Asset Allocation Strategy with Non-Hierarchical Clustering Risk Parity Portfolio

Kei Nakagawa [ID]**, Takanobu Kawahara, Akio Ito**

Nomura Asset Management Co., Ltd., Koto-ku, Japan
Email: kei.nak.0315@gmail.com

Abstract

Since the great financial crisis of 2008, many studies have pointed out that even in the portfolio where the asset allocation is sufficiently diversified; it is still possible that risk allocation is well concentrated to a few assets. One approach to this problem is risk parity strategies which equalize the risk contribution of each asset. However, even if we equalize the risk contribution, risk sources are not necessarily diversified. In this paper, we propose non-hierarchical clustering-risk parity strategy which will equalize risk contribution from and within each cluster. In addition, in order to ensure robustness of clustering, we also propose x-means++ algorithm which combines k-means++ with x-means. Assuming assets with similar movement have common risk sources; our approach will construct a portfolio which equalizes risk sources. Empirical analysis using actual price data of various asset classes shows that our proposed method will outperform risk-parity strategies or hierarchical clustering risk parity strategies.

Keywords

Non-Hierarchical Clustering Risk Parity, x-Means++, Asset Allocation

1. Introduction

Since the great financial crisis of 2008, many studies have pointed out that even in the portfolio where the asset allocation is sufficiently diversified; it is still possible that risk allocation is well concentrated to a few assets.

Traditionally portfolios are constructed using the men-variance approach [1]. To calculate optimal portfolio weights, this method performs optimization using expected returns and risks. However, as these numbers are hard to estimate with precision, the calculated weights tend to be biased and often not as diversified as initially intended [2].

One approach to this problem is risk parity strategies which equalize the risk

contribution of each asset [3]. Also, according to [4], unlike mean variance, risk parity does not depend on the estimation accuracy of the covariance matrix for performance. While pension funds generally have a diversified portfolio of stocks and bonds, it is still said that their portfolios have a large bias in risk contribution.

As the risk of stocks is much larger than bonds, the majority of the portfolio risk comes from stocks. In the first place, we make diversified investments because we expect other assets to support the overall performance when one asset is in a poor condition.

When the majority of the portfolio risk is occupied by stock, other assets do not make up for the stock market slump, and the expected effects of diversified investment cannot be obtained. Risk parity strategies have been proposed as an alternative to these conditions.

The concept of risk parity applies to the asset allocation problem and there are many previous studies. For example, [3] demonstrates that risk parity performance is more efficient as measured by Sharpe Ratio, compared to traditional balanced portfolios with a 60:40 equity and bond investment ratio.

On the other hand, some caveats are pointed out for risk parity strategy too. [5] and [6] have pointed out that even if the risk contributions are made equal, the sources of risk are not diversified.

Therefore, in this paper, we first group assets with similar movements using non-hierarchical clustering method. Then, we propose a non-hierarchical clustering/risk parity strategy in which the risk contributions are equal both in each cluster and within the cluster. We also propose x-means++ which is a combination of x-means algorithm [7] [8] and k-means++ algorithm [9] in order to secure the robustness of clustering.

Assuming assets with similar movement have common risk sources; our approach will construct a portfolio which equalizes risk sources. Empirical analysis using actual price data of various asset classes shows that our proposed method will outperform risk-parity strategies [3] or hierarchical clustering risk parity strategies [10].

The remaining sections of this paper are organized as follows. In Section 2, we briefly describe the related studies of the risk-based portfolio. In Section 3, we introduce the risk parity portfolio and non-hierarchical risk parity portfolio. In Section 4, we describe the x-mean++ clustering and in Section 5, we verify its effectiveness through empirical analysis with the actual financial market data. Finally, we conclude in Notation.

2. Related Work

Unlike the mean-variance portfolio, which uses both estimated return and risk, risk-based portfolios only use estimated risk to construct a portfolio. As predicting future returns is troublesome and also error maximization features of mean-variance optimization approach tend to construct a portfolio concentrated on a few securities [2], risk-based portfolios that do not use the expected return

have attracted attention of practitioners.

