Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

SUBSPACE PORTFOLIOS: DESIGN AND PERFORMANCE COMPARISON

by
Anqi Xiong

Data processing and engineering techniques enable people to observe and better understand the natural and human-made systems and processes that generate huge amounts of various data types. Data engineers collect data created in almost all fields and formats, such as images, audio, and text streams, biological and financial signals, sensing and many others. They develop and implement state-of-the art machine learning (ML) and artificial intelligence (AI) algorithms using big data to infer valuable information with social and economic value. Furthermore, ML/AI methodologies lead to automate many decision making processes with real-time applications serving people and businesses. As an example, mathematical tools are engineered for analysis of financial data such as prices, trade volumes, and other economic indicators of instruments including stocks, options and futures in order to automate the generation, implementation and maintenance of investment portfolios.

Among the techniques, subspace framework and methods are fundamental, and they have been successfully employed in widely used technologies and real-time applications spanning from Internet multimedia to electronic trading of financial products. In this dissertation, the eigendecomposition of empirical correlation matrix created from market data (normalized returns) for a basket of US equities plays a central role. Then, the merit of approximating such an empirical matrix by a Toeplitz matrix, where closed form solutions for its eigenvalues and eigenvectors exist, is investigated. More specifically, the exponential correlation model that populates such a Toeplitz matrix is used to approximate pairwise empirical correlations of asset returns in a portfolio. Hence, the analytically derived eigenvectors of such a

random vector process are utilized to design its eigenportfolios. The performances of the model based and the traditional eigenportfolios are studied and compared to validate the proposed portfolio design method. It is shown that the model based designs yield eigenportfolios that track the variations of the market statistics closely and deliver comparable or better performance.

The theoretical foundations of information theory and the rate-distortion theory that provide the basis for source coding methods, including transform coding, are revisited in the dissertation. This theoretical inquiry helps to construct the basic question of trade-offs between dimension of the eigensubspace versus the correlation structure of the random vector process it represents. The signal processing literature facilitates developing an efficient subspace partitioning algorithm to design novel portfolios by combining eigenportfolios of partitions for US equities that outperform the existing eigenportfolios (EP), market portfolios (MP), minimum variance portfolios (MVP), and hierarchical risk parity (HRP) portfolios for US equities. Additionally, the pdf-optimized quantizer framework is employed to sparse eigenportfolios in order to reduce the (trading) cost of their maintenance. Then, the concluding remarks are presented in the last section of the Dissertation.

SUBSPACE PORTFOLIOS: DESIGN AND PERFORMANCE COMPARISON

by
Anqi Xiong

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Electrical Engineering

Helen and John C. Hartmann Department of
Electrical and Computer Engineering

May 2020

APPROVAL PAGE

SUBSPACE PORTFOLIOS:
DESIGN AND PERFORMANCE COMPARISON

Anqi Xiong

Dr. Ali N. Akansu, Dissertation Advisor — Date
Professor of Electrical and Computer Engineering, NJIT

Dr. Ali Abdi, Committee Member — Date
Professor of Electrical and Computer Engineering, NJIT

Dr. Edip Niver, Committee Member — Date
Professor of Electrical and Computer Engineering, NJIT

Dr. Bipin Rajendran, Committee Member — Date
Reader of Engineering, King's College, London

Dr. Stephen M. Taylor, Committee Member — Date
Assistant Professor of Finance, NJIT

BIOGRAPHICAL SKETCH

Author: Anqi Xiong

Degree: Doctor of Philosophy

Date: May 2020

Undergraduate and Graduate Education:

- Doctor of Philosophy in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2020
- Master of Science in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2014
- Bachelor of Science in Electrical Engineering,
  Hunan University, Hunan, China, 2012

Major: Electrical Engineering

Presentations and Publications:

A. N. Akansu, A. Xiong, "Eigenportfolios of US Equities for Exponential Correlation Model," *Journal of Investment Strategies* 2020 (in press).

A. Xiong, A. N. Akansu, "On Sparsity of Eigenportfolios to Reduce Transaction Cost," *Journal of Capital Markets Studies*, Vol.3, No.1, pp 82-90, 2019.

A. N. Akansu, M. Avellaneda and A. Xiong, "Quant Investing in Cluster Portfolios," *Journal of Investment Strategies (in preparation for submission).*

A. Xiong, A. N. Akansu, "Performance Comparison of Minimum Variance, Market and Eigen Portfolios for US Equities," *53rd Annual Conference on Information Science and Systems (CISS)*, Baltimore, MD, 2019.

A. N. Akansu, A. Xiong, "Design of Eigenportfolios for US Equities Using Exponential Correlation Model," *53rd Annual Conference on Information Science and Systems (CISS)*, Baltimore, MD, 2019.

谨以此献给熊飞、贺建华
感谢你们的付出与支持

To my mother and father
because your endless support and encouragement
I can start my journey without any fear

ACKNOWLEDGMENT

The five years study at New Jersey Institute of Technology (NJIT) as a Ph.D. student, although it was a long journey with stress and lonely, I still enjoy and definitely will miss the time I spent here. Besides the knowledge, I have learned to be persistent and strong even in a hard time, which has been an unforgettable and valuable experience I gained. There are many people I would like to appreciate deeply from my heart. It was my pleasure to meet them in this journey and I feel so lucky to have their support.

First and foremost, I would like to convey my deep gratitude and respect to my advisor Prof. Ali N. Akansu for giving me the opportunity to start my doctoral study with him, for his continuously support and patience during my Ph.D. pursuit. He is always willing to share his immense knowledge with me both in professional domain and in life. I am and will always be grateful for him being my advisor and mentor not only in my Ph.D. study but also in my life.

Apart from my advisor, I would like to thank Prof. Stephen M. Taylor of the School of Management, who provided me the access to research facilities and shared his coding skills with me to collect the data for my research. I also would like to express the gratitude to Prof. Ali Abdi, Prof. Edip Niver of the ECE Department and Prof. Bipin Rajendran at Department of Engineering, King's College, London, for serving on my dissertation committee and for their insightful comments. Thank to the ECE Department for providing me the Teaching Assistantship that made this study possible.

I thank my labmates, Onur Yilmaz and Yanjia Sun who provided their help and support in my early Ph.D. study. To my friends, thank you for listening, offering me advice, and supporting me through this entire process, I look forward to continuing our friendships.

Finally, and perhaps most importantly, this journey would not have been possible without my parents, Fei Xiong and Jianhua He. Even you are on the opposite side of the world, when I have a hard time, it is your endless love and encouragement that helped me to have confidence and letting me to continue my adventure without any fear.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF TABLES
(Continued)

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

An investment portfolio is comprised of financial instruments such as stocks, bonds, futures, options, and others. It is designed based on risk limits and return expectations of investor by using historical market and relevant financial data. The portfolio manager allocates the total investment capital among the pre-selected assets and dynamically rebalances the allocations of the portfolio in order to minimize its risk for the targeted return performance [26, 4, 3].

Modern portfolio theory (MPT) builds on a mathematical method for portfolio optimization. It models the return of an asset as the normal (Gaussian) random variable and defines the investment risk as its standard deviation. Each asset in a portfolio has a weight, also called allocation coefficient, and the return of a portfolio is calculated as the weighted sum of asset returns, accordingly. Portfolio volatility is shown to be a function of pairwise correlations among asset returns. MPT provides closed-form solution for the risk optimization problem. The portfolio with minimum risk for the targeted return is called market portfolio (MP). Efficient frontier is generated by market portfolios on the risk-return plane. Similarly, the minimum variance portfolio (MVP) corresponds to the point with the minimum risk on the efficient frontier [26].

Financial signal processing (FSP) field is founded on signal processing methods and mathematical tools, and their engineering implementations with the focus on finance applications spanning from electronic trading to risk engineering. The engineers in FSP consider the financial data such as stock and option prices as signals. There is a great deal of overlap among techniques in signal processing and financial econometrics as revealed in the literature [3]. The objective is to infer useful

information from a set of noisy financial signals (market data) with the help of signal processing and statistical modeling methods with the goal of automating decision making processes for financial transactions [31, 4].

