Risk profile clustering strategy in portfolio diversification

Cathy Yi-Hsuan Chen
Wolfgang Karl Härdle
Alla Petukhina

Ladislaus von Bortkiewicz Chair of Statistics
Humboldt-Universität zu Berlin
lvb.wiwi.hu-berlin.de

Diversification

It is the part of a wise man to keep himself today for tomorrow, and not venture all his eggs in one basket (Don Quixote, M. Servantes)

OR

Put all your eggs in the one basket and WATCH THAT BASKET (Pudd'nhead Wilson, M. Twain)?

TEDAS with **Y**= S&P 500

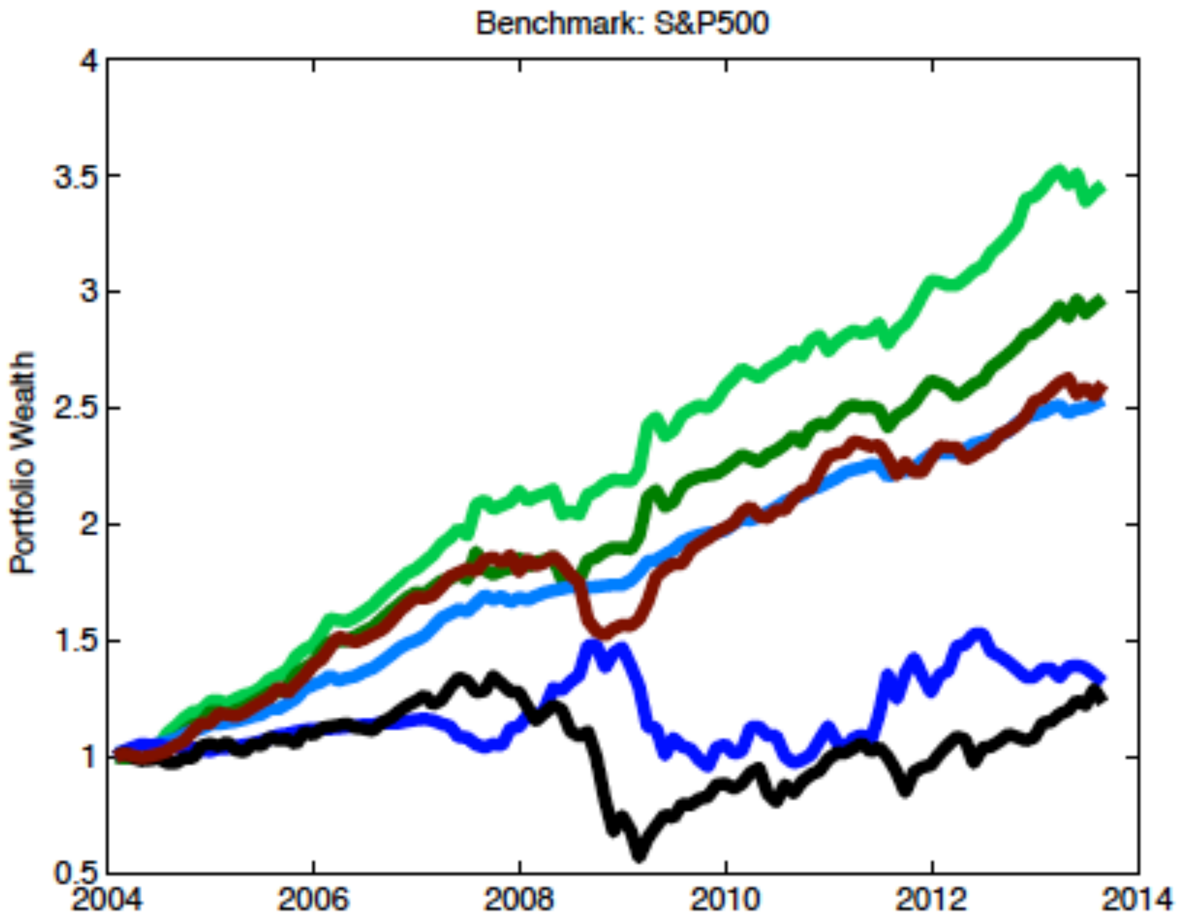


Figure 1. Cumulative portfolio wealth comparison: TEDAS 1 , TEDAS 3 , TEDAS 2 , RR , PESS , S&P 500 buy & hold; X = hedge funds' indices' returns matrix

TEDAS_strategies2

Challenges

- Risk-management challenges
  - Asset classes
  - Choice of risk measure
  - Liquidity issue
- Statistical challenges
  - Large assets' universe
  - Assets clustering

Objectives

- Improvement of portfolio diversification
- Risk-profile based consensus-way to detect assets' classes

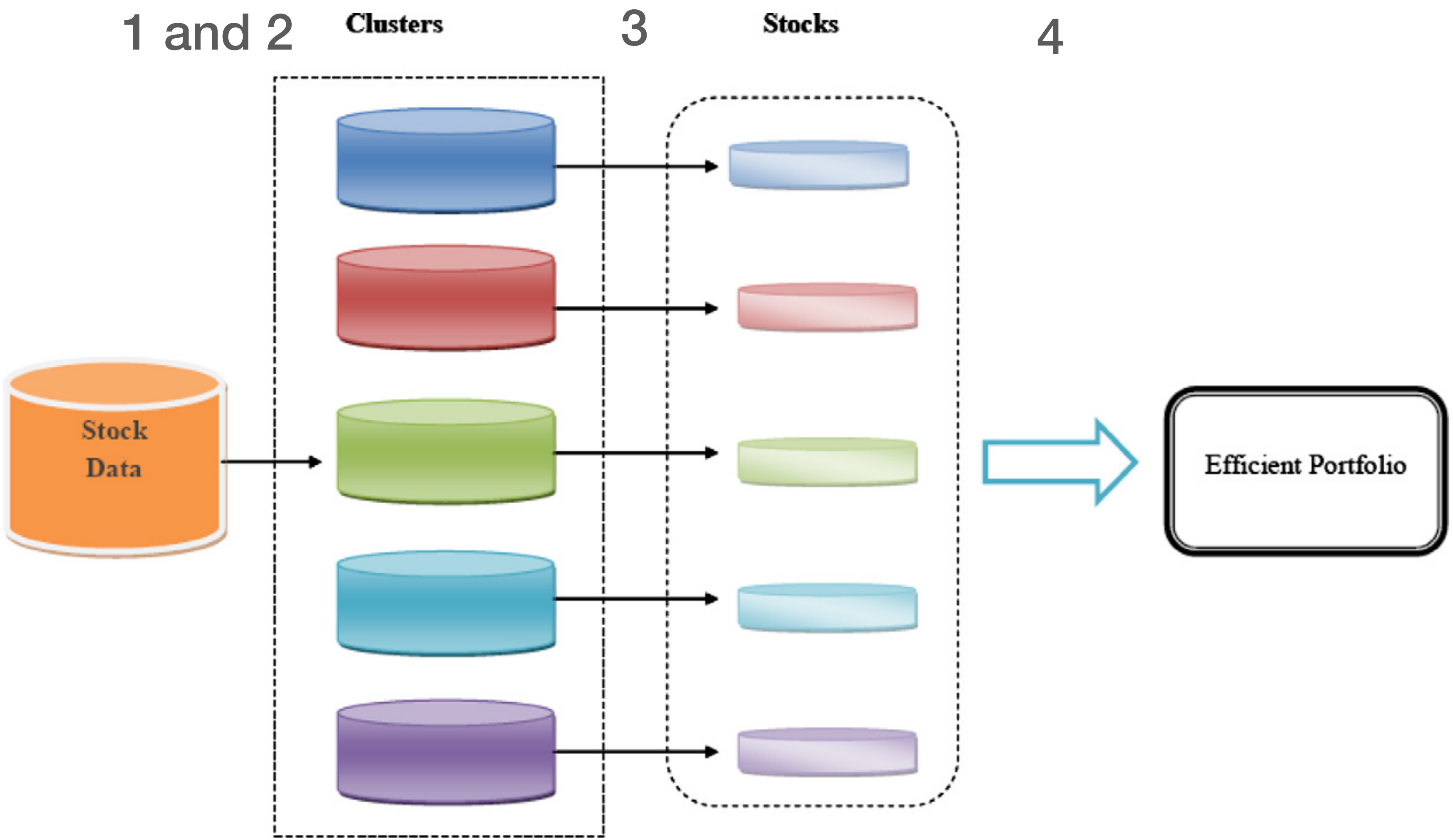
Outline

1. Motivation ✓
2. Methodology
3. Data
4. Empirical Results
5. Outlook
6. Technical details

Methodology

1. Construct risk profiles of assets (based on annual data)
   - CAPM β
   - Volatility
   - Skewness
   - Kurtosis
   - Value-at-Risk 5% ▸ Details
   - Expected Shortfall 5% ▸ Details

