhattan				
	A	**B**	**C**	**D**	**Total**
A	0.83	0.57	0.69	-	**0.76**
B	-	0.24	-	-	**0.24**
C	-	-	0.38	0.02	**0.03**
D	-	0.42	0.66	-	**0.66**
Total	**0.83**	**0.26**	**0.67**	**0.02**	**0.66**

Appendix – Distressed Securities

K-medoids

Two Clusters

		Manhattan					Manhattan		
		A	**B**	**Total**			**A**	**B**	**Total**
Euclidean	**A**	0.68	0.42	**0.65**		**A**	0.83	0.03	**0.79**
	B	0.61	0.71	**0.70**		**B**	0.61	0.65	**0.63**
	Total	**0.67**	**0.66**	**0.66**		**Total**	**0.68**	**0.63**	**0.66**

Three Clusters

		Manhattan						Manhattan			
		A	**B**	**C**	**Total**			**A**	**B**	**C**	**Total**
	A	0.68	0.66	0.38	**0.66**		**A**	0.79	0.82	0.53	**0.77**
Euclidean	**B**	0.78	0.71	0.02	**0.73**		**B**	0.37	0.65	0.60	**0.63**
	C	0.21	0.24	0.03	**0.16**		**C**	0.61	0.02	0.68	**0.64**
	Total	**0.67**	**0.67**	**0.03**	**0.66**		**Total**	**0.75**	**0.64**	**0.65**	**0.66**

Four Clusters

		Manhattan							Manhattan				
		A	**B**	**C**	**D**	**Total**			**A**	**B**	**C**	**D**	**Total**
	A	0.66	-	0.18	0.64	**0.65**		**A**	0.83	0.82	0.53	0.83	**0.82**
Euclidean	**B**	0.70	0.74	0.61	0.55	**0.72**		**B**	0.37	0.67	-	0.62	**0.65**
	C	0.26	0.02	0.13	-	**0.08**		**C**	0.61	-	0.74	0.24	**0.68**
	D	0.82	0.54	0.71	0.69	**0.68**		**D**	0.26	0.03	0.21	0.24	**0.17**
	Total	**0.66**	**0.68**	**0.64**	**0.65**	**0.66**		**Total**	**0.67**	**0.65**	**0.68**	**0.65**	**0.66**

Appendix – Distressed Securities

Figure 6: Common Funds Across Clusters Irrespective of the Clustering Method

Euclidean Distance Standard Deviation	Euclidean Distance Mean Absolute Deviation	Manhattan Distance Standard Deviation	Manhattan Distance Mean Absolute Deviation

Two Clusters

K-means

		A	B	Total		A	B	Total		A	B	Total		A	B	Total
K-medoids	**A**	41	14	**55**	**A**	29	9	**38**	**A**	35	10	**45**	**A**	40	16	**56**
	B	26	7	**33**	**B**	37	13	**50**	**B**	28	15	**43**	**B**	23	9	**32**
	Total	**67**	**21**	**88**	**Total**	**66**	**22**	**88**	**Total**	**63**	**25**	**88**	**Total**	**63**	**25**	**88**

Three Clusters

K-means

		A	B	C	Total		A	B	C	Total		A	B	C	Total		A	B	C	Total
K-medoids	**A**	26	1	1	**28**	**A**	22	5	0	**27**	**A**	16	11	19	**46**	**A**	19	5	4	**28**
	B	39	9	2	**50**	**B**	23	7	2	**32**	**B**	20	9	7	**36**	**B**	1	8	21	**30**
	C	0	9	1	**10**	**C**	20	7	2	**29**	**C**	0	5	1	**6**	**C**	16	12	2	**30**
	Total	**65**	**19**	**4**	**88**	**Total**	**65**	**19**	**4**	**88**	**Total**	**36**	**25**	**27**	**88**	**Total**	**36**	**25**	**27**	**88**

Four Clusters

K-means

		A	B	C	D	Total		A	B	C	D	Total		A	B	C	D	Total		A	B	C	D	Total
K-medoids	**A**	5	6	0	16	**27**	**A**	20	1	0	2	**23**	**A**	1	3	19	0	**23**	**A**	17	6	2	0	**25**
	B	21	4	0	0	**25**	**B**	7	3	1	17	**28**	**B**	17	6	0	3	**26**	**B**	1	5	20	3	**29**
	C	0	2	4	0	**6**	**C**	21	2	0	0	**23**	**C**	14	9	2	0	**25**	**C**	14	8	0	0	**22**
	D	21	6	0	3	**30**	**D**	0	11	3	0	**14**	**D**	5	5	4	0	**14**	**D**	5	4	3	0	**12**
	Total	**47**	**18**	**4**	**19**	**88**	**Total**	**48**	**17**	**4**	**19**	**88**	**Total**	**37**	**23**	**25**	**3**	**88**	**Total**	**37**	**23**	**25**	**3**	**88**