Subspace representation framework of linear algebra is one of the most powerful mathematical tools used for analysis and synthesis of data in many fields including machine learning, signal processing and finance. For example, the eigenanalysis, also known as Karhunen-Loève Transform (KLT) or principle component analysis (PCA), of correlation (covariance) matrix is performed to generate eigenvectors and eigenvalues as characteristics of a random vector process. They are used in signal processing, machine learning and pattern recognition for many tasks such as dimension reduction or lossy compression. The subspace methods are also utilized as mathematical tools for data visualization or data pre-processing before supervised learning is applied [16, 2, 14, 20, 8, 18].

Eigenanalysis of empirical correlation matrix calculated from returns of assets in a basket of instruments is used to design eigenportfolios where the portfolio returns are perfectly uncorrelated. [35, 7, 4].

This dissertation is structured as follows. In the next chapter, the mathematical preliminaries of this dissertation is introduced. The modern portfolio theory is revisited in Chapter 3. In Chapter 4, empirical correlations of asset returns are modeled as exponential function. Eigenportfolios as well as the MP and MVP are designed for such correlation matrix and their performances are evaluated and compared. The performances of model and empirical correlation matrices based eigenportfolios are also compared. It is demonstrated in the chapter that the exponential approximation to empirical correlations provide a good model to design eigenportfolios for U.S. equities. In Chapter 5, we develop an information theoretic portfolio sparsing method by designing pdf-optimized quantizers in order to reduce trading cost and show its merit with market data. In Chapter 6, a correlation based

subspace partitioning algorithm is developed. It is used to design novel protfolios by combining eigenportfolios of partitions. It is shown that these protfolios outperform the traditional eigenportfolios as well as known other portfolios such as MP, MVP, HRP and sector ETF for U.S. equities in terms of their PNLs and Sharpe ratios. The concluding remarks are expressed in the last chapter of the Dissertation.

CHAPTER 2

MATHEMATICAL PRELIMINARIES

In this chapter, the fundamental background of subspace transform are provided for the later discussion. The following sections include orthogonal subspace representation, eigen analysis of exponential correlation model, eigensubspace and gain of transform coding, pdf-optimized quantizer.

2.1 Orthogonal Subspace Representation

The orthogonal transforms have been used to process various types of data. The applications of such transforms spans from image processing, speech processing, analysis and design of communication system to feature selection and pattern recognition. Herein, we revisit the mathematical definitions of orthogonal subspace representation [1, 2].

Let $\mathbf{x}$ be an $N \times 1$ input signal

$$\mathbf{x} = [x_0, x_1, ..., x_{N-1}]^{\mathrm{T}}, \tag{2.1}$$

$\boldsymbol{\theta}$ be an $N \times 1$ coefficient

$$\boldsymbol{\theta} = [\theta_0, \theta_1, ..., \theta_{N-1}]^{\mathrm{T}}. \tag{2.2}$$

Let $\mathbf{\Phi}$ be an $N \times N$ orthogonal transform matrix

$$\Phi = [\boldsymbol{\phi}_0^{\mathrm{T}}, \boldsymbol{\phi}_1^{\mathrm{T}}, ..., \boldsymbol{\phi}_{N-1}^{\mathrm{T}}], \tag{2.3}$$

where $\boldsymbol{\phi}_k$ is an $N \times 1$ vector, and $\mathbf{\Phi}$ is an orthogonal transform, which means

$$\Phi\Phi^{-1} = \Phi\Phi^{\mathrm{T}} = \mathbf{I}, \tag{2.4}$$

where $\mathbf{I}$ is an identity matrix. We have the forward transform

$$\boldsymbol{\theta} = \Phi \mathbf{x} \tag{2.5}$$

and inverse transform

$$\mathbf{x} = (\Phi)^{-1} \boldsymbol{\theta} = \Phi^{\mathrm{T}} \boldsymbol{\theta} \tag{2.6}$$

to reconstruct the original signal.

2.2 Eigen Analysis of Exponential Correlation Model

Let the signal $\mathbf{x}$ defined in equation (2.1) comply with the exponential correlation model, which is defined as

$$R_{xx}(m) = E\left\{x_k x_l\right\} = \sigma_k \sigma_l \rho^{|m|};\ m = k - l;\ \forall\ k, l \tag{2.7}$$

where σ_k is the standard deviation of the signal x_k and ρ is the correlation coefficient with the range $-1 < \rho < 1$. Such a signal model has the correlation matrix expressed as a Toeplitz exponential correlation matrix

$$\mathbf{R} = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \cdots & 1 \end{bmatrix} \tag{2.8}$$

The inverse matrix is shown as [2]

$$\mathbf{R}^{-1} = \frac{1}{\beta^2} \begin{bmatrix} (1-\rho\alpha) & -\alpha & \cdots & \cdots & \cdots \\ -\alpha & 1 & -\alpha & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & -\alpha & 1 & -\alpha \\ \cdots & \cdots & 0 & -\alpha & (1-\rho\alpha) \end{bmatrix} \tag{2.9}$$

where

$$\beta^2 = \frac{1-\rho^2}{1+\rho^2}, \quad \alpha = \frac{\rho}{1+\rho^2} \tag{2.10}$$

The eigenvectors and eigenvalues of the (covariance or correlation) matrix $\mathbf{R}$ satisfy the set of equalities [41, 2]

$$\mathbf{R}\boldsymbol{\phi}_k = \lambda_k \boldsymbol{\phi}_k \tag{2.11}$$

Therefore, the eigen decomposition of the correlation matrix $\mathbf{R}$ is expressed as

$$\mathbf{R} = \mathbf{A}_{KLT}^T \boldsymbol{\Lambda} \mathbf{A}_{KLT} = \sum_{k=1}^{N} \lambda_k \boldsymbol{\phi}_k \boldsymbol{\phi}_k^T \tag{2.12}$$

the covariance of the coefficient vector $\boldsymbol{\theta}$ is diagonal matrix calculated as follows

$$\mathbf{R}_\theta = \mathbf{A}_{KLT} \mathbf{R} \mathbf{A}_{KLT}^T = \Lambda \tag{2.13}$$

where $\{\lambda_k, \boldsymbol{\phi}_k\}$ are the eigen pairs, and $\mathbf{A}_{KLT}$ with k^{th} column as $\boldsymbol{\phi}_k$, also known as Karhunen-Loève transform (KLT), defines the eigen subspace of $\mathbf{R}$ [2, 20].

The $N \times N$ eigenmatrix $\mathbf{A}_{KLT}$ is unitary, and inherently, it perfectly decorrelates transform coefficients as

$$E\{\theta_k \theta_l\} = \lambda_k \delta(k-l) = \sigma_{\theta_k}^2 \delta(k-l) \quad \forall k, l \tag{2.14}$$

The eigenvalues of Toeplitz matrix $\mathbf{R}$ as defined in (2.8) are expressed in the closed-form as [13, 33, 38]

$$\lambda_k = \frac{1-\rho^2}{1-2\rho\cos(\omega_k)+\rho^2}; 0 \leq k \leq N-1 \tag{2.15}$$

where $\{\omega_k\}$ are the positive roots of the equation

$$\tan(N\omega) = -\frac{(1-\rho^2)\sin(\omega)}{\cos(\omega) - 2\rho + \rho^2\cos(\omega)} \tag{2.16}$$

that is rewritten as [13, 38]

$$\begin{aligned} &\left[\tan\left(\omega\frac{N}{2}\right) + \gamma\tan\left(\frac{\omega}{2}\right)\right]\left[\tan\left(\omega\frac{N}{2}\right) - \frac{1}{\gamma}\cot\left(\frac{\omega}{2}\right)\right] = 0 \\ &\gamma = (1+\rho)\,/\,(1-\rho)\,. \end{aligned} \tag{2.17}$$

The roots of the above transcendental tangent equation, $\{\omega_k\}$, are required for the eigen kernel expression of such $N \times N$ Toeplitz matrix, and given as [33, 38]

$$\begin{aligned} &\mathbf{A}_{KLT} = [A(k,n)] = c_k \sin\left[\omega_k\left(n - \frac{N-1}{2}\right) + \frac{(k+1)\pi}{2}\right] \\ &c_k = \left(\frac{2}{N+\lambda_k}\right)^{1/2},\ 0 \le k, n \le N-1 \end{aligned} \tag{2.18}$$

An efficient root finding method for explicit solutions of transcendental equations (2.17) was investigated in [38]. It provides an efficient method to derive KLT kernel by utilizing the fast-fourier transform (FFT) algorithm.