Portfolio construction



Methodology

2. Cluster the assets (2-15 clusters)
   - Partitioning algorithms
     - k-means ▸ Details
     - FUZZY C-means ▸ Details
     - C-Medoids ▸ Details
   - Hierarchical algorithms ▸ Details
     - Agglomerative hierarchical clustering

3. Choose portfolio constituents from every cluster
   - Maximum Sharpe ratio
   - Random selection

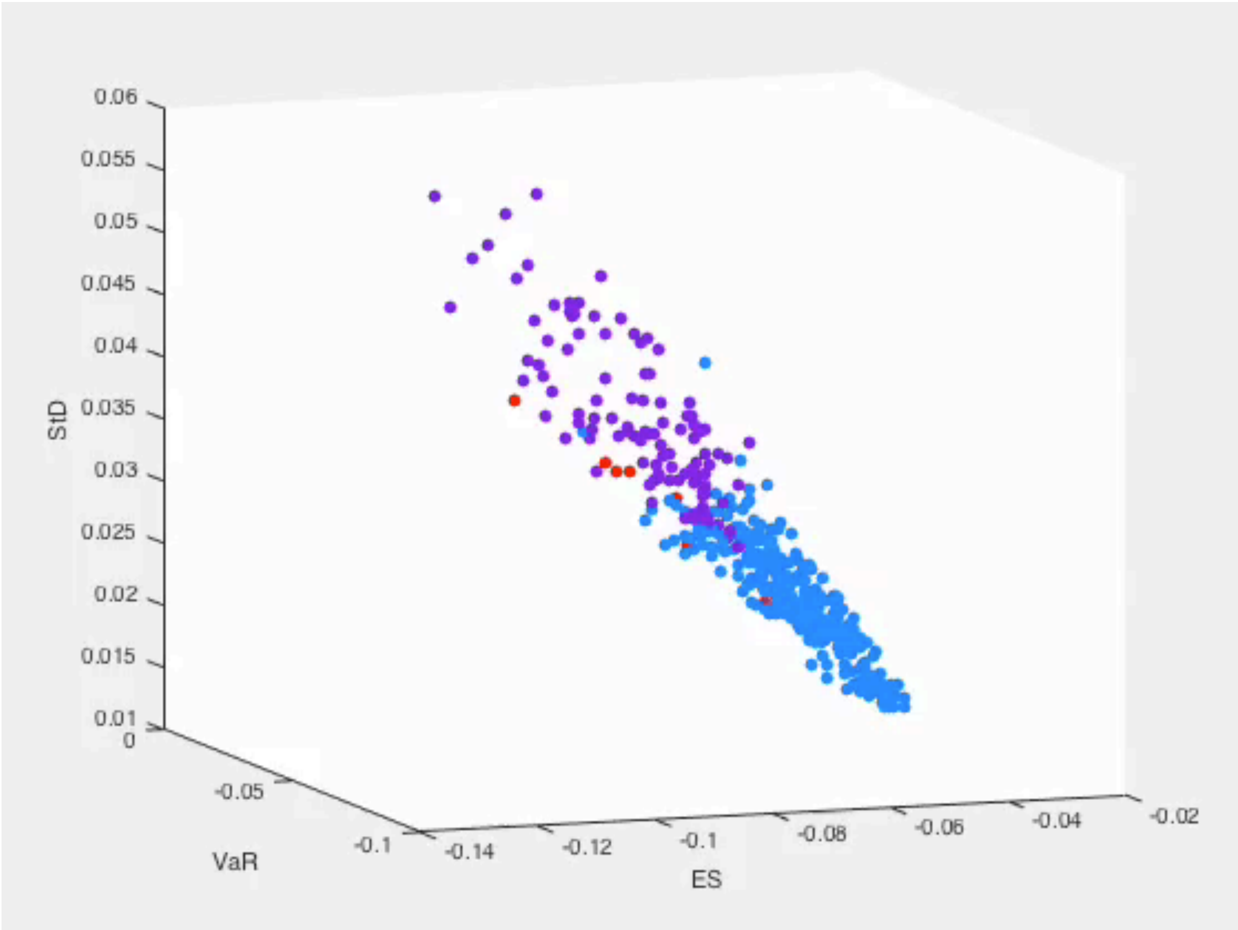
Methodology

4. Portfolio allocation
   - *1/n* rule
   - Mean-variance portfolios (Markowitz rule) Details
5. Rebalancing of portfolios
   - Every period t based on t -*1* clusters-detection and covariance matrix
   - Transaction costs are 1% of portfolio value

North American equity

- Daily data
  - STOXX North America 600 index
  - 435 - 593 stocks from STOXX North America 600 index as on 20160101
- Span: 19980101 - 20151231 (18 years)
- Source: Datastream

Risk profile communities: 3 agglomerative hierarchical clusters



Portfolios' performance

Number of clusters	k-means	Fuzzy C-means	C-medoids	Hierarchical clustering
2	3.8488	3.6062	7.0494	4.7722
3	2.1883	0.7723	3.8328	4.8754
5	2.3127	2.1768	3.1574	32.2337
7	3.3336	1.3769	4.5291	11.9582
9	9.7372	2.5469	2.9247	8.6865
11	3.4709	3.1204	2.5753	4.6962
13	2.9199	2.7403	2.4043	4.7678
15	4.2518	3.1997	2.4624	4.8438

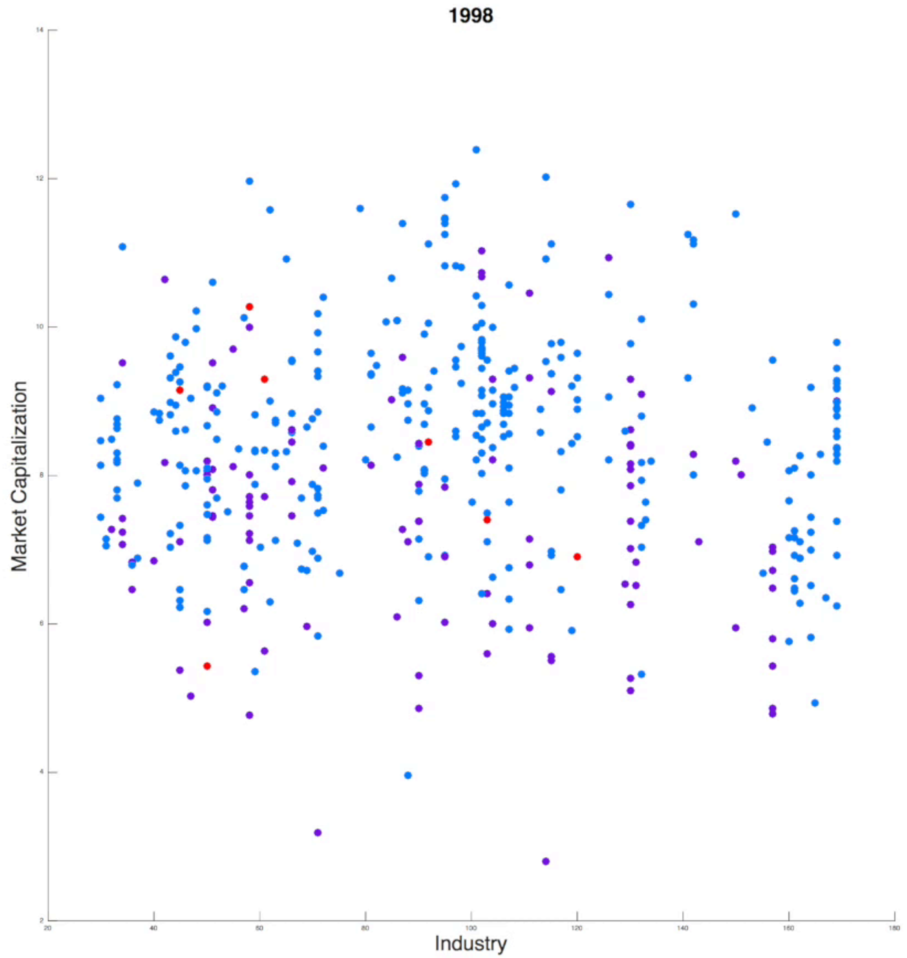
Table 1. 1/n portfolios cumulative return