Appendix – Event Driven

Figure 1: Cluster Size and Fund Transitions Across Clusters as a Function of the Number of Clusters
(Totals in Bold)

K-means

	Clusters								
	Two		**Three**			**Four**			
	A	B	A	B	C	A	B	C	D
Euc & Std	**121**	**60**	71A+1B = **72**	1A+55B = **56**	49A+4B = **53**	64A = **64**	32B = **32**	7A+1B+46C = **54**	1A+23B+7C = **31**
Euc & Mad	**116**	**65**	69A+2B = **71**	57B = **57**	47A+6B = **53**	70A+4B = **74**	28B = **28**	49C = **49**	1A+25B+4C = **30**
Man & Std	**102**	**79**	61A+4B = **65**	4A+61B = **65**	37A+14B = **51**	61A+3B+1C = **65**	33B = **33**	1A+42C = **43**	3A+29B+8C = **40**
Man & Mad	**93**	**88**	59A+4B = **63**	69B = **69**	34A+15B = **49**	58A+1B+1C = **60**	45B = **45**	41C = **41**	5A+23B+7C = **35**

K-medoids

	Clusters								
	Two		**Three**			**Four**			
	A	B	A	B	C	A	B	C	D
Euc & Std	**106**	**75**	51A+16B = **67**	13A+30B = **43**	42A+29B = **71**	60A = **60**	39B = **39**	52C = **52**	7A+4B+19C = **30**
Euc & Mad	**105**	**76**	55A+18B = **73**	11A+30B = **41**	39A+28B = **67**	63A = **63**	36B = **36**	49C = **49**	10A+5B+18C = **33**
Man & Std	**117**	**64**	84A = **84**	46B = **46**	33A+18B = **51**	50A+2B = **52**	20A+22B+6C = **48**	9B+45C = **54**	14A+13B = **27**
Man & Mad	**98**	**83**	64A = **64**	58B = **58**	34A+25B = **59**	58A = **58**	54B = **54**	42C = **42**	6A+4B+17C = **27**

Key

Distance Metric	**Standardization**
Euc = Euclidean Man = Manhattan	Std = Standard Deviation Mad = Mean Absolute Deviation

Appendix – Event Driven

Figure 2: Correlation Across Clusters (Centroids)

K-means

	Clusters									
	Two	**Three**			**Four**					
	A&B	A&B	A&C	B&C	A&B	A&C	A&D	B&C	B&D	C&D
Euc & Std	0.62	0.63	0.81	0.47	0.59	0.82	0.58	0.55	0.50	0.62
Euc & Mad	0.63	0.63	0.81	0.47	0.58	0.81	0.56	0.54	0.35	0.63
Man & Std	0.76	0.66	0.80	0.61	0.60	0.79	0.70	0.49	0.56	0.76
Man & Mad	0.79	0.69	0.79	0.63	0.58	0.78	0.80	0.51	0.56	0.79

K-medoids

	Clusters									
	Two	**Three**			**Four**					
	A&B	A&B	A&C	B&C	A&B	A&C	A&D	B&C	B&D	C&D
Euc & Std	0.49	0.59	0.38	0.08	0.59	0.38	0.59	0.08	0.40	0.78
Euc & Mad	0.49	-0.04	0.79	0.08	-0.04	0.79	0.76	0.08	0.33	0.79
Man & Std	0.81	0.81	0.80	0.81	0.50	0.80	0.53	0.47	0.44	0.43
Man & Mad	0.38	0.80	0.79	0.80	0.80	0.79	0.76	0.80	0.77	0.78

Key

Distance Metric	**Standardization**
Euc = Euclidean Man = Manhattan	Std = Standard Deviation Mad = Mean Absolute Deviation

Appendix – Event Driven

Figure 3: Correlation Ranges Within Clusters (Among Managers)

K-means

		Clusters								
		Two		**Three**			**Four**			
		A	B	A	B	C	A	B	C	D
Euc & Std	Max	0.92	0.86	0.93	0.87	0.92	0.93	0.90	0.91	0.66
	Med	**0.74**	**0.49**	**0.79**	**0.53**	**0.73**	**0.81**	**0.71**	**0.76**	**0.35**
	Min	0.29	-0.03	0.33	-0.06	0.24	0.53	0.36	0.38	-0.14
Euc & Mad	Max	0.93	0.87	0.93	0.87	0.92	0.93	0.89	0.92	0.66
	Med	**0.75**	**0.49**	**0.79**	**0.53**	**0.72**	**0.79**	**0.73**	**0.74**	**0.36**
	Min	0.30	-0.07	0.33	-0.07	0.25	0.50	0.38	0.37	-0.10
Man & Std	Max	0.91	0.86	0.93	0.87	0.90	0.93	0.89	0.90	0.71
	Med	**0.77**	**0.47**	**0.80**	**0.52**	**0.71**	**0.80**	**0.69**	**0.75**	**0.36**
	Min	0.52	-0.09	0.55	-0.06	0.24	0.51	0.36	0.37	-0.18
Man & Mad	Max	0.92	0.84	0.93	0.88	0.91	0.93	0.88	0.91	0.89
	Med	**0.78**	**0.51**	**0.80**	**0.51**	**0.72**	**0.81**	**0.64**	**0.75**	**0.53**
	Min	0.55	-0.11	0.56	-0.08	0.23	0.50	-0.04	0.36	0.09