Typical risk-based portfolios are the minimum variance portfolio [11], the risk parity portfolio [3], and the maximum diversification portfolio [12]. Each of these has shown to provide better performance than market capitalized portfolios and mean variance portfolios [13]. The minimum variance portfolio determines the asset allocation so that the variance of the portfolio is the smallest. This portfolio is located at the left end of the efficient frontier in the risk/return plane, and the expected return of this portfolio is also the smallest of the efficient frontier. However, it is known that minimum variance portfolios tend to have higher risk/return ratio ex-post. The maximum diversification portfolio is the portfolio with the largest diversification effect, and is obtained by maximizing the diversification ratio which is the weighted average of asset risk divided by portfolio risk.

Furthermore, it is known that these three portfolios can be written as a generalized risk-based portfolio [14]. Extensions to these three portfolios have been proposed. As an extension of the minimum variance portfolio, there are 1) those that include higher-order moments [15], 2) those that devise the method of estimating the co-variance matrix (Gaussian Process Latent Variable Model: GPLVM [16] and t-Process Laten Variable Model: TPLVM [17]), and those that use downside risk such as conditional value at risk (CVaR) [18] [19] which is an alternative to co-variance based risk measure.

As an extension of risk parity, there are principal component risk parity [5] and complex principal component risk parity [6] that focus on the source of risk. Furthermore, there is a hierarchical cluster risk parity [10] that divides the risk into clusters and distributes the risk to those clusters. This study proposes a non-hierarchical risk parity portfolio.

3. Non-Hierarchial Clustering Risk Parity Portfolio

3.1. Risk Parity

We consider a portfolio of n risky asset and let $R = (R_1, \cdots, R_N)^{\mathrm{T}}$ be the return (random variable) vector of each assets, $\mu = (\mu_1, \cdots, \mu_N)^{\mathrm{T}}$ be the vector of expected returns, and $\Sigma = E\left[(R-\mu)(R-\mu)^{\mathrm{T}}\right]$ be the covaraiance matrix of asset returns. Additionally, we denote weight vector of portfolio as $w = (w_1, \cdots, w_N)^{\mathrm{T}}$.

To derive the specific form of the risk parity portfolio, we will introduce Marginal Risk Contribution (*MRC*) as a derivative of portfolio risk $\sigma_P = \sqrt{w^{\mathrm{T}} \Sigma w}$ by weight w.

$$MRC = \frac{\partial \sigma_P}{\partial w} = \frac{\Sigma w}{\sigma_P}, MRC_i = \frac{(\Sigma w)_i}{\sigma_P} \tag{1}$$

We will be able to decompose portfolio risk using *MRC* as following.

$$\sigma_P = \sum_{i=1}^{N} w_i \times MRC_i = w^{\mathrm{T}} MRC \tag{2}$$

We will additionally define Risk Contribution (*RC*) as below.

$$RC_i = \frac{w_i \times MRC_i}{\sigma_P} \tag{3}$$

Finally Risk Parity Portfolio can be defined as a portfolio which RCs from each asset i are equal.

$$RC_i = RC_j, \text{ for all } i, j \tag{4}$$

Restricting short-selling and usage of leverage, [20] showed that portfolio weights can be calculated efficiently by solving optimization problem (5)-(6). As (5)-(6) are a convex optimization problem this will have a unique solution.

$$\min_{w} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(RC_i - RC_j \right) \tag{5}$$

$$\text{s.t, } \sum_{i=1}^{N} w_i = 1, \, w_i > 0 \tag{6}$$

3.2. Non-Hierarchical Clustering Risk Parity

Essence of risk parity portfolio is controlling risk allocation. While constructing a risk parity portfolio we choose to allocate risk contribution equally to each asset, but we can consider alternative way of allocating risk, which is called risk budgeting strategy [21].