2.3 Eigensubspace and Gain of Transform Coding

The gain of transform coding is a widely used metric in signal processing field to evaluate objective performance of a subspace. For the given signal type with the dimension N, G_{TC}^N over pulse code modulation (PCM), for the given $\mathbf{R}$ is defined as follows [2]

$$G_{TC}^N = \frac{\frac{1}{N}\sum_{k=0}^{N-1}\sigma_{\theta_k}^2}{\left(\prod_{k=0}^{N-1}\sigma_{\theta_k}^2\right)^{\frac{1}{N}}} \tag{2.19}$$

where $\sigma^2_{\theta_k}$ is the variance of k^{th} coefficients defined in equation (2.14).

The eigensubspace optimally repacks the signal energy into the largest L, $L \leq N$, coefficients (maximized explained variance) as expressed in the information theoretic performance metric [2].

It is closely related to the metric called explained variance. It is emphasized that eigensubspace is the optimum subspace that simultaneously satisfies the desired conditions of equations (2.14) and (2.19) where $\{\sigma^2_{\theta_k} = \lambda_k \, \forall k\}$ [2].

2.4 PDF-Optimized Quantizers

Quantization of coefficients in the transform domain, called transform coding (TC), is defined as [43]

$$\hat{\boldsymbol{\theta}}_{TC} = Q_\theta \left\{ \boldsymbol{\theta} \right\}, \tag{2.20}$$

the coefficient $\boldsymbol{\theta}$ has been quantized to $\hat{\boldsymbol{\theta}}_{TC}$

$$\tilde{\boldsymbol{\theta}}_{TC} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{TC}, \tag{2.21}$$

where $\tilde{\boldsymbol{\theta}}_{TC}$ is the quantization error due to the quantizer $Q_\theta \{\cdot\}$, and $\hat{\boldsymbol{\theta}}_{TC}$ is used to reconstruct the original signal by applying the inverse transform

$$\Phi^{\mathrm{T}} \hat{\boldsymbol{\theta}}_{TC} = \hat{\mathbf{x}}_{TC}, \tag{2.22}$$

so the reconstruction error can be written as

$$\tilde{\mathbf{x}}_{TC} = \mathbf{x} - \hat{\mathbf{x}}_{TC}. \tag{2.23}$$

The average mean square error between the original signal and reconstructed signal due to the quantization of coefficient is written as [2]

$$\sigma^2_{\epsilon,TC} = \frac{1}{N} E \left\{ \tilde{\mathbf{x}}^{\mathrm{T}}_{TC} \tilde{\mathbf{x}}_{TC} \right\} = \frac{1}{N} \sum_{k=0}^{N-1} \sigma^2_{\epsilon k}. \tag{2.24}$$

Similarly, the average mean square error between the original coefficient and quantized coefficient is

$$\sigma_{q,TC}^2 = \frac{1}{N} E\left\{\tilde{\boldsymbol{\theta}}_{TC}^{\mathrm{T}} \tilde{\boldsymbol{\theta}}_{TC}\right\} = \frac{1}{N} \sum_{k=0}^{N-1} \sigma_{qk}^2. \tag{2.25}$$

Noted that the mean square error of the signal and coefficient are the same, since

$$E\left\{\tilde{\mathbf{x}}_{TC}^{\mathrm{T}} \tilde{\mathbf{x}}_{TC}\right\} = E\left\{\tilde{\mathbf{x}}_{TC}^{\mathrm{T}} \Phi^{\mathrm{T}} \Phi \tilde{\mathbf{x}}_{TC}\right\} = E\left\{\tilde{\boldsymbol{\theta}}_{TC}^{\mathrm{T}} \tilde{\boldsymbol{\theta}}_{TC}\right\}. \tag{2.26}$$

So we have no doubt with

$$\sigma_{\epsilon,TC}^2 = \sigma_{q,TC}^2 \tag{2.27}$$

The Lloyd-Max quantizer minimizes separately each σ_{qk}^2, and so does $\sum_k \sigma_{qk}^2$. It minimizes the quantization error in the mean square sense. The pdf optimized quantizer is calculated iteratively as described in [28, 23]. The random input x has a pdf $f(x)$ with zero-mean and unit variance, the end values of N-level quantizer are x_k and x_{k+1}, and y_k represents all numbers fall into the k^{th} interval (bin) $[x_k, x_{k+1}]$, where $k = 1, 2, ..., N$, and $x_1 = -\infty$ and $x_{N+1} = \infty$. The quantization error in mse for such quantizer is calculated as

$$\sigma_q^2 = \sum_{k=1}^{N} \int_{x_k}^{x_{k+1}} (x - y_k)^2 f(x)\, dx. \tag{2.28}$$

The Lloyd-Max quantizer design algorithm updates the intervals $[x_k, x_{k+1}]$ and y_k iteratively, by satisfying two conditions [28, 23]

$$\begin{aligned} \frac{\partial \sigma_q^2}{\partial x_k} &= 0;\ k = 2, 3, ..., N \\ \frac{\partial \sigma_q^2}{\partial y_k} &= 0;\ k = 1, 2, 3, ..., N \end{aligned} \tag{2.29}$$

which also implies that

$$x_k = \frac{1}{2}(y_k + y_{k-1});\ k = 2, 3, ..., N$$
$$y_k = \frac{\int_{x_k}^{x_{k+1}} x f(x)\, dx}{\int_{x_k}^{x_{k+1}} f(x)\, dx};\ k = 1, 2, 3, ..., N \tag{2.30}$$

The first order entropy for an N output-level pdf-optimized quantizer is calculated as [43]

$$H = -\sum_{k=1}^{N} F_k log_2 F_k$$
$$F_k = \int_{x_k}^{x_{k+1}} f(x)\, dx. \tag{2.31}$$

Note that the noise variances of all bins are the same in a pdf-optimized quantizer.

2.5 Chapter Summary

The orthogonal subspace has been used for signal processing and quantitative finance, in this chapter, the forward and inverse transform with the orthogonal subspace is discussed. The exponential correlation model that describe the auto-correlation of an N dimensional signal is defined. A close form expression of eigenvectors and the corresponding eigenvalues for a Toeplitz correlation matrix that generated from the model is summarized. They will be used to design the eigenportfolios in Chapter 4. The eigensubspace that maximized the gain of transform coding G_{TC}, this parameter will also be utilized to design the subband portfolios in Chapter 6. At last, the pdf-optimized quantizers is introduced, in Chapter 5, it will be applied to sparse the eigenportfolios.

CHAPTER 3

MODERN PORTFOLIO THEORY

In order to have a better illustration of subspace portfolio design problem and compare the performance with other well known portfolios in the following chapters, the fundamental concept of famous Modern Portfolio Theory (MPT) are briefly introduced, it suggests a mathematical method for portfolio optimization and provides the solution. Basic knowledge in finance include asset and portfolio return, risk (volatility) and Sharpe ratio are defined and the derivation of two optimal portfolios are showed in this chapter.