Portfolios’ performance

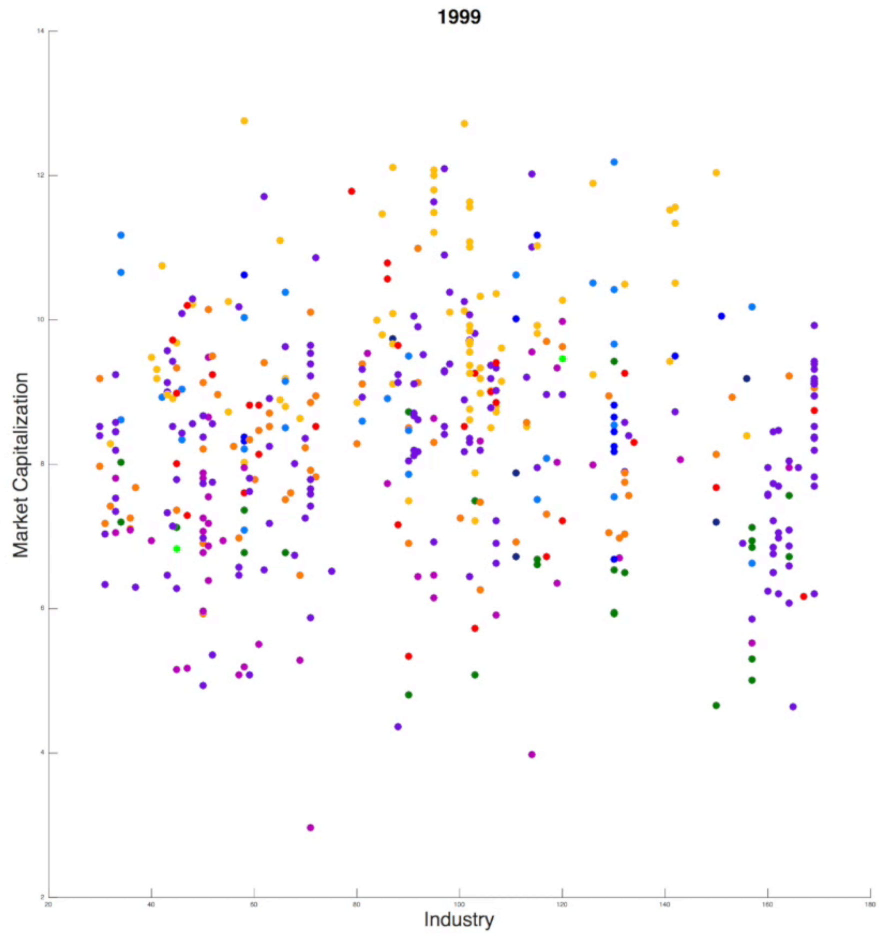
Number of clusters	k-means	Fuzzy C-means	C-medoids	Hierarchical clustering
2	1.1485	1.3903	0.9618	1.1857
3	1.5374	1.2923	2.9060	1.5435
5	4.9130	5.5700	1.2583	9.6457
7	17.4741	13.1557	1.7954	34.9652
9	1.7550	3.1684	2.6919	1.4887
11	1.2773	1.5133	1.9223	1.3254
13	2.2544	2.6394	2.9076	4.1655
15	3.4324	5.9887	10.3958	3.2033

Table 2. Markowitz-portfolios cumulative return

Risk profile communities vs size and industry



Risk profile communities vs size and industry



k-means clusters' portfolios

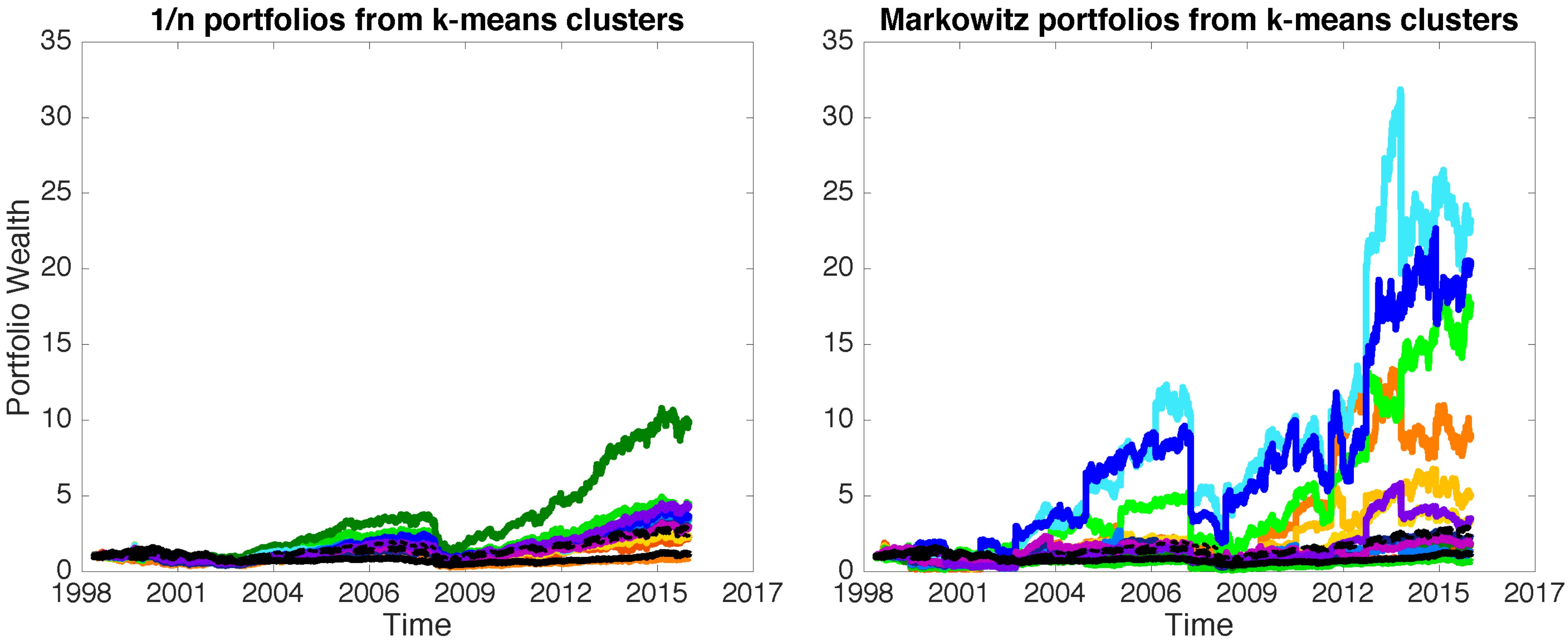


Figure 1. Cumulative portfolio wealth comparison (Distance measure: squared Euclidean): Black - Buy&hold STOXX600 NA(solid), Markowitz (dashed), 1/n (doted)

FUZZY C-means clusters' portfolios

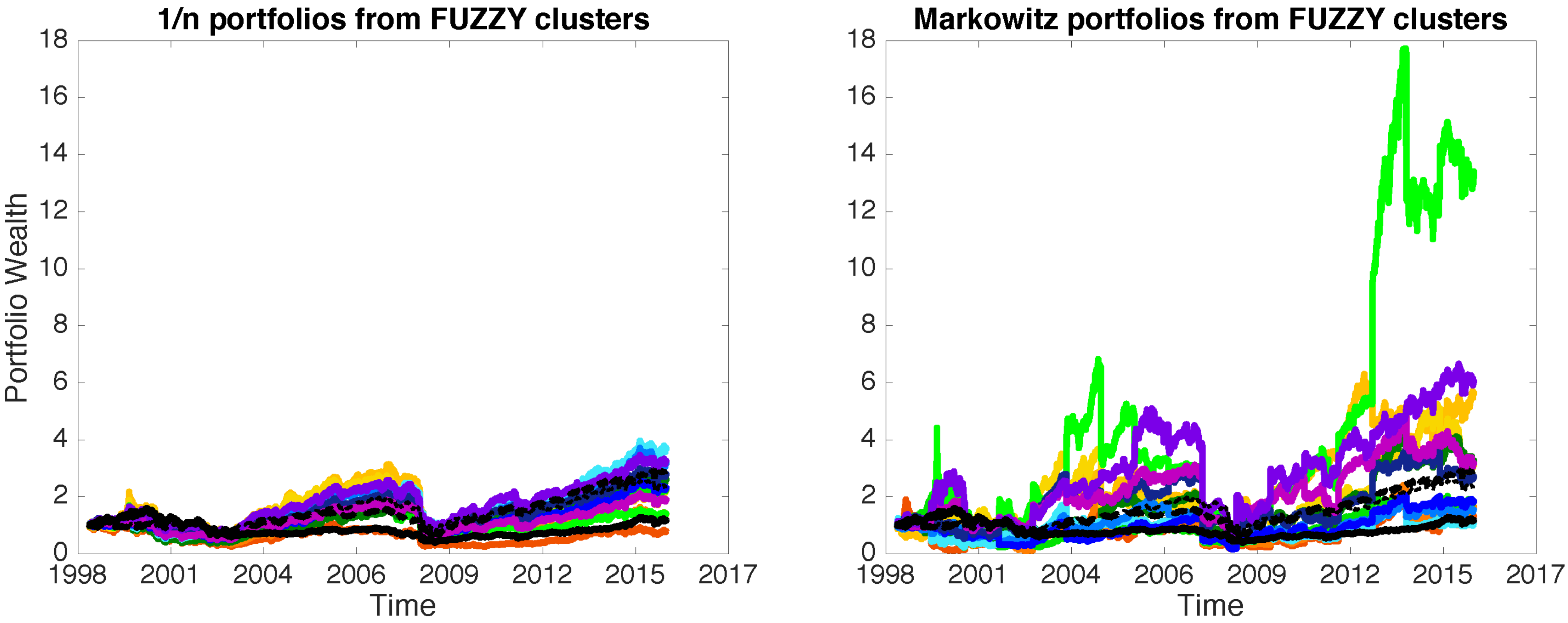


Figure 2. Cumulative portfolio wealth comparison: Black - Buy&hold STOXX600 NA(solid), Markowitz (dashed), 1/n (doted)