In this article we aim to equalize risk contribution from each cluster and at the same time equalize risk contribution from each asset within every clusters.

To achieve this goal, we will first perform non-hierarchical clustering using asset returns to determine risk clusters.

And using this cluster we will solve optimization problem below to get the portfolio weights of non-hierarchical clustering risk parity portfolio. k stands for number of clusters and N_k stands for number of assets in each cluster. We can see that risk contribution from each cluster is equalized and risk contributions from each asset within each cluster are equalized in this portfolio. We will introduce this method as non-hierarchical clustering risk parity strategy.

$$\min_{w} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(RC_i - \frac{1}{k} \times \frac{1}{N_k} \right) \tag{7}$$

$$\text{s.t, } \sum_{i=1}^{N} w_i = 1, \, w_i > 0 \tag{8}$$

4. *x*-Means++

The k-means is a standard algorithm of a hierarchical clustering method which is easy to implement and has high calculation efficiency.

A cluster refers to a collection of data points aggregated together according to certain distances and a centroid C_i is a center point in each cluster.

We first define a target centroid number k.

The k-means divides the data into k clusters so as to minimize the following

evaluation function in which $d(x, y)$ is the distance function.

$$\sum_{i=1}^{k} \sum_{x \in C_i} \left(d(x, C_i) \right)^2 \tag{9}$$

However, the *k*-means algorithm has two shortcomings. First, the result may depend on the initial clusters, so the algorithm does not guarantee the optimal clustering. Second, the algorithm needs to set the numbers of clusters *k* initially.

The initialization method called *k*-means++ [9] was proposed for the first shortcoming. And *x*-means [7] [8] was proposed for the second one. In this paper, we combine the *k*-means++ initialization with *x*-means. Next section describes the *k*-means++ and *x*-means algorithm which are the components of our proposed. Table 1 shows the comparing each algorithm.

4.1. *k*-Means++

The feature of *k*-means++ is the initialization of centroids C_i. The *k*-means++ algorithm decides the *k* clusters as follows:

Step 1:

Choose one data point at random in data as an initial centroid C_1.

Step 2:

For each data point x_i, compute $d(x_i)$, the distance between x_i and the nearest centroid that has already been chosen.

Step 3:

Choose one new data point x_p at random as a new centroid with the following probability

$$\left(d(x_p) \right)^2 \Big/ \sum_{i=1}^{n-1} \left(d(x_i) \right)^2 \tag{10}$$

here, the data already selected as the cluster has the probability 0 because the distance between the data and the nearest centroid is 0.

Step 4:

Repeat the step 2 and 3 until *k* centroids have been chosen.

Table 1. Comparison of clustering algorithms.

	Pros	Cons
k-means	easy to implement and has high calculation efficiency.	depends on the initial clusters. needs to set the numbers of clusters *k* initially.
k-means++	can decrease the dependency on initial clusters.	needs to set the numbers of clusters *k* initially.
x-means	can determine the optimal number of clusters.	depends on the initial clusters.
x-means++ (Ours)	can determine the optimal number of clusters. can decrease the dependency on initial clusters.	-

4.2. *x*-Means

The *x*-means algorithm can determine the optimal number of clusters unlike *k*-means algorithm which the number of clusters has to be given in advance.

The process of *x*-means clustering is to perform *k*-means repeatedly from $k = 2$ until a Bayesian information criterion (*BIC*) does not improve.

This study applies the following algorithm proposed by [8].

Step 1:

We prepare *p*-dimensional data whose sample size is *n*.

Step 2:

We apply *k*-means ($k = 2$) to all data. We name the divided clusters as C_1, C_2.

Step 3:

We repeat the following procedure from step 4 to step 9 by setting $i = 1, 2$.

Step 4:

For a cluster C_i, we apply *k*-means ($k = 2$). We name the divided clusters as C_i^1, C_i^2.