K-medoids

		Clusters								
		Two		**Three**			**Four**			
		A	B	A	B	C	A	B	C	D
Euc &Std	Max	1.00	0.69	0.62	1.00	0.81	0.62	1.00	0.79	1.00
	Med	**0.39**	**0.42**	**0.36**	**0.41**	**0.56**	**0.37**	**0.45**	**0.52**	**0.68**
	Min	-0.27	-0.17	-0.09	0.04	-0.14	-0.09	0.04	-0.14	0.07
Euc & Mad	Max	1.00	0.69	1.00	1.00	0.81	1.00	1.00	0.79	1.00
	Med	**0.39**	**0.42**	**0.67**	**0.42**	**0.57**	**0.67**	**0.45**	**0.55**	**0.70**
	Min	-0.27	-0.17	0.12	0.04	-0.14	0.23	0.04	-0.14	0.02
Man & Std	Max	1.00	0.77	1.00	0.77	0.85	0.98	0.54	0.71	0.72
	Med	**0.55**	**0.47**	**0.61**	**0.51**	**0.38**	**0.56**	**0.39**	**0.42**	**0.51**
	Min	-0.14	-0.24	-0.14	-0.24	-0.17	-0.14	-0.05	-0.27	0.02
Man & Mad	Max	0.67	0.77	1.00	0.67	0.81	1.00	0.67	0.79	1.00
	Med	**0.37**	**0.38**	**0.71**	**0.38**	**0.55**	**0.71**	**0.39**	**0.50**	**0.69**
	Min	-0.09	-0.27	0.09	-0.27	-0.14	0.09	-0.27	-0.14	0.23

Key

Distance Metric	**Standardization**
Euc = Euclidean Man = Manhattan	Std = Standard Deviation Mad = Mean Absolute Deviation

Appendix – Event Driven

Figure 4: Common Funds Across Clusters Irrespective of the Distance Metric

K-means

Standardization = Standard Deviation

Standardization = Mean Absolute Deviation

Two Clusters

		Manhattan		**Total**			Manhattan		**Total**
		A	**B**				**A**	**B**	
Euclidean	**A**	101	20	**121**		**A**	93	23	**116**
	B	1	59	**60**		**B**	0	65	**65**
	Total	**102**	**79**	**181**		**Total**	**93**	**88**	**181**

Three Clusters

		Manhattan			**Total**			Manhattan			**Total**
		A	**B**	**C**				**A**	**B**	**C**	
Euclidean	**A**	63	8	1	**72**		**A**	62	8	1	**71**
	B	0	56	0	**56**		**B**	0	57	0	**57**
	C	2	1	50	**53**		**C**	1	4	48	**53**
	Total	**65**	**65**	**51**	**181**		**Total**	**63**	**69**	**49**	**181**

Four Clusters

		Manhattan				**Total**			Manhattan				**Total**
		A	**B**	**C**	**D**				**A**	**B**	**C**	**D**	
Euclidean	**A**	58	1	0	5	**64**		**A**	58	3	0	13	**74**
	B	0	30	0	2	**32**		**B**	0	26	0	2	**28**
	C	7	1	43	3	**54**		**C**	2	0	41	6	**49**
	D	0	1	0	30	**31**		**D**	0	16	0	14	**30**
	Total	**65**	**33**	**43**	**40**	**181**		**Total**	**60**	**45**	**41**	**35**	**181**

Appendix – Event Driven

K-medoids

Standardization = Standard Deviation

Standardization = Mean Absolute Deviation

Two Clusters

		Manhattan					Manhattan		
		A	**B**	**Total**			**A**	**B**	**Total**
Euclidean	**A**	73	33	**106**		**A**	91	14	**105**
	B	44	31	**75**		**B**	7	69	**76**
	Total	**117**	**64**	**181**		**Total**	**98**	**83**	**181**

Three Clusters

		Manhattan						Manhattan			
		A	**B**	**C**	**Total**			**A**	**B**	**C**	**Total**
Euclidean	**A**	26	33	8	**67**		**A**	56	12	5	**73**
	B	13	5	25	**43**		**B**	1	29	11	**41**
	C	45	8	18	**71**		**C**	7	17	43	**67**
	Total	**84**	**46**	**51**	**181**		**Total**	**64**	**58**	**59**	**181**