3.1 N-Asset Portfolio

The normalized return of the k^{th} asset in N-asset portfolio at the discrete-time n is defined as [4, 39]

$$r_k(n) = \frac{p_k(n)}{p_k(n-1)} - 1;\ k = 1, 2, \ldots, N \tag{3.1}$$

where $p_k(n)$ is its price. The mean and variance of $r_k(n)$ are calculated with the ergodicity assumption for a measurement window of W samples [30]

$$\mu_k = E\{r_k(n)\} = \frac{1}{W}\sum_{m=0}^{W-1} r_k(n-m) \tag{3.2}$$

$$\sigma_k^2 = E\left\{r_k^2(n)\right\} - \mu_k^2 = \left[\frac{1}{W}\sum_{m=0}^{W-1} r_k^2(n-m)\right] - \mu_k^2 \tag{3.3}$$

where σ_k is the volatility of the k^{th} asset. The return vector of portfolio assets is defined as

$$\mathbf{r}(n) = [r_k(n)]\,; k = 1, 2, \ldots, N \tag{3.4}$$

The sampling time index n is omitted for convenience in the following discussions. The mean vector of asset returns, $\boldsymbol{\mu}$, is of $N \times 1$ and populated with the expected asset returns

$$\boldsymbol{\mu} = E\{\mathbf{r}\} = [E\{r_k(n)\}] = [\mu_k] \tag{3.5}$$

Then, the diagonal risk matrix is defined as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N \end{bmatrix} \tag{3.6}$$

It is $N \times N$ with elements $\{\sigma_k\}$. Hence, the risk normalized return also known as Sharpe ratio vector of assets is expressed as

$$\mathbf{S} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \begin{bmatrix} \frac{\mu_1}{\sigma_1} \\ \frac{\mu_2}{\sigma_2} \\ \vdots \\ \frac{\mu_N}{\sigma_N} \end{bmatrix} \tag{3.7}$$

The normalized asset returns are calculated as

$$\widehat{\mathbf{r}} = [\widehat{r}_k] = \left[\frac{r_k - \mu_k}{\sigma_k}\right] = \boldsymbol{\Sigma}^{-1}(\mathbf{r} - \boldsymbol{\mu})\,. \tag{3.8}$$

The return of an N-asset portfolio is found as

$$r_p = \mathbf{q}^{\mathrm{T}}\mathbf{r} \tag{3.9}$$

where $\mathbf{q} = [q_k]\,; k = 1, 2, \ldots, N$ is the investment allocation vector. Portfolio risk is defined as follows [39]

$$\sigma_p = \left(E\left\{r_p^2\right\} - \mu_p^2\right)^{1/2} = \left(\mathbf{q}^{\mathrm{T}}\mathbf{C}\mathbf{q}\right)^{1/2} = \left(\mathbf{q}^{\mathrm{T}}\boldsymbol{\Sigma}^{\mathrm{T}}\mathbf{R}\boldsymbol{\Sigma}\mathbf{q}\right)^{1/2} \tag{3.10}$$

where $\mu_p = E\left\{r_p\right\} = \mathbf{q}^{\mathrm{T}}E\left\{\mathbf{r}\right\} = \mathbf{q}^{\mathrm{T}}\boldsymbol{\mu}$ is the expected return of the portfolio, $\mathbf{C}$ is $N \times N$ covariance matrix of asset returns, and $\mathbf{R}$ is their $N \times N$ correlation matrix with elements $[R_{ij}] = E\{\widetilde{r}_i\widetilde{r}_j\} = \rho_{ij}$.

3.2 Portfolio Optimization

Modern Portfolio Theory (MPT) introduces a method to create an optimal portfolio by minimizing the portfolio risk σ_p with the constraints of constant expected portfolio return $\mu_p = \mu$ and normalized investment allocation vector [26, 4]

$$\begin{aligned} \min \quad & \mathbf{q}^{\mathrm{T}}\mathbf{C}\mathbf{q} \\ \mathbf{s.t.} \quad & \mathbf{q}^{\mathrm{T}}\boldsymbol{\mu} = \mu \\ & \mathbf{q}^{\mathrm{T}}\mathbf{1} = 1 \end{aligned} \tag{3.11}$$

where $\mathbf{1}$ is an $N \times 1$ vector with elements equal to 1.

The risk minimization problem is solved by introducing two Lagrangian multipliers as

$$L\left(\mathbf{q}, \lambda_1, \lambda_2\right) = \frac{1}{2}\mathbf{q}^T\mathbf{C}\mathbf{q} + \lambda_1\left(\mu - \mathbf{q}^T\boldsymbol{\mu}\right) + \lambda_2\left(1 - \mathbf{q}^T\mathbf{1}\right) \tag{3.12}$$

The solution for the optimum investment allocation vector is obtained as [26]

$$\mathbf{q}^* = \frac{\begin{vmatrix} \mu & \mathbf{1}^T\mathbf{C}^{-1}\boldsymbol{\mu} \\ 1 & \mathbf{1}^T\mathbf{C}^{-1}\mathbf{1} \end{vmatrix}\mathbf{C}^{-1}\boldsymbol{\mu} + \begin{vmatrix} \boldsymbol{\mu}^T\mathbf{C}^{-1}\boldsymbol{\mu} & \mu \\ \boldsymbol{\mu}^T\mathbf{C}^{-1}\mathbf{1} & 1 \end{vmatrix}\mathbf{C}^{-1}\mathbf{1}}{\begin{vmatrix} \boldsymbol{\mu}^T\mathbf{C}^{-1}\boldsymbol{\mu} & \mathbf{1}^T\mathbf{C}^{-1}\boldsymbol{\mu} \\ \boldsymbol{\mu}^T\mathbf{C}^{-1}\mathbf{1} & \mathbf{1}^T\mathbf{C}^{-1}\mathbf{1} \end{vmatrix}}. \tag{3.13}$$

This solution forms a curve on the (σ,μ) plane that is called Markowitz bullet [26, 4].

3.2.1 Minimum Variance Portfolio

The investment allocation vector for the minimum variance portfolio (MVP) is derived by solving the following optimization problem

$$\begin{aligned} \min \quad & \mathbf{q}^{\mathrm{T}}\mathbf{C}\mathbf{q} \\ \mathbf{s.t.} \quad & \mathbf{q}^{\mathrm{T}}\mathbf{1} = 1 \end{aligned} \tag{3.14}$$

by using the Lagrangian multiplier of the form

$$L\left(\mathbf{q},\lambda\right) = \frac{1}{2}\mathbf{q}^{T}\mathbf{C}\mathbf{q} + \lambda\left(1 - \mathbf{q}^{T}\mathbf{1}\right) \tag{3.15}$$

The solution is found as [26, 4]

$$\mathbf{q}_{min} = \frac{\mathbf{C}^{-1}\mathbf{1}}{\mathbf{1}^{T}\mathbf{C}^{-1}\mathbf{1}} \tag{3.16}$$

The return of the minimum variance portfolio is found as [42]

$$r_{min} = \mathbf{q}_{min}^{T}\mathbf{r}, \tag{3.17}$$

where, the mean and variance of its return are calculated as

$$\mu_{min} = \mathbf{q}_{min}^{T}\boldsymbol{\mu} = \frac{\boldsymbol{\mu}^{T}\mathbf{C}^{-1}\mathbf{1}}{\mathbf{1}^{T}\mathbf{C}^{-1}\mathbf{1}} \tag{3.18}$$

$$\sigma_{min}^{2} = \mathbf{q}_{min}^{T}\mathbf{C}\mathbf{q}_{min} = \frac{1}{\mathbf{1}^{T}\mathbf{C}^{-1}\mathbf{1}}. \tag{3.19}$$