C - medoids clusters' portfolios

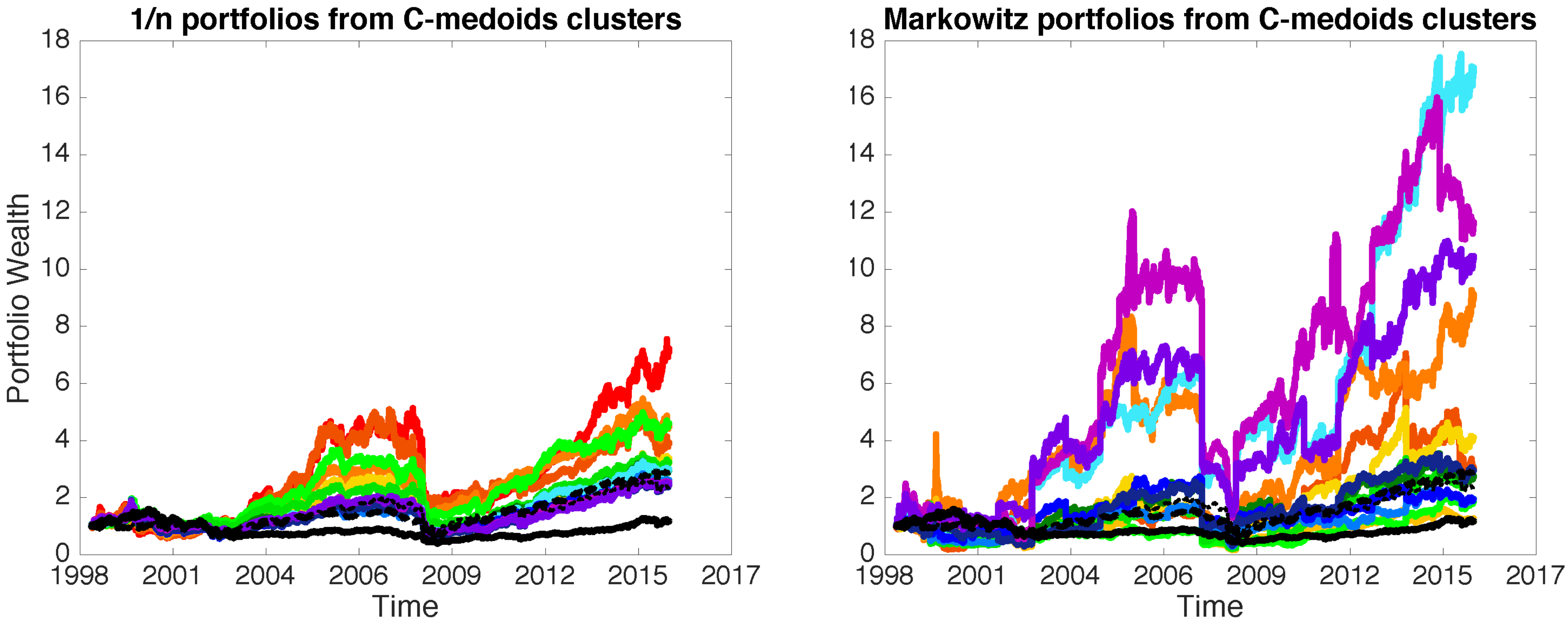
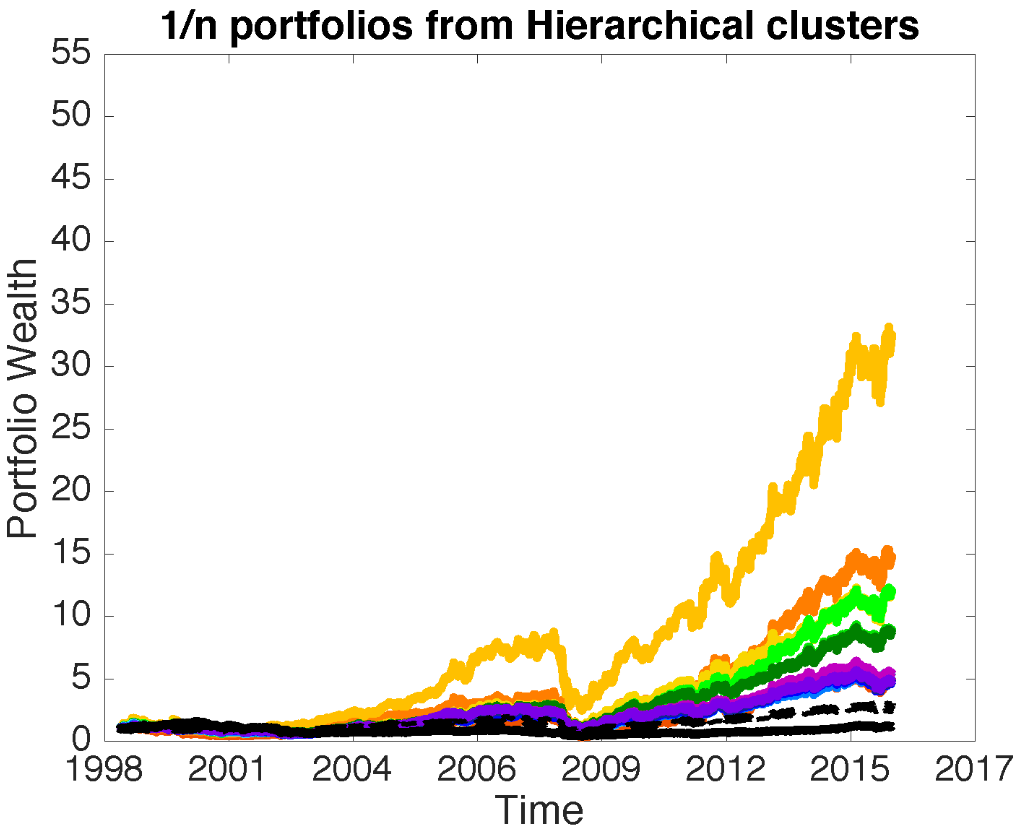


Figure 3. Cumulative portfolio wealth comparison (Distance measure: squared Euclidean): Black - Buy&hold STOXX600 NA(solid), Markowitz (dashed), 1/n (doted)

Hierarchical clusters' portfolios



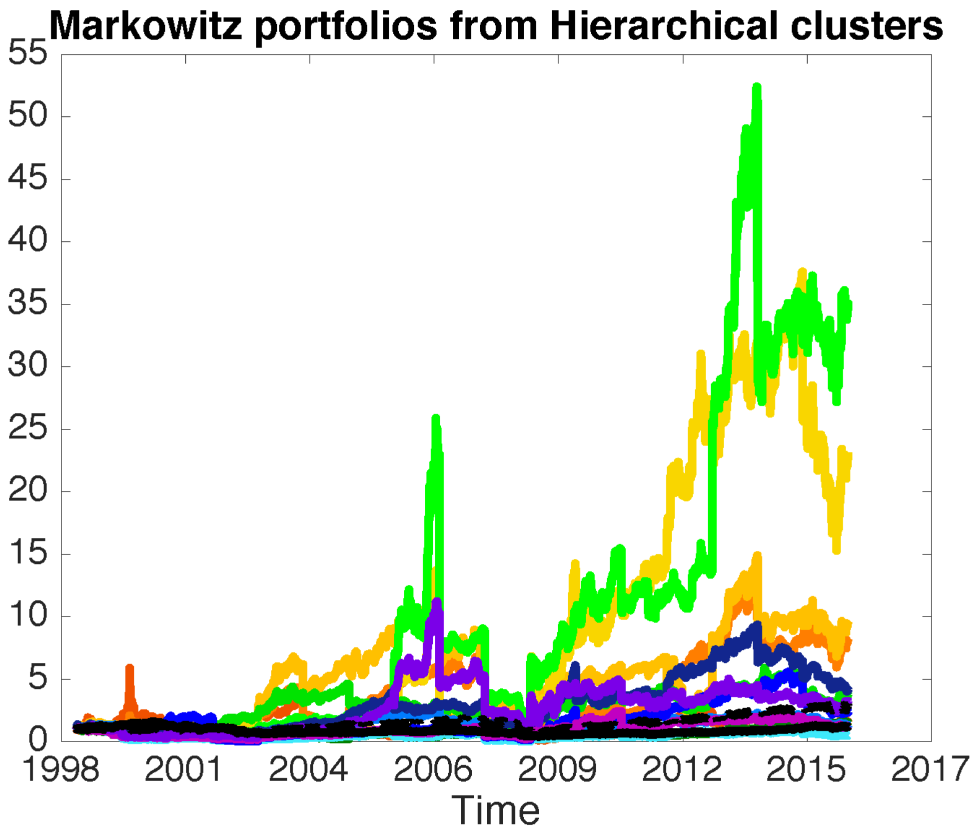


Figure 4. Cumulative portfolio wealth comparison (Distance measure: Euclidean, Agglomeration method: weighted): Black - Buy&hold STOXX600 NA(solid), Markowitz (dashed), 1/n (doted)