Four Clusters

		Manhattan							Manhattan				
		A	**B**	**C**	**D**	**Total**			**A**	**B**	**C**	**D**	**Total**
Euclidean	**A**	5	26	10	19	**60**		**A**	47	10	4	2	**63**
	B	9	5	24	1	**39**		**B**	1	25	8	2	**36**
	C	27	4	15	6	**52**		**C**	5	14	29	1	**49**
	D	11	13	5	1	**30**		**D**	5	5	1	22	**33**
	Total	**52**	**48**	**54**	**27**	**181**		**Total**	**58**	**54**	**42**	**27**	**181**

Appendix – Event Driven

Figure 5 Median Correlations of Common Funds in Figure 4 with Broad Benchmark

K-means

Standardization = Standard Deviation | Standardization = Mean Absolute Deviation

Two Clusters

		Manhattan					Manhattan		
		A	**B**	**Total**			**A**	**B**	**Total**
Euclidean	**A**	0.74	0.44	**0.71**		**A**	0.76	0.55	**0.71**
	B	0.66	0.26	**0.26**		**B**	-	0.27	**0.27**
	Total	**0.74**	**0.36**	**0.61**		**Total**	**0.76**	**0.39**	**0.61**

Three Clusters

		Manhattan						Manhattan			
		A	**B**	**C**	**Total**			**A**	**B**	**C**	**Total**
Euclidean	**A**	0.78	0.58	0.66	**0.76**		**A**	0.78	0.60	0.66	**0.76**
	B	-	0.27	-	**0.27**		**B**	-	0.27	-	**0.27**
	C	0.79	0.40	0.62	**0.62**		**C**	0.80	0.40	0.63	**0.62**
	Total	**0.78**	**0.34**	**0.62**	**0.61**		**Total**	**0.78**	**0.35**	**0.64**	**0.61**

Four Clusters

		Manhattan							Manhattan				
		A	**B**	**C**	**D**	**Total**			**A**	**B**	**C**	**D**	**Total**
Euclidean	**A**	0.79	0.56	-	0.58	**0.77**		**A**	0.79	0.59	-	0.63	**0.75**
	B	-	0.43	-	0.38	**0.43**		**B**	-	0.43	-	0.38	**0.43**
	C	0.74	0.77	0.66	0.47	**0.66**		**C**	0.83	-	0.66	0.44	**0.64**
	D	-	0.19	-	0.19	**0.19**		**D**	-	0.16	-	0.22	**0.18**
	Total	**0.78**	**0.44**	**0.66**	**0.26**	**0.61**		**Total**	**0.79**	**0.28**	**0.66**	**0.40**	**0.61**

Appendix – Event Driven

K-medoids

Standardization = Standard Deviation

Standardization = Mean Absolute Deviation

Two Clusters

		Manhattan		
		A	**B**	**Total**
Euclidean	**A**	0.72	0.66	**0.70**
	B	0.50	0.41	**0.49**
	Total	**0.65**	**0.58**	**0.61**

	Manhattan		
	A	**B**	**Total**
A	0.71	0.40	**0.70**
B	0.60	0.44	**0.49**
Total	**0.70**	**0.42**	**0.61**

Three Clusters

		Manhattan			
		A	**B**	**C**	**Total**
Euclidean	**A**	0.79	0.63	0.60	**0.67**
	B	0.46	0.33	0.25	**0.33**
	C	0.70	0.73	0.60	**0.68**
	Total	**0.70**	**0.61**	**0.42**	**0.61**

	Manhattan			
	A	**B**	**C**	**Total**
A	0.70	0.49	0.60	**0.67**
B	0.29	0.27	0.47	**0.29**
C	0.67	0.55	0.72	**0.68**
Total	**0.69**	**0.40**	**0.67**	**0.61**

Four Clusters

		Manhattan				
		A	**B**	**C**	**D**	**Total**
Euclidean	**A**	0.79	0.68	0.64	0.77	**0.67**
	B	0.46	0.29	0.31	0.33	**0.33**
	C	0.67	0.37	0.66	0.70	**0.67**
	D	0.78	0.72	0.07	0.34	**0.68**
	Total	**0.67**	**0.66**	**0.51**	**0.70**	**0.61**

	Manhattan				
	A	**B**	**C**	**D**	**Total**
A	0.68	0.57	0.65	0.56	**0.66**
B	0.29	0.27	0.47	0.38	**0.29**
C	0.67	0.65	0.69	0.61	**0.67**
D	0.78	0.26	0.52	0.72	**0.70**
Total	**0.68**	**0.41**	**0.63**	**0.70**	**0.61**