Step 5:

We assume the following *p*-dimensional normal distribution for the data $x_i \in C_i$:

$$f(x; \theta_i) = \frac{1}{(2\pi)^{\frac{p}{2}} \sqrt{\det |V_i|}} \exp{-\frac{1}{2}(x - \mu_i)^{\mathrm{T}} V_i^{-1} (x - \mu_i)} \qquad (11)$$

Then, we calculate *BIC* as

$$BIC = -2 \log L(\theta_i; x_i \in C_i) + q \log n_i \qquad (12)$$

where $\theta_i = [\mu_i, V_i]$ is the maximum likelihood estimate of the *p*-dimensional normal distribution; μ_i is *p*-dimensional means vector, and V_i is $p \times p$ dimensional covariance matrix; q is the number of the parameters dimension, and it becomes $q = p(p+3)/2$. n_i is the number of elements in C_i. L is the likelihood function which indicates $L(\cdot) = \prod f(\cdot)$.

Step 6:

We assume the *p*-dimentional normal distributions with their parameters $\theta_i^{(1)}, \theta_i^{(2)}$ for C_i^1, C_i^2 respectively. The probability density function of this 2-division model becomes

$$g(\theta_i^{(1)}, \theta_i^{(2)}; x) = \alpha_i \left[f(\theta_i^{(1)}; x) \right]^{\delta_i} \left[f(\theta_i^{(2)}; x) \right]^{1 - \delta_i}, \qquad (13)$$

where

$$\delta_i = \begin{cases} 1, & \text{if } x \text{ is included in } C_i^1 \\ 0, & \text{if } x \text{ is included in } C_i^2 \end{cases} \qquad (14)$$

x_i will be included in either C_i^1 or C_i^2; α_i is a constant which lets equation (12) be a probability density function ($1/2 \leq \alpha_i \leq 1$). We approximate α_i as follows:

$$\alpha_i = 0.5/K(\beta_i), \qquad (15)$$

where β_i is a normalized distance between the two clusters, shown by

$$\beta_i = \sqrt{\frac{\|\mu_1 - \mu_2\|^2}{|V_1| + |V_2|}}, \tag{16}$$

$K(\cdot)$ stands for a lower probability of normal distribution. The *BIC* for this model is

$$BIC' = -2\log L'(\theta_i'; x_i \in C_i) + q\log n_i \tag{17}$$

where $\theta_i' = \left[\theta_i^{(1)}, \theta_i^{(2)}\right]$ is the maximum likelihood estimate of the *p*-dimensional normal distribution; since there are two parameters of mean and covariance for each *p* variable, the number of parameters dimension becomes $q' = p(p+3)$. L' is the likelihood function which indicates $L'(\cdot) = \prod g(\cdot)$.

Step 7:

If $BIC > BIC'$, we prefer the two-divide model, and decide to continue the division; we set $C_i \leftarrow C_i^1$. As for C_i^2, we push the *p*-dimensional data, the cluster centers, the log likelihood and the *BIC* onto the stack. We return to Step 4.

Step 8:

If $BIC \leq BIC'$, we prefer not to divide clusters anymore, and decide to stop. We extract the stacked data, which is stored in Step 7, and we set $C_i \leftarrow C_i^2$. We return to Step 4. If the stack is empty, go to Step 9.

Step 9:

The 2-division procedure for C_i is completed. We renumber the cluster identification such that it becomes unique in C_i

Step 10:

The two-division procedure for initial $k = 2$ divided clusters is completed. We renumber all cluster identifications such that they become unique.

Step 11:

We note the outputs of the cluster identification, the center of each cluster, the log likelihood of each cluster, and the number of elements in each cluster.

5. Empirical Results

This section describes the empirical study with real market data.

5.1. Datasets

We perform empirical analysis using equity and bond futures price data. The indices we use in this study are summarized in Table 2. We use 15 equity futures and 12 bond futures from May 2005 to May 2020. The summary of statistics of each index is reported in Table 3.