Best Performing Methods and Distances

Algo	Euclidean	Sq. Euclid.	Cityblock	Minkowski
Single	19.4857	9.9379	30.9323	19.4857
Complete	24.5083	13.6496	10.7680	24.5084
Average	20.9679	36.8007	19.5127	20.9679
Weighted	16.5012	34.9651	17.0920	16.5012
Centroid	27.1234	83.1752	19.7400	27.1234
Median	11.6537	7.7195	29.8658	11.6537
Ward	19.1667	8.3456	17.6150	19.1667

Table 3. Best performing agglomeration Method and Distances (Markowitz portfolios, Maximum Sharpe portfolio selection)

Best Performing Methods and Distances

Algo	Euclidean	Sq. Euclid.	Cityblock	Minkowski
Complete	19.3	11.0	9.8	9.8
Average	16.2	17.8	17.7	17.8
Weighted	16.9	47.6	36.2	49.3
Centroid	16.8	46.7	28.9	188.5
Median	16.3	43.2	28.9	34.8
Ward	7.8	25.8	27.9	55.6

Table 4. Best performing agglomeration Method and Distances (Markowitz portfolios, Random portfolio selection)

Portfolios' portraits

Year	1998	1999	2000	2001	2002	2003
W_1	0.2666	0.1892	0.4598	0	0	0.0028
W_2	0.5252	0.5806	0.2210	0.4901	0.1066	0
W_3	0.2082	0.0022	0	0.0123	0	0.2251
W_4	0	0.2281	0	0.3914	0.0003	0.6186
W_5	0	0	0	0	0.8931	0
W_6	0	0	0.3193	0.1061	0	0.1271
W_7	0	0	0	0	0	0.0264

Table 5. Weights of clusters in Markowitz portfolio (Distance measure: squared Euclidean, agglomeration method: weighted)

Portfolios' portraits

Year	1998	1999	2000	2001	2002	2003
ES_1	-0.0729	-0.0515	-0.0933	-0.1459	-0.1992	-0.0844
ES_2	-0.0335	-0.0792	-0.0918	-0.0674	-0.1441	-0.0698
ES_3	-0.0300	-0.1229	-0.1975	-0.1580	-0.0367	-0.0183
ES_4	-0.0753	-0.0964	-0.1955	-0.0900	-0.0434	-0.0169
ES_5	-0.1079	-0.0659	-0.0697	-0.0525	-0.0478	-0.1101
ES_6	-0.0401	-0.1531	-0.0208	-0.0525	-0.1794	-0.1099
ES_7	-0.0641	-0.0443	-0.0526	-0.0856	-0.1925	-0.0316

Table 6. Expected shortfalls of stocks-constituents of Markowitz portfolios

k - means clusters' random portfolios

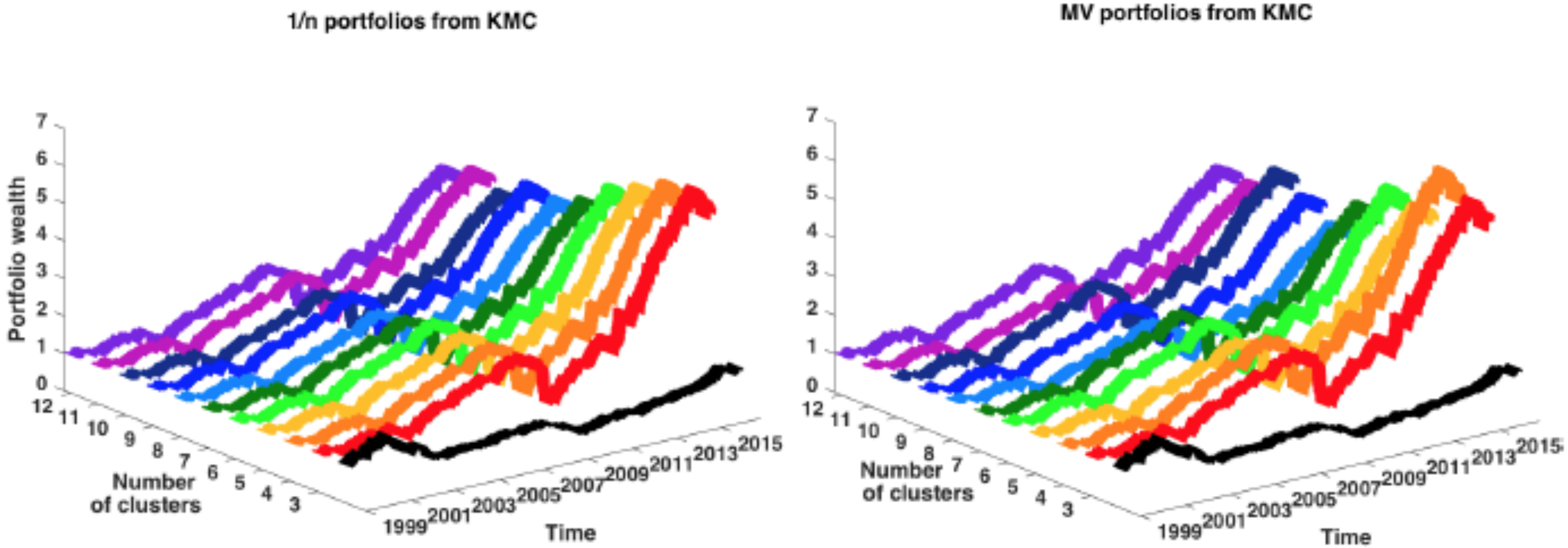


Figure 5. Average cumulative return over 100 randomly selected portfolios: 1/n portfolios (left), Markowitz portfolios (right), Black - STOXX600 NA

FUZZY C - means clusters' random portfolios

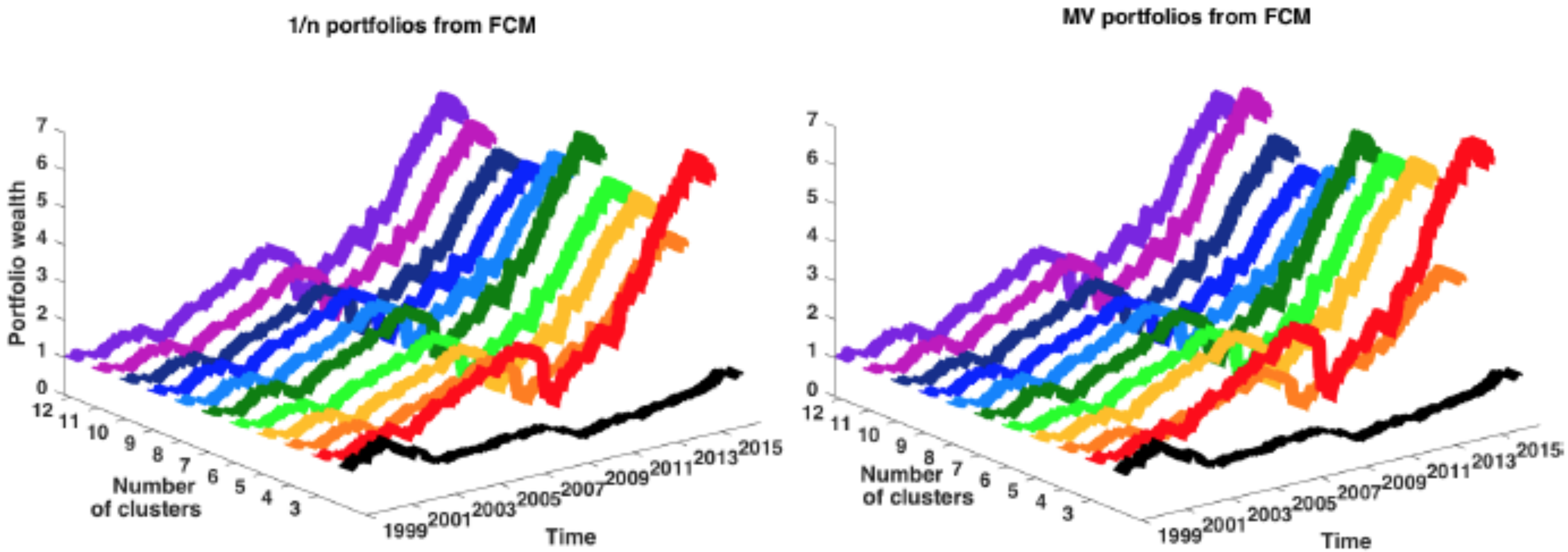


Figure 6. Average cumulative return over 100 randomly selected portfolios: 1/n portfolios (left), Markowitz portfolios (right), Black - STOXX600 NA