Appendix – Event Driven

Figure 6: Common Funds Across Clusters Irrespective of the Clustering Method

Euclidean Distance Standard Deviation | **Euclidean Distance Mean Absolute Deviation** | **Manhattan Distance Standard Deviation** | **Manhattan Distance Mean Absolute Deviation**

Two Clusters

K-means

		A	B	Total		A	B	Total		A	B	Total		A	B	Total
K-medoids	**A**	93	13	**106**	**A**	90	15	**105**	**A**	66	51	**117**	**A**	75	23	**98**
	B	28	47	**75**	**B**	26	50	**76**	**B**	36	28	**64**	**B**	18	65	**83**
	Total	**121**	**60**	**181**	**Total**	**116**	**65**	**181**	**Total**	**102**	**79**	**181**	**Total**	**93**	**88**	**181**

Three Clusters

K-means

		A	B	C	Total		A	B	C	Total		A	B	C	Total		A	B	C	Total
K-medoids	**A**	27	2	38	**67**	**A**	28	4	41	**73**	**A**	56	21	7	**84**	**A**	27	5	32	**64**
	B	2	40	1	**43**	**B**	2	39	0	**41**	**B**	3	6	37	**46**	**B**	6	42	10	**58**
	C	43	14	14	**71**	**C**	41	14	12	**67**	**C**	6	38	7	**51**	**C**	30	22	7	**59**
	Total	**72**	**56**	**53**	**181**	**Total**	**71**	**57**	**53**	**181**	**Total**	**65**	**65**	**51**	**181**	**Total**	**63**	**69**	**49**	**181**

Four Clusters

K-means

		A	B	C	D	Total		A	B	C	D	Total		A	B	C	D	Total		A	B	C	D	Total
K-medoids	**A**	22	0	34	4	**60**	**A**	21	2	40	0	**63**	**A**	30	8	3	11	**52**	**A**	23	0	25	10	**58**
	B	1	27	1	10	**39**	**B**	2	9	0	25	**36**	**B**	16	2	22	8	**48**	**B**	6	34	8	6	**54**
	C	22	3	15	12	**52**	**C**	27	13	7	2	**49**	**C**	7	23	10	14	**54**	**C**	16	7	7	12	**42**
	D	19	2	4	5	**30**	**D**	24	6	2	1	**33**	**D**	12	0	8	7	**27**	**D**	15	4	1	7	**27**
	Total	**64**	**32**	**54**	**31**	**181**	**Total**	**74**	**30**	**49**	**28**	**181**	**Total**	**65**	**33**	**43**	**40**	**181**	**Total**	**60**	**45**	**41**	**35**	**181**

Appendix – Merger Arbitrage

Figure 1: Cluster Size and Fund Transitions Across Clusters as a Function of the Number of Clusters
(Totals in Bold)

K-means

	Clusters								
	Two		Three			Four			
	A	B	A	B	C	A	B	C	D
Euc & Std	**59**	**57**	44A = **44**	7A+38B = **45**	8A+19C = **27**	41A+4B = **45**	39B+1C = **40**	3A+2B+19C = **24**	7C = **7**
Euc & Mad	**59**	**57**	45A = **45**	7A+38B = **45**	7A+19C = **26**	43A+2B = **45**	41B = **41**	2A+2B+11C = **15**	15C = **15**
Man & Std	**59**	**57**	46A+3B = **49**	36B = **36**	13A+18B = **31**	42A+1C = **43**	27B = **27**	5A+1B+23C = **29**	2A+8B+7C = **17**
Man & Mad	**59**	**57**	47A+3B = **50**	37B = **37**	12A+17B = **29**	40A+1C = **41**	2A+25B = **27**	5A+2B+20C = **27**	3A+10B+8C = **21**

K-medoids

	Clusters								
	Two		Three			Four			
	A	B	A	B	C	A	B	C	D
Euc & Std	**64**	**52**	47A = **47**	48B = **48**	17A+4B = **21**	43A = **43**	45B = **45**	14C = **14**	4A+3B+7C = **14**
Euc & Mad	**61**	**55**	36A = **36**	2A+47B = **49**	23A+8B = **31**	27A+1B+10C = **38**	43B+2C = **45**	1A+4B+15C = **20**	8A+1B+4C = **13**
Man & Std	**64**	**52**	45A+1B = **46**	2A+47B = **49**	17A+4B = **21**	32A+1B = **33**	1A+44B+2C = **47**	1B+14C = **15**	13A+3B+5C = **21**
Man & Mad	**59**	**57**	40A = **40**	44B = **44**	19A+13B = **32**	20A+1B+6C = **27**	1A+38B+7C = **46**	6A+12C = **18**	13A+5B+7C = **25**

Key

Distance Metric	**Standardization**
Euc = Euclidean Man = Manhattan	Std = Standard Deviation Mad = Mean Absolute Deviation

Appendix – Merger Arbitrage

Figure 2: Correlation Across Clusters (Centroids)