5.2. Parameters Settings

We compare risk parity (RP) [3], hierarchical risk parity (HRP) [10], clustering risk parity (CRP) using *k*-means++ which the number of clusters is fixed, and non-hierarchical risk parity using *x*-means++ (XRP). We set *k* of *k*-means++ from 2 to 8. Our simulation process is given below.

Table 2. Investment assets.

	Investment assets							
Equity furure (15 assets)	S&P500 (SP)	NAS DAQ (NQ)	CA (PT)	GB (Z)	FR (CF)	DE (GX)	EU (VG)	ES (IB)
	NL (EO)	CH (SM)	NIKKEI (NK)	TOPIX (TP)	HK (HI)	AU (XP)	SG (QZ)	
Bond furure (12 assets)	US2Y (TU)	US5Y (FV)	US10Y (TY)	US20Y (US)	AU3Y (YM)	AU10Y (XM)	CA10Y (CN)	DE2Y (DU)
	DE5Y (OE)	DE10Y (RX)	GB10Y (G)	JP10Y (JB)				

[a]Words in parentheses denote tickers.

Table 3. Summary of statistics of investment assets.

Performance Statistics	SP	NQ	PT	Z	CF	GX	VG	IB
Return (%, Ann)	5.32	8.53	4.49	3.13	2.90	3.34	2.89	2.37
Risk (%, Ann)	19.34	26.30	18.70	18.42	22.38	22.86	23.42	23.04
R/R	0.28	0.32	0.24	0.17	0.13	0.15	0.12	0.10

Performance Statistics	EO	SM	NK	TP	HI	XP	QZ
Return (%, Ann)	2.97	5.38	4.85	3.74	6.22	5.08	4.80
Risk (%, Ann)	22.02	18.36	23.87	22.58	23.05	16.34	19.51
R/R	0.13	0.29	0.20	0.17	0.27	0.31	0.25

Performance Statistics	OE	RX	G	JB
Return (%, Ann)	2.50	4.28	3.44	1.93
Risk (%, Ann)	3.03	5.17	5.79	2.95
R/R	0.82	0.83	0.59	0.65

Performance Statistics	TU	FV	TY	US	YM	XM	CN	DU
Return (%, Ann)	1.33	3.03	4.31	5.76	2.80	2.80	3.92	0.83
Risk (%, Ann)	1.42	3.72	5.75	9.88	7.17	7.17	5.41	1.15
R/R	0.94	0.82	0.75	0.58	0.39	0.39	0.72	0.72

First, we estimate covariance matrix and perform clustering methods using 250 days of asset return data. Then, we construct each portfolio every 20 business days. Our simulation period is from April, 2001 to May, 2020.

5.3. Performance Measures

For evaluating an investment strategy, we use the following measures that are widely used in financial space. Returns are annualized, risk is calculated as standard deviation of return and R/R stands for return/risk ratio. In this paper, each portfolio will have different risk levels as we utilize wide range of assets with

various risk levels. We think R/R which is the efficiency of the portfolio performance, is more appropriate measure for performance evaluation for this study than return alone.

$$\text{Return} = \frac{250}{T}\sum_{t=1}^{T} r_t \tag{18}$$

$$\text{Risk} = \sqrt{\frac{250}{T-1}\sum_{t=1}^{T}\left(r_t - \mu\right)^2} \tag{19}$$

$$\text{R/R} = \text{Return/Risk} \tag{20}$$

$$\text{maxDD} = \min_{k\in[1,T]}\left(0, \frac{W_k}{\max\limits_{j\in[1,k]} W_j} - 1\right) \tag{21}$$

Here, r_t denotes the portfolio return at time t, μ denotes average of r_t and W_k denotes the wealth of portfolio at time k.

5.4. Results

Table 4 shows the result of simulation. The upper row shows the results for the entire period, the middle row shows the results for the first half, and the lower row shows the results for the second half. We compare risk parity (RP), hierarchical risk parity (HRP), clustering risk parity (CRPx) with x denoting the number of clusters, and non-hierarchical risk parity using by our proposed x-means++ (XRP).