Hierarchical clusters' random portfolios

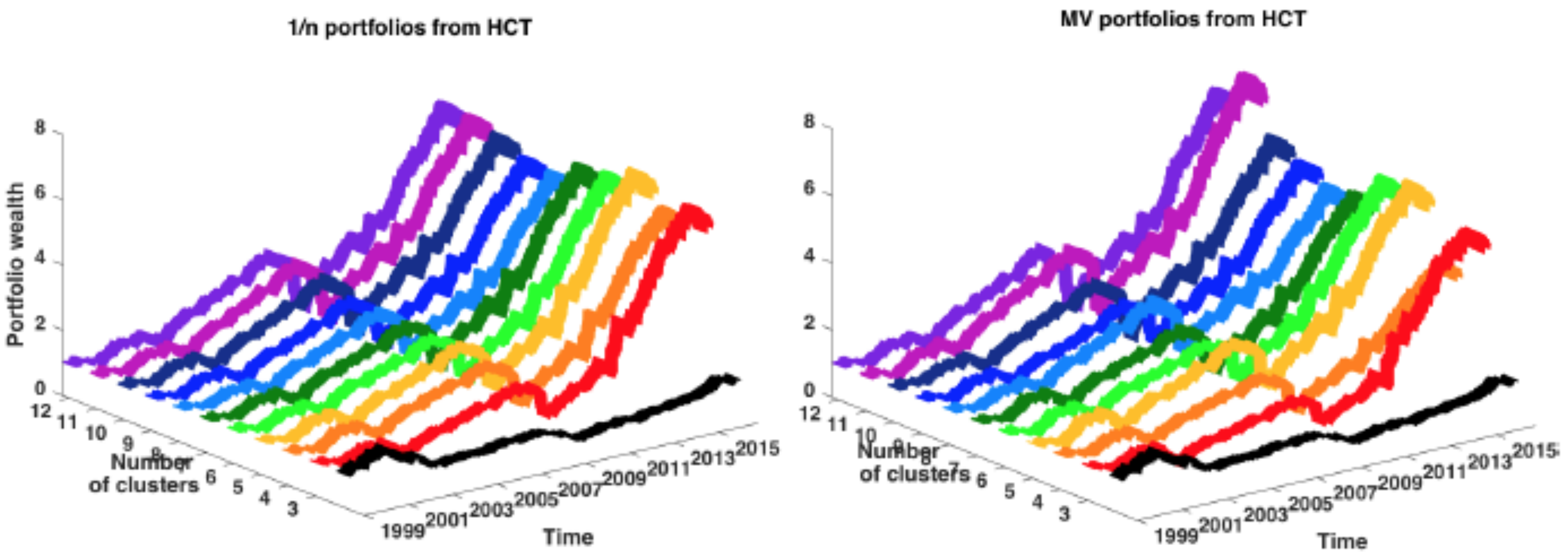


Figure 7. Average cumulative return over 100 randomly selected portfolios: 1/n portfolios (left), Markowitz portfolios (right), Black - STOXX600 NA

Validation of partition: Silhouette width

Details

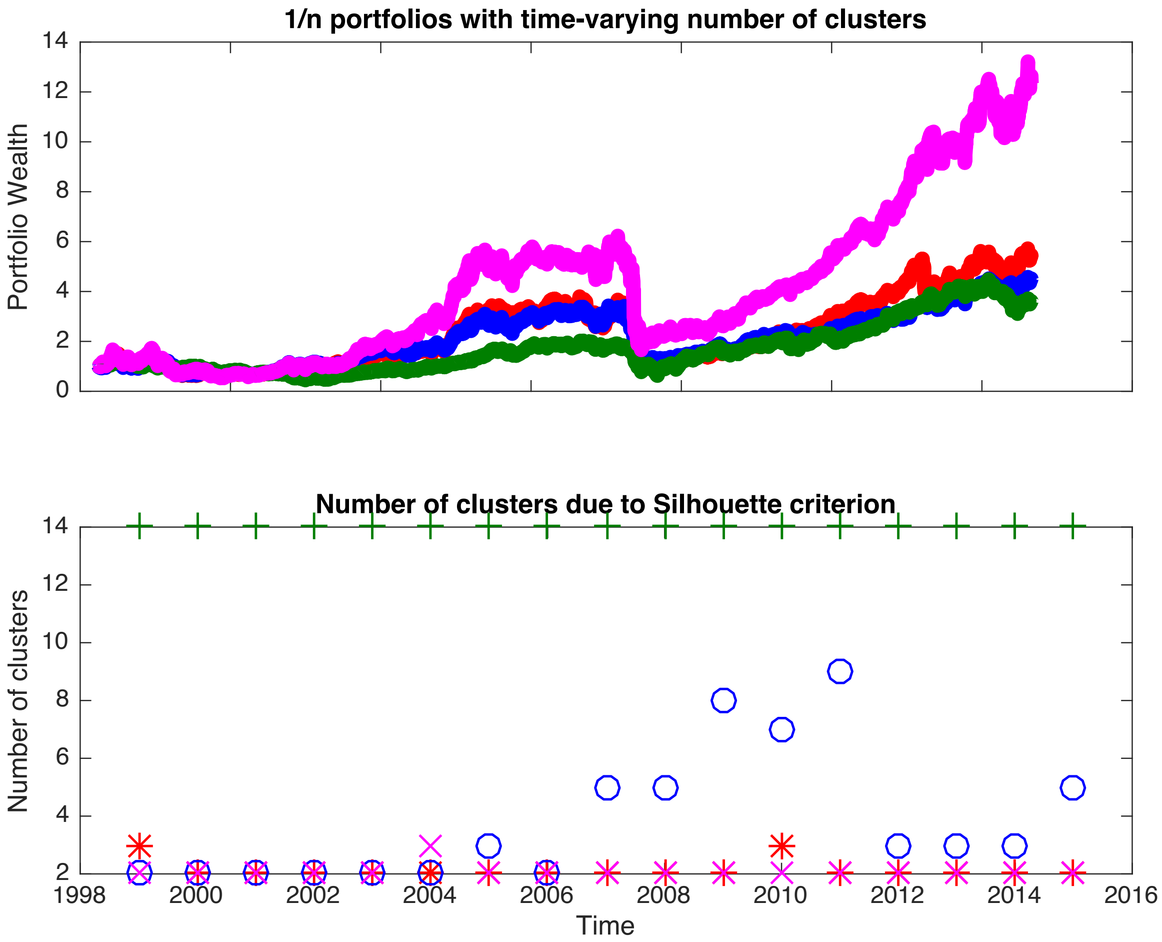


Figure 8. k-means , FUZZY c-means , Hierarchical and C-medoids clusters' portfolios