K-means

	Clusters									
	Two	**Three**			**Four**					
	A&B	A&B	A&C	B&C	A&B	A&C	A&D	B&C	B&D	C&D
Euc & Std	0.68	0.72	0.38	0.57	0.70	0.57	-0.44	0.65	-0.22	-0.37
Euc & Mad	0.68	0.72	0.37	0.56	0.70	0.59	-0.05	0.59	0.34	0.21
Man & Std	0.73	0.74	0.55	0.65	0.57	0.67	0.65	0.84	0.58	0.53
Man & Mad	0.71	0.73	0.50	0.63	0.70	0.28	0.79	0.52	0.73	0.41

K-medoids

	Clusters									
	Two	**Three**			**Four**					
	A&B	A&B	A&C	B&C	A&B	A&C	A&D	B&C	B&D	C&D
Euc & Std	0.65	0.65	0.57	0.84	0.65	0.57	0.26	0.84	0.42	0.23
Euc & Mad	0.83	0.74	0.47	0.31	0.83	0.50	0.17	0.52	0.21	0.02
Man & Std	0.68	0.70	0.17	0.21	0.68	0.17	0.81	-0.10	0.58	0.13
Man & Mad	0.74	0.83	0.79	0.78	0.68	0.17	0.56	-0.10	0.31	0.19

Key

Distance Metric	**Standardization**
Euc = Euclidean	Std = Standard Deviation
Man = Manhattan	Mad = Mean Absolute Deviation

Appendix – Merger Arbitrage

Figure 3: Correlation Ranges Within Clusters (Among Managers)

K-means

		Clusters								
		Two		**Three**			**Four**			
		A	B	A	B	C	A	B	C	D
Euc & Std	Max	0.95	0.94	0.97	0.96	0.71	0.95	0.97	0.78	0.58
	Med	**0.70**	**0.72**	**0.79**	**0.73**	**0.48**	**0.74**	**0.82**	**0.64**	**0.55**
	Min	0.38	-0.57	0.54	0.31	-0.33	0.34	0.52	0.28	0.24
Euc & Mad	Max	0.94	0.94	0.97	0.96	0.00	0.95	0.97	0.86	0.58
	Med	**0.69**	**0.71**	**0.79**	**0.73**	**0.72**	**0.73**	**0.83**	**0.74**	**0.39**
	Min	0.37	-0.58	0.54	0.33	0.46	0.35	0.54	0.49	0.02
Man & Std	Max	0.95	0.97	0.95	0.98	0.97	0.96	0.86	0.98	0.90
	Med	**0.67**	**0.73**	**0.74**	**0.82**	**0.82**	**0.79**	**0.48**	**0.86**	**0.69**
	Min	-0.73	-0.08	0.37	0.18	0.61	0.44	-0.52	0.62	0.23
Man & Mad	Max	0.95	0.96	0.96	0.86	0.73	0.95	0.97	0.68	0.80
	Med	**0.67**	**0.73**	**0.72**	**0.49**	**0.40**	**0.83**	**0.83**	**0.40**	**0.65**
	Min	-0.72	-0.08	0.43	-0.52	-0.41	0.44	0.58	-0.23	0.26

K-medoids

		Clusters								
		Two		**Three**			**Four**			
		A	B	A	B	C	A	B	C	D
Euc &Std	Max	0.80	0.77	0.80	0.77	1.00	0.80	0.77	1.00	1.00
	Med	**0.48**	**0.50**	**0.53**	**0.50**	**0.36**	**0.55**	**0.50**	**0.51**	**0.52**
	Min	-0.61	-0.09	0.10	-0.09	-0.52	0.10	-0.09	0.17	-0.27
Euc & Mad	Max	1.00	0.77	1.00	0.72	0.60	1.00	0.77	1.00	0.45
	Med	**0.70**	**0.50**	**0.80**	**0.49**	**0.22**	**0.75**	**0.54**	**0.52**	**0.21**
	Min	-0.61	-0.09	-0.04	0.12	-0.47	0.20	-0.09	-0.45	0.00
Man & Std	Max	0.78	0.72	0.81	0.77	0.49	0.78	0.72	0.45	0.79
	Med	**0.51**	**0.48**	**0.57**	**0.50**	**0.16**	**0.62**	**0.48**	**0.21**	**0.38**
	Min	-0.75	0.08	-0.10	0.05	-0.27	-0.10	0.08	-0.27	0.10
Man & Mad	Max	1.00	0.72	1.00	0.76	1.00	0.78	0.72	0.49	0.48
	Med	**0.71**	**0.47**	**0.73**	**0.51**	**0.46**	**0.69**	**0.50**	**0.19**	**0.29**
	Min	-0.61	-0.06	0.05	0.05	-0.48	0.39	0.16	-0.27	-0.02