Table 4. Performance statistics of portfolios.

Performance Measures	RP	HRP	CRP2	CRP3	CRP4	CRP5	CRP6	CRP7	CRP8	XRP
All Period (from Apr, 2001 to May, 2020)										
Return (%, Ann)	2.31	1.22	1.05	2.48	2.50	2.67	2.60	3.31	2.71	3.06
Risk (%, Ann)	2.04	1.35	3.59	2.65	2.78	3.38	2.48	2.83	3.20	2.25
R/R	1.14	0.90	0.29	0.93	0.90	0.79	1.05	1.17	0.85	1.36
maxDD (%)	7.24	3.93	18.07	11.20	9.96	8.66	8.48	7.48	9.23	7.07
First Half Period (from Apr, 2001 to Apr, 2010)										
Return (%, Ann)	2.28	1.74	−0.06	2.21	2.42	2.31	2.30	2.34	2.81	2.91
Risk (%, Ann)	2.42	1.82	4.26	3.14	3.54	3.23	2.83	3.00	3.95	2.62
R/R	0.94	0.95	−0.01	0.71	0.68	0.71	0.81	0.78	0.71	1.11
maxDD (%)	7.24	3.93	18.07	11.20	9.96	8.66	8.48	7.48	9.23	7.07
Second Half Period (from May, 2010 to May, 2020)										
Return (%, Ann)	2.34	0.75	2.04	2.71	2.58	2.99	2.86	4.17	2.62	3.18
Risk (%, Ann)	1.61	0.68	2.86	2.12	1.85	3.50	2.11	2.67	2.32	1.85
R/R	1.45	1.10	0.71	1.28	1.39	0.86	1.36	1.56	1.13	1.72
maxDD (%)	−4.79	−1.23	−8.11	−5.68	−5.59	−8.86	−5.71	−5.81	−5.37	−4.23

In terms of R/R for all period, XRP is the most efficient among RP, HRP and all CRPs. In addition, the return level of XRP is higher than all methods excluding CRP7. Also, maxDD of XRP is smaller than all methods excluding HRP. Our result shows that XRP has the best performance of all. XRP also gives the best R/R and second best maxDD in both the first half and the second half.

6. Conclusions

Our study makes the following contributions:

- We propose non-hierarchical clustering-risk parity strategy in which the risk contributions are equal both in each cluster and within the cluster.
- We also propose x-means++ algorithm which combines k-means++ algorithm with x-means algorithm to ensure robustness of clustering.
- Empirical analysis shows that the portfolio equalized risk contribution from each risk sources by our proposed approach, outperforms risk parity strategies or hierarchical clustering risk parity strategies.

Our future tasks are to perform empirical analysis using larger dataset such as individual stocks to verify the robustness of our proposed strategy and to apply our method to a complex valued risk diversification strategy.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Markowitz, H. (1952) Portfolio Selection. *Journal of Finance*, **7**, 77-91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

[2] Michaud, R.O. (1989) The Markowitz Optimization Enigma: Is "Optimized" Optimal? *Financial Analysts Journal*, **45**, 31-42. https://doi.org/10.2469/faj.v45.n1.31

[3] Qian, E. (2005) Risk Parity Portfolios: Efficient Portfolios through True Diversification. Panagora Asset Management, Boston.

[4] Nakagawa, K., Imamura, M. and Yoshida, K. (2018) Risk-Based Portfolios with Large Dynamic Covariance Matrices. *International Journal of Financial Studies*, **6**, 52. https://doi.org/10.3390/ijfs6020052

[5] Meucci, A. (2009) Managing Diversification. *Risk*, **22**, 74-79.