Validation of partition: Calinski-Harabasz criterion

Details

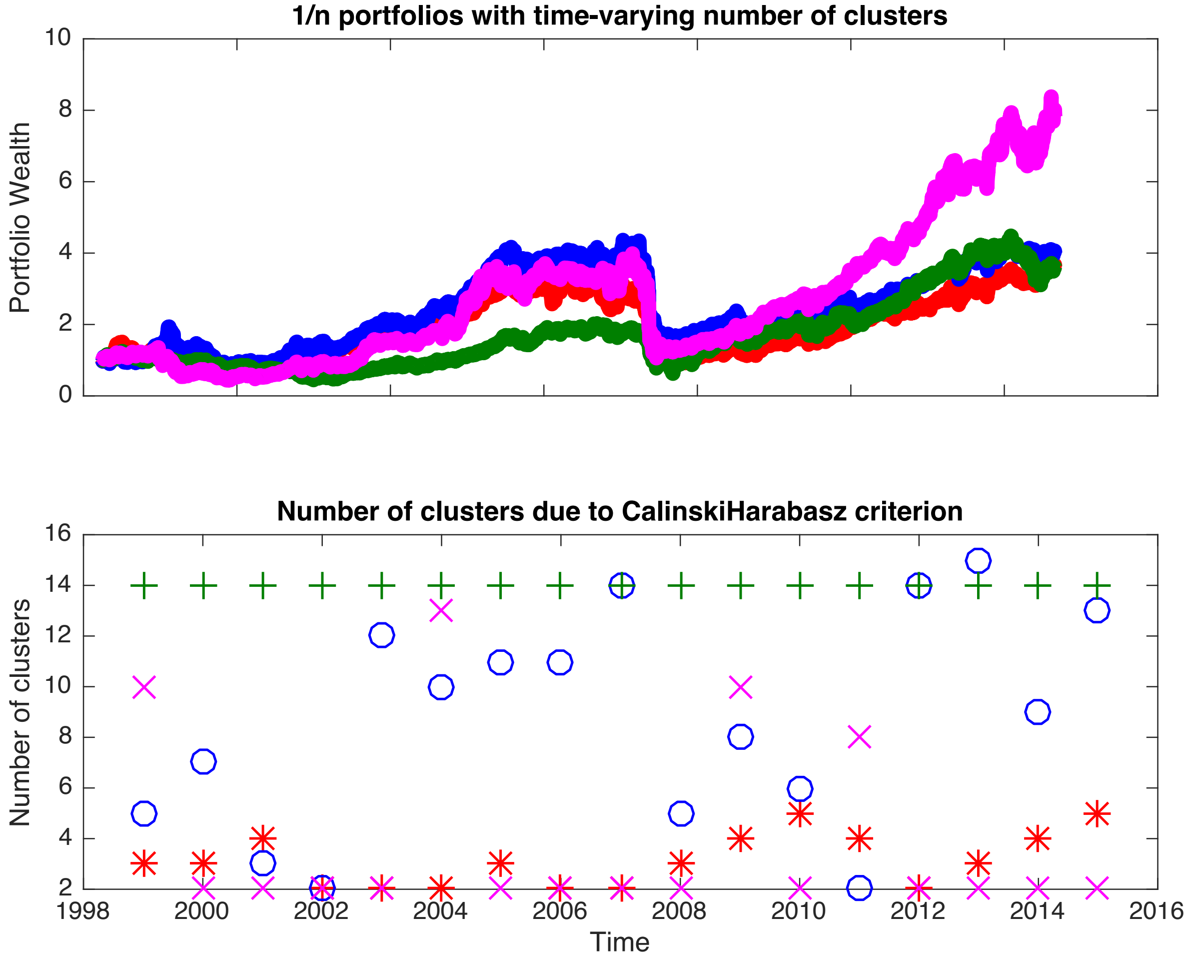


Figure 9. k-means , FUZZY c-means, Hierarchical and C-medoids clusters' portfolios

Validation of partition: Davies-Bouldin index

Details

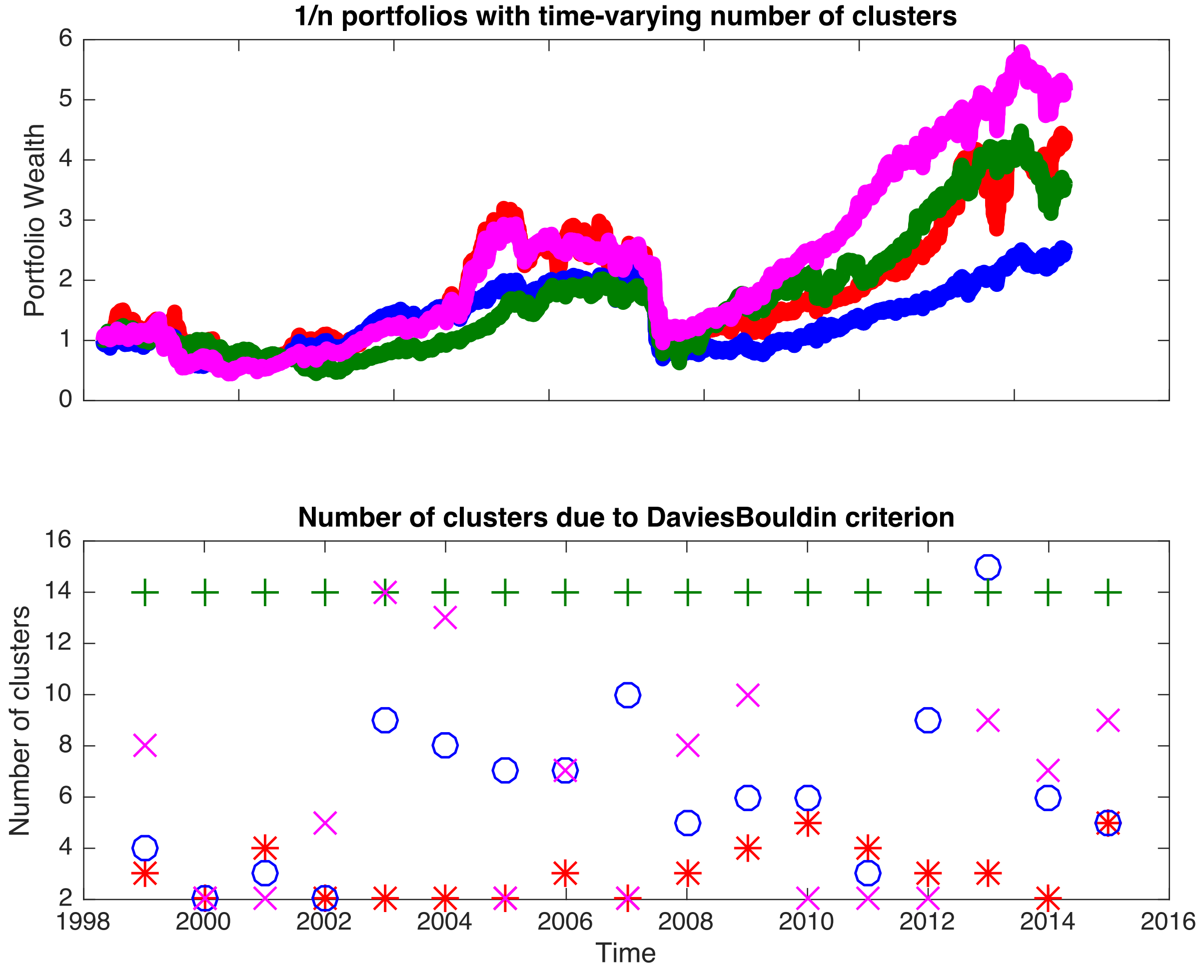


Figure 10. k-means, FUZZY c-means, Hierarchical and C-medoids clusters' portfolios

Conclusion

- Improvement of portfolio diversification
  - outperforms benchmarks in out-of-sample framework
- Risk-profile clustering strategy
  - dimension reduction of assets’ universe
  - multiple risk measures
  - hierarchical clustering portfolios demonstrate best performance

Outlook

- Other datasets
  - Mutual funds
  - Hedge funds
- Other clustering methods
- Other risk measures

Value at Risk (VaR)

- Portfolio loss X
  - Given pdf $f(x)$ and cdf $F(x)$
- Value at Risk

$$VaR_\alpha = x_\alpha = F^{-1}(\alpha) \tag{1}$$

Back to "Methodology"

Expected shortfall

Let L_i, $i \in \{1, \ldots, t\}$, be a (continuous) series of portfolio losses and q_θ the θ-quantile of these losses

$$\mathrm{ES}_t = \mathrm{E}[L_t | L_t > q_\theta] \tag{2}$$

Back to "Methodology"

k - means Clustering

- Fix k Clusters a priori
- Assign observations to cluster j with mean $\overline{x}_j$
- Computationally hard/iterative
- Standard Algorithms do not yield unique allocation
- time investment $\mathcal{O}(n^{\rho k+1} \log n)$

Back to "Methodology"

k - means Clustering

Minimize the within cluster sum of squares w.r.t.
$\mathcal{S} = \{S_1, \ldots, S_k\}, \bigcup_{j=1}^{k} S_j = \{1, 2, \ldots, n\}$:

$$\hat{\mathcal{S}} = \underset{\mathcal{S}}{\operatorname{argmin}} = \sum_{j=1}^{k} \sum_{i \in \mathcal{S}_j} ||x_i - \mu_j||_2^2 \tag{3}$$

The k - means standard algorithm is iterative starting from random partitions/points.

Back to "Methodology"

Standard Algorithm

Fix an initial set $\{\mu_j^{(t)}\}_{j=1}^k$, $t = 1$

Assign: $\hat{j}(i) = \underset{j}{\operatorname{argmin}} ||x_i - \mu_j^{(t)}||^2$

x_i belongs then to cluster $\hat{j}(i)$ resulting in (new) partition

$$\bigcup_{j=1}^{k} \mathcal{S}_j^{(t)} = \{1, \ldots, n\}$$

Update: $\mu_j^{(t+1)} = \left(\#\mathcal{S}_j^{(t)}\right)^{-1} \sum_{i \in \mathcal{S}_j^{(t)}} x_i$

Iterate: assign, update until convergence.

Back to "Methodology"

FUZZY c-means clustering (FCM)

Jim Bezdek 1981
FCM is a clustering method that allows each data point to belong to multiple clusters with varying degrees of membership.

- Randomly initialise the cluster membership values, μ_{ij}
- Calculate the cluster centers

$$c_j = \frac{\sum_{i=1}^{D} \mu_{ij}^m x_i}{\sum_{i=1}^{D} \mu_{ij}^m} \tag{4}$$

Back to "Methodology"

FUZZY c-means clustering (FCM)

- Update μ_{ij} according to the following

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{N} \left(\frac{||x_i - c_j||}{||x_i - c_k||} \right)^{\frac{2}{m-1}}} \tag{5}$$

- Calculate the objective function, J_m

$$J_m = \sum_{i=1}^{D} \sum_{j=1}^{N} \mu_{ij}^{m} ||x_i - c_j||^2 \tag{6}$$

- Iterate: assign, update until convergence.