The Sharpe ratio of the minimum variance portfolio for the given $\mathbf{C}$ and $\boldsymbol{\mu}$ is expressed as

$$S_{min} = \frac{\mu_{min}}{\sigma_{min}} = \frac{\boldsymbol{\mu}^{T}\mathbf{C}^{-1}\mathbf{1}}{\left(\mathbf{1}^{T}\mathbf{C}^{-1}\mathbf{1}\right)^{\frac{1}{2}}} \tag{3.20}$$

3.2.2 Market Portfolio

Market portfolio (MP) maximizes the risk normalized return, Sharpe Ratio, with the constraint of $\mathbf{q}^T\mathbf{1} = 1$ as

$$\begin{aligned} \max \quad & \frac{\mathbf{q}^T\boldsymbol{\mu}}{(\mathbf{q}^T\mathbf{C}\mathbf{q})^{\frac{1}{2}}} \\ \mathbf{s.t.} \quad & \mathbf{q}^{\mathrm{T}}\mathbf{1} = 1 \end{aligned} \tag{3.21}$$

This problem can be solved by using the Lagrangian as

$$L(\mathbf{q}, \lambda) = \frac{\mathbf{q}^T\boldsymbol{\mu}}{(\mathbf{q}^T\mathbf{C}\mathbf{q})^{\frac{1}{2}}} + \lambda\left(1 - \mathbf{q}^T\mathbf{1}\right), \tag{3.22}$$

and set its partial derivatives, with respect to $\mathbf{q}$ and λ, to be zero lead to the solution of the market portfolio [26, 4]

$$\mathbf{q}_m = \frac{\mathbf{C}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\mathbf{C}^{-1}\boldsymbol{\mu}}. \tag{3.23}$$

Note that for a constant $\boldsymbol{\mu}$, the minimum variance and market portfolios are identical. The return of the market portfolio is calculated as

$$r_m = \mathbf{q}_m^T\mathbf{r} \tag{3.24}$$

with its mean and variance

$$\mu_m = \mathbf{q}_m^T\boldsymbol{\mu} = \frac{\boldsymbol{\mu}^T\mathbf{C}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\mathbf{C}^{-1}\boldsymbol{\mu}} \tag{3.25}$$

$$\sigma_m^2 = \mathbf{q}_m^T\mathbf{C}\mathbf{q}_m = \frac{\boldsymbol{\mu}^T\mathbf{C}^{-1}}{\mathbf{1}^T\mathbf{C}^{-1}\boldsymbol{\mu}}\mathbf{C}\frac{\mathbf{C}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\mathbf{C}^{-1}\boldsymbol{\mu}} = \frac{\boldsymbol{\mu}^T\mathbf{C}^{-1}\boldsymbol{\mu}}{\left(\mathbf{1}^T\mathbf{C}^{-1}\boldsymbol{\mu}\right)^2}. \tag{3.26}$$

Therefore, the Sharpe ratio of the market portfolio for the given $\mathbf{C}$ and $\boldsymbol{\mu}$ is expressed as

$$S_m = \frac{\mu_m}{\sigma_m} = sign\left[\mathbf{1}^T\mathbf{C}^{-1}\boldsymbol{\mu}\right]\left(\boldsymbol{\mu}^T\mathbf{C}^{-1}\boldsymbol{\mu}\right)^{\frac{1}{2}}, \tag{3.27}$$

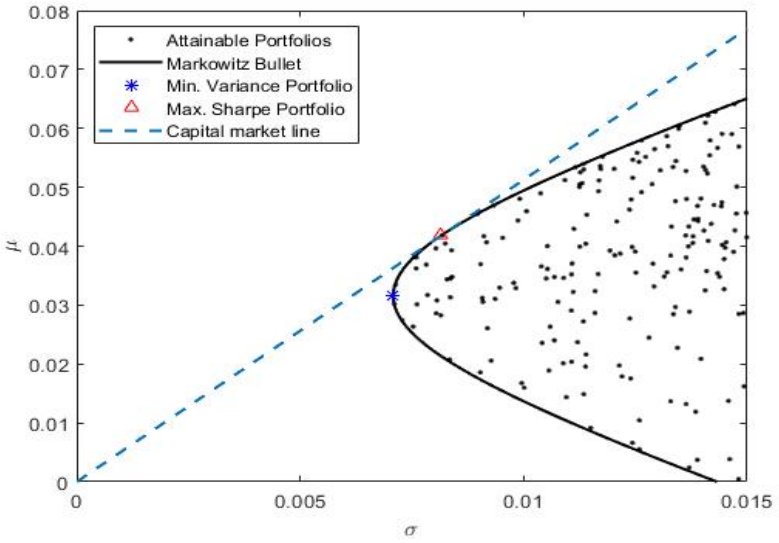


Figure 3.1 Markowitz bullet. All the attainable portfolios (weight randomly generated with $\mathbf{q}^{\mathrm{T}}\mathbf{1} = 1$) lie in and on the frontier. Maximum Sharpe portfolio is located at the tangency point of CML to Markowitz bullet with $\rho_{12} = 0.6$, $\rho_{13} = 0.2$, $\rho_{23} = 0.7$, $\mu_1 = 0.07$, $\mu_2 = 0.03$, and $\mu_3 = 0.02$, $\sigma_1 = 0.02$, $\sigma_2 = 0.03$, $\sigma_3 = 0.01$.

where $sign\,[\cdot]$ denotes the sign of the result in the square brackets.

In Figure 3.1 The line that connects the risk free asset and market portfolio is called capital market line (CML). It can be shown that tangency point of the CML to the Markowitz Bullet is the market portfolio [4].

3.3 Chapter Summary

The Modern Portfolio Theory (MPT) is briefly revisited in this chapter. It provides the framework to evaluate the portfolio return and risk (volatility) in an analytic approach by using the expectation value and covariance of N-asset returns. This framework is also utilized in the following chapters to make the evaluation and comparisons of eigenportfolios. The MPT also enables us to design the two optimal portfolios, namely, the minimum variance portfolio (MVP) and the market portfolio (MP), the analytic solution of them are shown, the closed-form expression of their expected return and risk (volatility) are given in this chapter as well.

CHAPTER 4

EIGENPORTFOLIOS DESIGN AND COMPARISONS

Eigenanalysis of covariance (correlation) matrix describing a random vector process, also called principal component analysis (PCA) or Karhunen-Loève transform (KLT), has been successfully employed in various fields spanning from signal processing to quantitative finance [32, 21, 24, 41, 2, 20, 12, 34]. In this chapter, we will focus on eigenportfolios that are generated from the empirical correlation matrix of asset returns in a basket of instruments and approximate it by a Toeplitz matrix whose elements are defined by an exponential function. This approach enables us to use the currently available closed form expressions for the eigenvectors and eigenvalues of such matrix type to build the model based eigenportfolios and study their performance [13, 33, 2, 38]. This is an extension of the earlier work where a framework is presented to analyze performance of eigenportfolios based on metrics like Sharpe ratio (risk-normalized return) and market exposure. This chapter show its merit for 28-stock basket of DJIA index for EOD returns from July 1, 1999 to November 1, 2018, and its subintervals, and three other baskets[1]. [37, 5, 38, 44].

Asset return vector of a basket is often modeled as jointly normal random vector process [40]. Although histograms of asset returns for market data show fat-tails, excessive kurtosis and asymmetry properties, the normality assumption makes statistical modeling and analysis tractable. Eigen decomposition of the covariance matrix of a normal vector process defines its eigen subspace with the corresponding eigenvectors (and the eigenvalues) where the projection coefficients, also known as the transform or eigencoefficients, of the process have zero cross-correlations. This is an important property to achieve the statistical independence of eigencoefficients It

[1]Each basket is comprised of the most allocated 15 stocks for the sector ETFs XLF (Finance), XLI (Industrial) and XLV (Health Care).

is emphasized that these eigencoefficients are equivalent to eigenportfolio returns in this analysis. Therefore, eigen decomposition of empirical correlation matrix of asset returns results in eigenvectors, and their components are used as the capital allocation coefficients of N eigenportfolios with perfectly decorrelated portfolio returns. Note that signs of eigenvector components bear significant information. These portfolios with uncorrelated returns are particularly used as the independent variables of regression to predict asset returns in trading strategies like statistical arbitrage [7, 4].