Key

Distance Metric	**Standardization**
Euc = Euclidean Man = Manhattan	Std = Standard Deviation Mad = Mean Absolute Deviation

Appendix – Merger Arbitrage

Figure 4: Common Funds Across Clusters Irrespective of the Distance Metric

K-means

Standardization = Standard Deviation

Standardization = Mean Absolute Deviation

Two Clusters

Standardization = Standard Deviation

		Manhattan		
		A	**B**	**Total**
Euclidean	**A**	53	6	**59**
	B	6	51	**57**
	Total	**59**	**57**	**116**

Standardization = Mean Absolute Deviation

	Manhattan		
	A	**B**	**Total**
A	53	6	**59**
B	6	51	**57**
Total	**59**	**57**	**116**

Three Clusters

Standardization = Standard Deviation

		Manhattan			
		A	**B**	**C**	**Total**
	A	41	0	3	**44**
Euclidean	**B**	6	36	3	**45**
	C	2	0	25	**27**
	Total	**49**	**36**	**31**	**116**

Standardization = Mean Absolute Deviation

	Manhattan			
	A	**B**	**C**	**Total**
A	43	0	2	**45**
B	6	37	2	**45**
C	1	0	25	**26**
Total	**50**	**37**	**29**	**116**

Four Clusters

Standardization = Standard Deviation

		Manhattan				
		A	**B**	**C**	**D**	**Total**
	A	42	0	2	1	**45**
Euclidean	**B**	0	27	2	11	**40**
	C	1	0	20	3	**24**
	D	0	0	5	2	**7**
	Total	**43**	**27**	**29**	**17**	**116**

Standardization = Mean Absolute Deviation

	Manhattan				
	A	**B**	**C**	**D**	**Total**
A	39	0	4	2	**45**
B	1	27	1	12	**41**
C	1	0	14	0	**15**
D	0	0	8	7	**15**
Total	**41**	**27**	**27**	**21**	**116**

Appendix – Merger Arbitrage

K-medoids

Standardization = Standard Deviation | Standardization = Mean Absolute Deviation

Two Clusters

		Manhattan					Manhattan		
		A	**B**	**Total**			**A**	**B**	**Total**
Euclidean	**A**	60	4	**64**		**A**	57	4	**61**
	B	4	48	**52**		**B**	2	53	**55**
	Total	**64**	**52**	**116**		**Total**	**59**	**57**	**116**

Three Clusters

		Manhattan						Manhattan			
		A	**B**	**C**	**Total**			**A**	**B**	**C**	**Total**
Euclidean	**A**	38	2	7	**47**		**A**	27	0	9	**36**
	B	2	44	2	**48**		**B**	0	41	8	**49**
	C	6	3	12	**21**		**C**	13	3	15	**31**
	Total	**46**	**49**	**21**	**116**		**Total**	**40**	**44**	**32**	**116**

Four Clusters

		Manhattan							Manhattan				
		A	**B**	**C**	**D**	**Total**			**A**	**B**	**C**	**D**	**Total**
Euclidean	**A**	25	0	2	16	**43**		**A**	21	2	4	11	**38**
	B	1	42	1	1	**45**		**B**	2	41	1	1	**45**
	C	1	2	9	2	**14**		**C**	2	3	2	13	**20**
	D	6	3	3	2	**14**		**D**	2	0	11	0	**13**
	Total	**33**	**47**	**15**	**21**	**116**		**Total**	**27**	**46**	**18**	**25**	**116**

Appendix – Merger Arbitrage

Figure 5 Median Correlations of Common Funds in Figure 4 with Broad Benchmark

K-means

Standardization = Standard Deviation

Standardization = Mean Absolute Deviation

Two Clusters

		Manhattan		
		A	**B**	**Total**
Euclidean	**A**	0.60	0.65	**0.60**
	B	-0.01	0.54	**0.52**
	Total	**0.60**	**0.56**	**0.57**

	Manhattan		
	A	**B**	**Total**
A	0.60	0.65	**0.60**
B	-0.01	0.54	**0.52**
Total	**0.60**	**0.56**	**0.57**

Three Clusters

		Manhattan			
		A	**B**	**C**	**Total**
Euclidean	**A**	0.61	-	0.37	**0.61**
	B	0.61	0.60	0.56	**0.61**
	C	0.43	-	0.23	**0.24**
	Total	**0.61**	**0.60**	**0.28**	**0.57**

	Manhattan			
	A	**B**	**C**	**Total**
A	0.61	-	0.39	**0.60**
B	0.61	0.61	0.44	**0.61**
C	0.46	-	0.23	**0.23**
Total	**0.61**	**0.61**	**0.26**	**0.57**