Back to "Methodology"

Hierarchical Algorithms, Agglomerative Techniques

1.Construct the finest partition, i.e. each point is one cluster.

2.Compute the distance matrix $\mathcal{D}$.

DO

3.Find the two clusters with the closest distance.

4.Unite the two clusters into one cluster.

5.Compute the distance between the new groups and obtain a reduced distance matrix $\mathcal{D}$.

UNTIL all clusters are agglomerated.

Back to "Methodology"

Agglomerative Techniques

After unification of P and Q one obtains the following distance to another group (object) R

$$d(R, P+Q) =$$
$$\delta_1 d(R,P) + \delta_2 d(R,Q) + \delta_3 d(P,Q) + \delta_4 |d(R,P) - d(R,Q)|$$

δ_j - weighting factors

Denote by $n_P = \sum_{i=1}^{n} \mathbf{I}(x_i \in P)$ number of objects in group ***P***

Back to "Methodology"

Agglomeration methods

Name	δ_1	δ_2	δ_3	δ_4
Single linkage	1/2	1/2	0	-1/2
Complete linkage	1/2	1/2	0	1/2
Average linkage (unweighted)	1/2	1/2	0	0
Average linkage (weighted)	$\frac{n_P}{n_P+n_Q}$	$\frac{n_Q}{n_P+n_Q}$	0	0
Centroid	$\frac{n_P}{n_P+n_Q}$	$\frac{n_Q}{n_P+n_Q}$	$-\frac{n_P n_Q}{(n_P+n_Q)^2}$	0
Median	1/2	1/2	-1/4	0
Ward	$\frac{n_R+n_P}{n_R+n_P+n_Q}$	$\frac{n_R+n_Q}{n_R+n_P+n_Q}$	$-\frac{n_R}{n_R+n_P+n_Q}$	0

where $n_P = \sum_{i=1}^{n} \mathbf{I}(x_i \in P)$ denotes the number of objects in group P.

Back to "Methodology"

Distance Measures

Distance	$d(x, y)$
Euclidean	$\|x - y\|$ (L_2-Metric)
Maximum	$max_i \lvert x_i - y_i \rvert$ (L_∞-Metric)
Manhattan	$\sum_i \lvert x_i - y_i \rvert$ (L_1-Metric)
Canberra	$\sum_i \frac{\lvert x_i - y_i \rvert}{\lvert x_i \rvert + \lvert y_i \rvert}$
Minkowski	$\|x - y\|_p$

Back to "Methodology"

Distance Measures

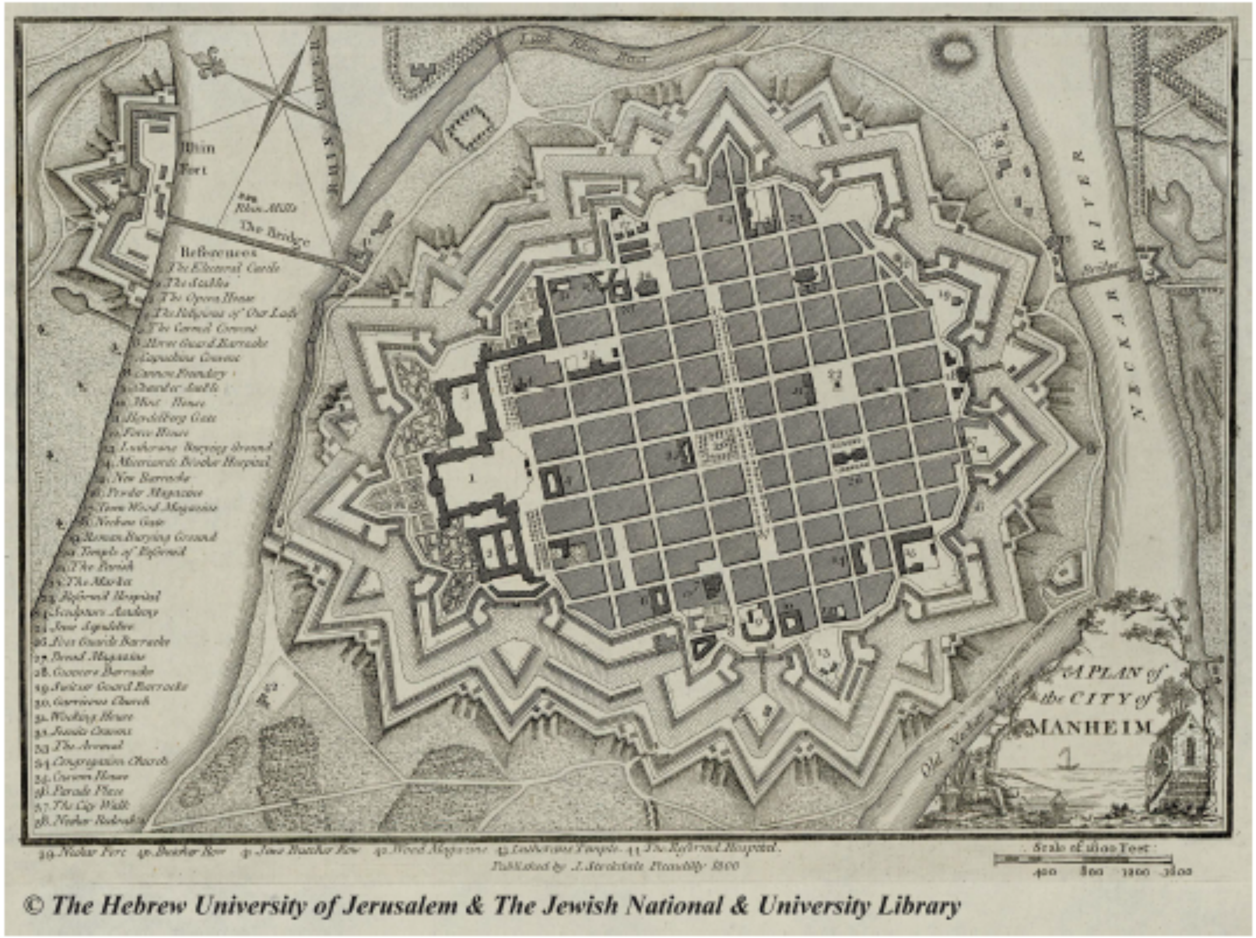


Figure Map of Mannheim around 1800

▸ Back to "Methodology"

Source: http://historic-cities.huji.ac.il

Markowitz rule

Log returns $X_t \in \mathbb{R}^p$:

$$
\begin{aligned}
\min_{w_t \in \mathbb{R}^p} \quad & \sigma^2_{P,t}(w_t) \stackrel{\text{def}}{=} w_t^\top \Sigma_t w_t \\
\text{s.t.} \quad & \mu_{P,t}(w_t) = r_T, \\
& w_t^\top 1_p = 1, \\
& w_{i,t} \geq 0
\end{aligned} \tag{7}
$$

where r_T "target" return,
$\Sigma_t \stackrel{\text{def}}{=} E_{t-1}\{(X_t - \mu)(X_t - \mu)^\top\}$ Back to "Methodology"

C - medoids

C-medoids clustering is related to the k-means. Both attempt to minimize the distance between points labeled to be in a cluster and a point designated as the center of that cluster. In contrast to the k-means, C-medoids chooses datapoints as centers (medoids) and works with an arbitrary matrix of distances.

Back to "Results"

Silhouette Value

The silhouette value for each point is a measure of how similar that point is to points in its own cluster, when compared to points in other clusters. The silhouette value for the i-th point, S_i , is defined as

$$S_i = (b_i - a_i)/max(a_i, b_i)$$

where a_i is the average distance from the i-th point to the other points in the same cluster as i
b_i is the minimum average distance from the i-th point to points in a different cluster, minimized over clusters

▸ Back to "Results"

Calinski-Harabasz criterion

The Calinski-Harabasz criterion is sometimes called the variance ratio criterion (VRC). The Calinski-Harabasz index is defined as

$$VRC_k = \frac{SS_B}{SS_W} \cdot \frac{N-k}{k-1}$$

where SS_B is the overall between-cluster variance,
SS_W is the overall within-cluster variance,
k is the number of clusters,
N is the number of observations

Back to "Results"

Davies-Bouldin Criterion

The Davies-Bouldin criterion is based on a ratio of within-cluster and between-cluster distances

$$DB = \frac{1}{k}\sum_{i=1}^{k} max_{j\neq i}\{D_{i,j}\}$$

where $D_{i,j}$ is the within-to-between cluster distance ratio for the i-th and j-th clusters.

$$\{D_{i,j}\} = \frac{\bar{d}_i + \bar{d}_j}{d_{i,j}}$$

d_i/d_j are average distance between each point in the i-th/j-th cluster and centroid of the i-th/j-th cluster
$d_{i,j}$ is the Euclidean distance between the centroids of the i-th and j-th clusters.

▸ Back to "Results"