Market exposure dictates the market risk of a portfolio. Hence, market-neutrality is a desired feature for lower risk portfolios. Some of the eigenportfolios may inherently be considered as market-neutral investment tools due to the zero mean of their allocation coefficients. They are expected not to be highly affected from the market fluctuations. For a typical basket of correlated assets, most eigenportfolios are almost uncorrelated from the market moves, except the first eigenportfolio with mostly long positions (high market exposure), and a few others with their smaller non-zero means of asset allocations. Momentum based strategies like index investing aim to mimic the market where the first eigenportfolio with high market exposure may serve the purpose [34]. In contrast, most of the remaining eigenportfolios have their built-in self-hedging against the market trend [7].

In this chapter, The exponential correlation to model co-movements of asset returns for a basket, and the kernels for its eigenvectors and eigenvalues are introduced. The performance of eigenportfolios designed by using those eigenvectors are investigated. The performances of the model and the measurement based eigenportfolios are presented to highlight the merit of the exponential approximation to the correlations of returns. PNLs and Sharpe ratios of the minimum variance (MVP), market (MP) and eigenportfolios (EP) along with several ETFs is displayed to compare their market performances [26, 4]. Sharpe ratios of eigenportfolios that are designed by eigen decomposition of model based and empirical measurements

based correlation matrices are also derived for several independent subintervals of the same basket. In addition, similar performance comparisons are made for three different stock baskets in order to emphasize the robustness and merit of the proposed framework to analyze eigenportfolios.

4.1 Exponential Modeling of Return Correlations and Eigen Decomposition of Toeplitz Matrix

The historical cross-correlations of normalized returns for assets k and l at time n, with a measurement window of W past samples are modeled as follows

$$R_{rr}(m) = E\left\{r_k(n)r_l(n)\right\} = \frac{1}{W}\sum_{n=0}^{W-1} r_k(n)r_l(n) = \sigma_k^2\rho^{|m|};\ m = k-l;\ \forall\ k,l \qquad (4.1)$$

with the correlation coefficient $-1 < \rho < 1$. The resulting correlation matrix of size $N \times N$, for normalized asset volatilities $\sigma_k = \sigma_l = 1, \forall\ k,l$, is real, symmetric and Toeplitz as (2.8) shows.

As a special case, $\mathbf{R}$ represents Gaussian wide sense stationary (WSS) stochastic process with multivariate normal distribution $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{R})$ for finite dimensions where [30, 22]

$$\boldsymbol{\mu} = [\mu_k = \mu]\,; k = 1, 2, \ldots, N \qquad (4.2)$$

One may focus on the general case where $\mu_k \neq \mu_l\ \forall\, k,l$ that is more realistic in most applications.

4.1.1 Eigenportfolios of N-assets

Eigenanalysis of empirical correlation matrix $\mathbf{R}_E$ of returns in a basket of instruments enables us to build a set of eigenportfolios with perfectly decorrelated returns [4, 2]. The decorrelation of eigenportfolio returns is a desirable feature to design certain types of portfolios and trading algorithms [7]. Moreover, except the first

eigenportfolio with long only positions in a typical scenario, eigenportfolios may have market-neutrality whenever their eigenvectors, capital allocation coefficients with long and short positions for assets with similar correlations to the market, have zero mean. A market neutral portfolio is self-hedged against market fluctuations, and it may be used in low-risk investment strategies. The design steps of eigenportfolios are summarized as follows.

The return vector at time n is created as $\mathbf{r} = [r_k]\,; k = 1, 2, \ldots, N$ [39, 5]. Each asset return is normalized to be zero mean and unit variance as defined in equation (3.8). The covariance (correlation) matrix ($\mathbf{C} = \mathbf{R}$ due to normalization) of returns is expressed as

$$\mathbf{R}_E \triangleq \left[E\left\{\widehat{\mathbf{r}}\widehat{\mathbf{r}}^T\right\}\right] = [R_{k,l}] \tag{4.3}$$
$$= \begin{bmatrix} R_{1,1} & R_{1,2} & \cdots & R_{1,N} \\ R_{2,1} & R_{2,2} & \cdots & R_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ R_{N,1} & R_{N,2} & \cdots & R_{N,N} \end{bmatrix}$$

with the elements calculated as

$$[R_{k,l}] = E\left\{\widehat{r}_k \widehat{r}_l\right\} = \frac{1}{W}\sum_{m=0}^{W-1} \widehat{r}_k(n-m)\widehat{r}_l(n-m) \tag{4.4}$$

It represents measured cross-correlations for the historical data window size of W samples. It is noted that the value of W impacts the validity of the stationarity assumption in that empirical measurement interval. $\mathbf{R}_E$ is assumed to be a real, symmetric and positive definite matrix. The first row of empirical correlation matrix for EOD returns of 28 stocks[2] in DJIA index with $W = 40$ on July 25, 2005, and its approximation by exponential correlation model are shown in Figure 4.1 as an example. The significance of descending ordering of measured correlations for

[2]AAPL, AXP, BA, CAT, CSCO, CVX, DIS, DWDP, GS, HD, IBM, INTC, JNJ, JPM, KO, MCD, MMM, MRK, MSFT, NKE, PFE, PG, TRV, UNH, UTX, VZ, WMT, XOM.

improved model fitting is emphasized here. The variation of the first order correlation coefficient ρ for the time period from July 1, 1999 to Nov. 1, 2018, with $W = 40$, is displayed in Figure 4.2.

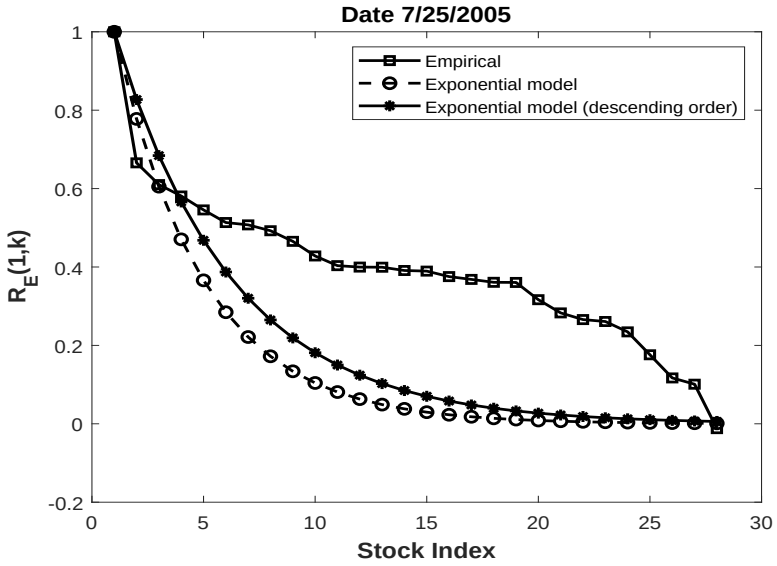


Figure 4.1 First row of empirical correlation matrix $\mathbf{R}_E$ (in descending order) for EOD returns of 28 stocks in DJIA index with $W = 40$ on July 25, 2005 along with its approximations by exponential correlation model for the descending and the randomly ordered empirical correlations

The eigen decomposition of $\mathbf{R}_E$ is performed according to equation (2.12). Then, the eigenvectors $\{\phi_k\}\, k = 1, 2, ..., N$ are used to create the eigenportfolios. The vector components $\{\phi_k(n),\ n = 1, 2, ..., N\}$ are repurposed as the capital allocation coefficients of the k^{th} eigenportfolio [2, 41].