Four Clusters

		Manhattan				
		A	**B**	**C**	**D**	**Total**
Euclidean	**A**	0.61	-	0.35	0.61	**0.61**
	B	-	0.61	0.38	0.56	**0.59**
	C	0.60	-	0.43	0.23	**0.43**
	D	-	-	-0.21	0.01	**-0.05**
	Total	**0.61**	**0.61**	**0.34**	**0.52**	**0.57**

	Manhattan				
	A	**B**	**C**	**D**	**Total**
A	0.61	-	0.51	0.63	**0.61**
B	0.44	0.62	0.33	0.54	**0.61**
C	0.60	-	0.48	-	**0.50**
D	-	-	-0.02	0.18	**0.03**
Total	**0.60**	**0.62**	**0.37**	**0.52**	**0.57**

Appendix – Merger Arbitrage

K-medoids

Standardization = Standard Deviation | Standardization = Mean Absolute Deviation

Two Clusters

		Manhattan					Manhattan		
		A	**B**	**Total**			**A**	**B**	**Total**
Euclidean	**A**	0.56	0.57	**0.56**		**A**	0.54	0.61	**0.56**
	B	0.40	0.60	**0.59**		**B**	0.53	0.60	**0.60**
	Total	**0.54**	**0.60**	**0.57**		**Total**	**0.54**	**0.60**	**0.57**

Three Clusters

		Manhattan						Manhattan			
		A	**B**	**C**	**Total**			**A**	**B**	**C**	**Total**
Euclidean	**A**	0.59	0.54	0.34	**0.57**		**A**	0.59	-	0.68	**0.61**
	B	0.51	0.60	0.25	**0.59**		**B**	-	0.60	0.62	**0.60**
	C	0.18	0.39	0.55	**0.39**		**C**	0.26	0.53	0.28	**0.33**
	Total	**0.58**	**0.60**	**0.47**	**0.57**		**Total**	**0.55**	**0.60**	**0.58**	**0.57**

Four Clusters

		Manhattan							Manhattan				
		A	**B**	**C**	**D**	**Total**			**A**	**B**	**C**	**D**	**Total**
Euclidean	**A**	0.61	-	0.62	0.46	**0.58**		**A**	0.63	0.58	0.24	0.40	**0.57**
	B	0.44	0.60	-0.05	0.37	**0.60**		**B**	0.53	0.60	-0.05	0.42	**0.60**
	C	0.18	0.36	0.58	0.40	**0.55**		**C**	0.65	0.39	-0.58	0.46	**0.43**
	D	0.51	0.46	-0.57	0.45	**0.48**		**D**	0.55	-	0.58	-	**0.58**
	Total	**0.59**	**0.60**	**0.52**	**0.44**	**0.57**		**Total**	**0.62**	**0.60**	**0.41**	**0.42**	**0.57**

Appendix – Merger Arbitrage

Figure 6: Common Funds Across Clusters Irrespective of the Clustering Method

	Euclidean Distance Standard Deviation				Euclidean Distance Mean Absolute Deviation				Manhattan Distance Standard Deviation				Manhattan Distance Mean Absolute Deviation			

Two Clusters

K-means

K-medoids		A	B	Total		A	B	Total		A	B	Total		A	B	Total
	A	10	54	64	**A**	7	54	61	**A**	9	55	64	**A**	5	54	59
	B	49	3	52	**B**	52	3	55	**B**	50	2	52	**B**	54	3	57
	Total	59	57	116	**Total**	59	57	116	**Total**	59	57	116	**Total**	59	57	116

Three Clusters

K-means

K-medoids		A	B	C	Total		A	B	C	Total		A	B	C	Total		A	B	C	Total
	A	1	38	8	47	**A**	0	34	2	36	**A**	7	29	10	46	**A**	3	26	11	40
	B	41	1	6	48	**B**	44	2	3	49	**B**	40	0	9	49	**B**	40	0	4	44
	C	2	6	13	21	**C**	1	9	21	31	**C**	2	7	12	21	**C**	7	11	14	32
	Total	44	45	27	116	**Total**	45	45	26	116	**Total**	49	36	31	116	**Total**	50	37	29	116

Four Clusters

K-means

K-medoids		A	B	C	D	Total		A	B	C	D	Total		A	B	C	D	Total		A	B	C	D	Total
	A	5	34	1	3	43	**A**	2	31	4	1	38	**A**	3	19	9	2	33	**A**	0	22	2	3	27
	B	39	1	4	1	45	**B**	41	1	1	2	45	**B**	38	0	9	0	47	**B**	39	0	5	2	46
	C	1	5	7	1	14	**C**	2	1	5	12	20	**C**	0	1	7	7	15	**C**	0	1	5	12	18
	D	0	0	12	2	14	**D**	0	8	5	0	13	**D**	2	7	4	8	21	**D**	2	4	15	4	25
	Total	45	40	24	7	116	**Total**	45	41	15	15	116	**Total**	43	27	29	17	116	**Total**	41	27	27	21	116