[6] Uchiyama, Y., Kadoya, T. and Nakagawa, K. (2019) Complex Valued Risk Diversification. *Entropy*, **21**, 119. https://doi.org/10.3390/e21020119

[7] Pelleg, D. and Moore, A.W. (2000) X-Means: Extending K-Means with Efficient Estimation of the Number of Clusters. *Proceedings of the* 17*th International Conference on Machine Learning*, June 2000, San Francisco, 727-734.

[8] Ishioka, T. (2006) An Expansion of X-Means: Progressive Iteration of K-Means and Merging of the Clusters. *Journal of Computational Statistics*, **18**, 3-13.

[9] Arthur, D. and Vassilvitskii, S. (2006) k-Means++: The Advantages of Careful Seeding. Stanford.

[10] De Prado, M.L. (2016) Building Diversified Portfolios that Outperform out of Sample. *The Journal of Portfolio Management*, **42**, 59-69.

https://doi.org/10.3905/jpm.2016.42.4.059

[11] Clarke, R.G., De Silva, H. and Thorley, S. (2006) Minimum-Variance Portfolios in the US Equity Market. *The Journal of Portfolio Management*, **33**, 10-24. https://doi.org/10.3905/jpm.2006.661366

[12] Choueifaty, Y. and Coignard, Y. (2008) Toward Maximum Diversification. *The Journal of Portfolio Management*, **35**, 40-51. https://doi.org/10.3905/JPM.2008.35.1.40

[13] Lee, W. (2011) Risk-Based Asset Allocation: A New Answer to an Old Question? *Journal of Portfolio Management*, **37**, 11-28. https://doi.org/10.3905/jpm.2011.37.4.011

[14] Jurczenko, E., Michel, T. and Teiletche, J. (2013) Generalized Risk-Based Investing.

[15] Martellini, L. and Ziemann, V. (2010) Improved Estimates of Higher-Order Comoments and Implications for Portfolio Selection. *The Review of Financial Studies*, **23**, 1467-1502.

[16] Nirwan, R.S. and Bertschinger, N. (2019) Applications of Gaussian Process Latent Variable Models in Finance. In: Bi, Y., Bhatia, R. and Kapoor, S., Eds., *Intelligent Systems and Applications*, *IntelliSys* 2019, *Advances in Intelligent Systems and Computing*, Springer, Cham, 1209-1221. https://doi.org/10.1007/978-3-030-29513-4_87

[17] Uchiyama, Y. and Nakagawa, K. (2020) TPLVM: Portfolio Construction by Student's *t*-Process Latent Variable Model. *Mathematics*, **8**, 449. https://doi.org/10.3390/math8030449

[18] Rockafellar, R.T. and Uryasev, S. (2000) Optimization of Conditional Value-at-Risk. *Journal of Risk*, **2**, 21-42. https://doi.org/10.21314/JOR.2000.038

[19] Nakagawa, K., Noma, S. and Abe, M. (2020) RM-CVaR: Regularized Multiple β-CVaR Portfolio. Proc*eedings of the* 29*th International Joint Conference on Artificial Intelligence* (*IJCAI* 2020), Kyoto, January 2020, 4562-4568. https://doi.org/10.24963/ijcai.2020/629

[20] Maillard, S., Roncalli, T. and Teïletche, J. (2010) The Properties of Equally Weighted Risk Contribution Portfolios. *The Journal of Portfolio Management*, **36**, 60-70. https://doi.org/10.3905/jpm.2010.36.4.060

[21] Roncalli, T. (2013) Introduction to Risk Parity and Budgeting. CRC Press, Boca Raton. https://doi.org/10.2139/ssrn.2272973

Notation

R: Return (random variable) vector

μ : Vector of expected returns

Σ : Covaraiance matrix

w: Weight vector of portfolio

σ_P : Portfolio risk

MRC: Marginal Risk Contribution

RC: Risk Contribution

C_i : Center point in cluster i

k: Number of clusters

n_i : Number of elements in C_i

$d(\cdot)$: Distance function

$f(x;\theta_i)$: p-dimensional normal distribution for the data x and paramater θ_i