4.2 Performance Analysis of Eigenportfolios

4.2.1 Performance of N-asset Eigenportfolios

The correlation matrix of the returns is expressed as $\mathbf{R} = E\left\{\widehat{\mathbf{r}}\widehat{\mathbf{r}}^T\right\} = \mathbf{\Sigma}^{-1}\mathbf{C}\mathbf{\Sigma}^{-1}$ where $\mathbf{C}$ is the covariance, and $\mathbf{\Sigma}$ is the risk matrix of asset returns as defined in equation (3.6). The eigen decomposition of correlation matrix is shown in equation (2.12).

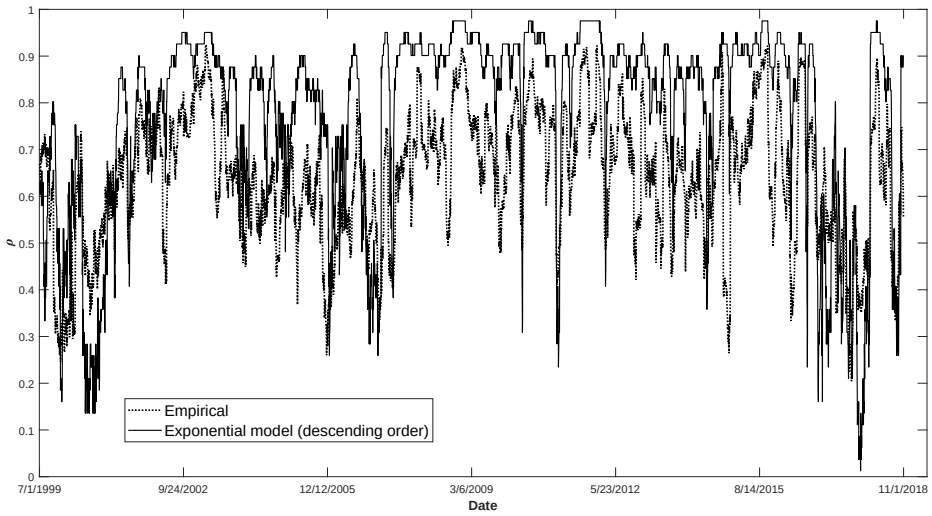


Figure 4.2 Time variation of ρ for empirical and exponential model cases using EOD returns of 28 stocks in DJIA index for the time period from July 1, 1999 to Nov. 1, 2018 with $W = 40$. Their means and standard deviations for the entire period are $\mu_\rho^E = 0.64$ and $\sigma_\rho^E = 0.142$, and $\mu_\rho^{EM} = 0.77$ and $\sigma_\rho^{EM} = 0.2$ for empirical and exponential model cases, respectively.

The eigenvectors $\{\phi_k\}$ are modified twice in order to generate eigenportfolios. First, the components of eigenvectors are normalized by the volatilities of the corresponding asset returns. It yields risk normalized capital allocation coefficients for assets of eigenportfolios. Although this normalization destroys the orthogonality of the original eigenvectors, the perfect decorrelation property of eigenportfolio returns is still preserved. Second, summation of the absolute values of risk normalized capital allocation coefficients, the total investment amount, is normalized to one for each eigenportfolio. This normalization also does not compromise the perfect decorrelation property of their returns. The resulting eigen matrix with such modifications is written as

$$\widetilde{\mathbf{A}}_{KLT} = \mathbf{\Sigma}^{-1}\mathbf{A}_{KLT}\mathbf{\Psi}^{-1} \tag{4.5}$$

where $\mathbf{\Sigma}$ is given in equation (3.6), and $\mathbf{\Psi}$ is the diagonal normalization matrix with elements $\psi_k = \sum_n \left|\frac{\phi_k(n)}{\sigma_n}\right| \;\; k = 1, 2, ..., N$. Then, the eigenportfolio returns are calculated as

$$\mathbf{r}^{ep} = \widetilde{\mathbf{A}}_{KLT}^{T}\mathbf{r} \tag{4.6}$$

The mean values of eigenportfolio returns are derived as

$$\boldsymbol{\mu}^{ep} = E\left\{\mathbf{r}^{ep}\right\} = E\left\{\widetilde{\mathbf{A}}_{KLT}^{T}\mathbf{r}\right\} = \boldsymbol{\Psi}^{-1}\mathbf{A}_{KLT}^{T}\mathbf{S} \tag{4.7}$$

where the Sharpe ratio vector of assets $\mathbf{S}$ is given in equation (3.7). Therefore, the diagonal covariance matrix of eigenportfolio returns is expressed as

$$\begin{aligned}
\mathbf{C}^{ep} &= E\left\{\left(\mathbf{r}^{ep} - \boldsymbol{\mu}^{ep}\right)\left(\mathbf{r}^{ep} - \boldsymbol{\mu}^{ep}\right)^{T}\right\} \\
&= E\left\{\widetilde{\mathbf{A}}_{KLT}^{T}\left(\mathbf{r} - \boldsymbol{\mu}\right)\left(\mathbf{r} - \boldsymbol{\mu}\right)^{T}\widetilde{\mathbf{A}}_{KLT}\right\} \\
&= \boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}\boldsymbol{\Psi}^{-1} \\
&= \begin{bmatrix} \sigma^2_{r_1^{ep}} & 0 & \cdots & 0 \\ 0 & \sigma^2_{r_2^{ep}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{r_N^{ep}} \end{bmatrix} = \begin{bmatrix} \frac{\lambda_1}{\psi_1^2} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{\psi_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_N}{\psi_N^2} \end{bmatrix}
\end{aligned} \tag{4.8}$$

where $\left[\sigma^2_{r_k^{ep}}\right]$ is the variance of the k^{th} eigenportfolio return. Hence, the diagonal risk matrix of eigenportfolio returns is calculated as

$$\boldsymbol{\Sigma}^{ep} = \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Psi}^{-1} \tag{4.9}$$

where its k^{th} element is shown to be $\sigma_{r_k^{ep}} = \frac{\sqrt{\lambda_k}}{\psi_k}$. Then, the diagonal correlation matrix of the normalized eigenportfolio returns is derived as

$$\mathbf{R}^{ep} = \boldsymbol{\Sigma}^{ep^{-1}}\mathbf{C}^{ep}\boldsymbol{\Sigma}^{ep^{-1}} = \mathbf{I} \tag{4.10}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix.

Similarly, the Sharpe ratios of the eigenportfolio returns are calculated in the matrix form as follows

$$\boldsymbol{S}^{ep} = \boldsymbol{\Sigma}^{ep^{-1}}\boldsymbol{\mu}^{ep} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{A}_{KLT}^{T}\mathbf{S} \tag{4.11}$$

where $\mathbf{S} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ as shown in equation (3.7). Hence,

$$\boldsymbol{S}^{ep} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{A}_{KLT}^{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \tag{4.12}$$

4.2.2 Market Exposure of Eigenportfolio

Market exposure is the amount of investment exposed to the market risk (unhedged investment against market fluctuations) and calculated as

$$\boldsymbol{M}^{ep} = [M_k^{ep}]\,; k = 1, 2, \ldots, N \tag{4.13}$$

$$M_k^{ep} = \sum_{n=1}^{N} \widetilde{\boldsymbol{\phi}}_k(n) \tag{4.14}$$

where M_k^{ep} is the market exposure of the k^{th} eigenportfolio (k^{th} column of $\widetilde{\mathbf{A}}_{KLT}$), and $\widetilde{\boldsymbol{\phi}}_k(n)$ is the n^{th} component of the k^{th} eigenportfolio. Note that this metric assumes all assets have the same cross-correlation (co-movement) with the market. It can easily be extended for the case of uneven cross-correlations.

4.3 Eigenportfolio Performance

Eigenportfolios of exponential correlation model are evaluated with respect to Sharpe ratios of their returns, profit and loss (PNL) curves, and market exposures. As the first case, we created the empirical correlation matrix for a basket of 28 DJIA stocks, and generated its approximation by using exponential correlation model in order to show the merit of the proposed framework to design and evaluate the resulting eigenportfolios.

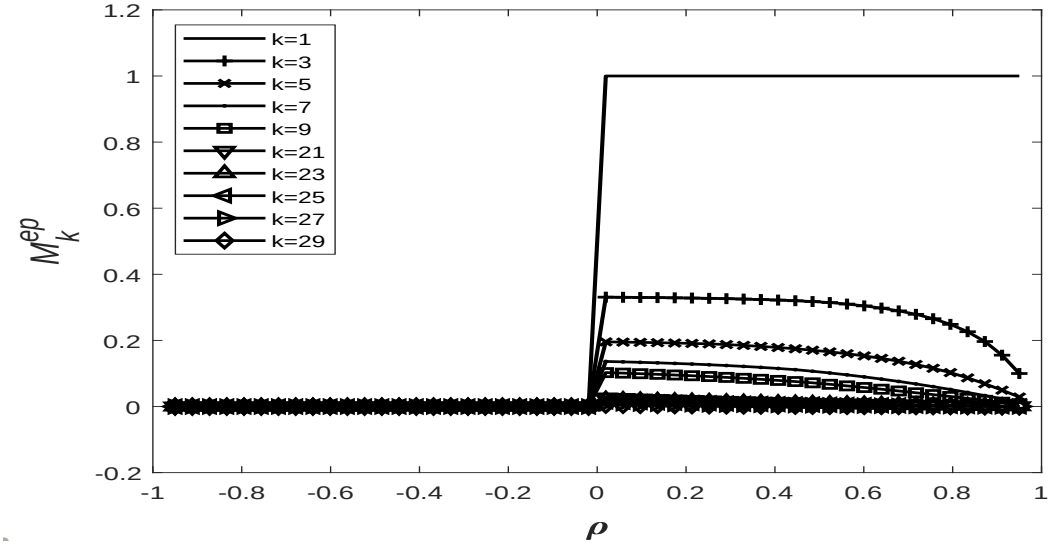
4.3.1 Performance of Eigenportfolios Based on Exponential Correlation Model

The variations of market exposures M_k^{ep}, and Sharpe ratios S_k^{ep} of the first and the last five odd indexed eigenportfolios generated from the exponential correlation model, with $\mu_k = 1$, $\sigma_k = 1$ and $N = 30$, as a function of ρ are displayed in Figures 4.3a and 4.3b, respectively. It is noted that the first eigenportfolio has *long* positions for all assets, and it has the highest market exposure according to equation (4.13). Its Sharpe ratio decreases when the value of ρ increases. An eigenportfolio with the zero sum of its *long* and *short* positions has no market exposure. It is observed from Figure 4.3a that odd indexed eigenportfolios have zero market exposure for $\rho < 0$. Hence, their Sharpe ratios are zero in that range of ρ as shown in Figure 4.3b.

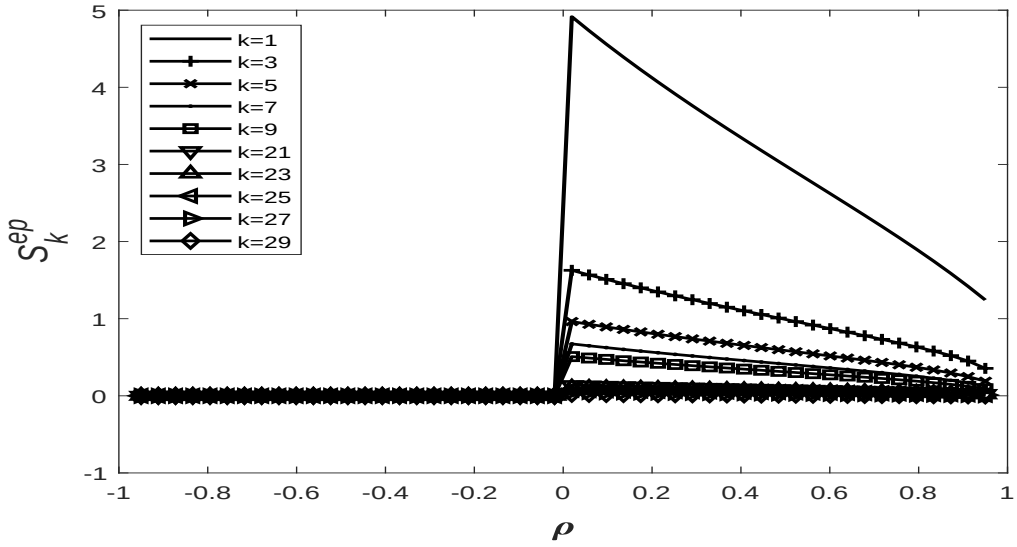
The market exposures M_k^{ep}, and Sharpe ratios S_k^{ep} for the first and the last four odd indexed eigenportfolios generated from the exponential correlation model, with $\mu_k = 1$, $\sigma_k = 1$ and $\rho = 0.9$, as a function of portfolio size N are displayed in Figures 4.4a and 4.4b, respectively. The Sharpe ratios of eigenportfolios with non-zero market exposures increase with portfolio size. On the other hand, the market exposure is less sensitive to the portfolio size in particular when N >100. In contrast, the even indexed eigenportfolios of the exponential correlation model have zero market exposure for $\rho > 0$. Therefore, their Sharpe ratios are always zero regardless of the portfolio size.

4.3.2 Performance of Eigenportfolios Based on Empirical Correlations of DJIA Stocks

PNL curves of the first eigenportfolios generated from the empirical correlation matrix and its exponential model based Toeplitz approximation for 28 stocks of the DJIA index using EOD data from July 1, 1999 to November 1, 2018 are displayed in Figure 4.5a. Their Sharpe ratios are 0.55 and 0.56, respectively, as tabulated in Table 4.1. Similarly, PNLs of the first five eigenportfolios for the same data set are shown in Figure 4.5b, and their average returns, volatilities and Sharpe ratios are also given

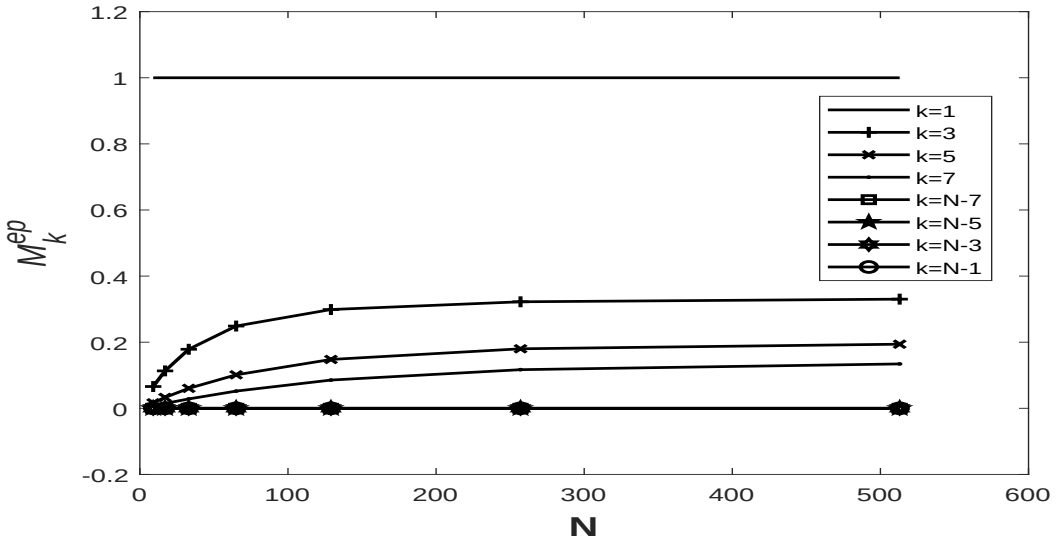


(a)

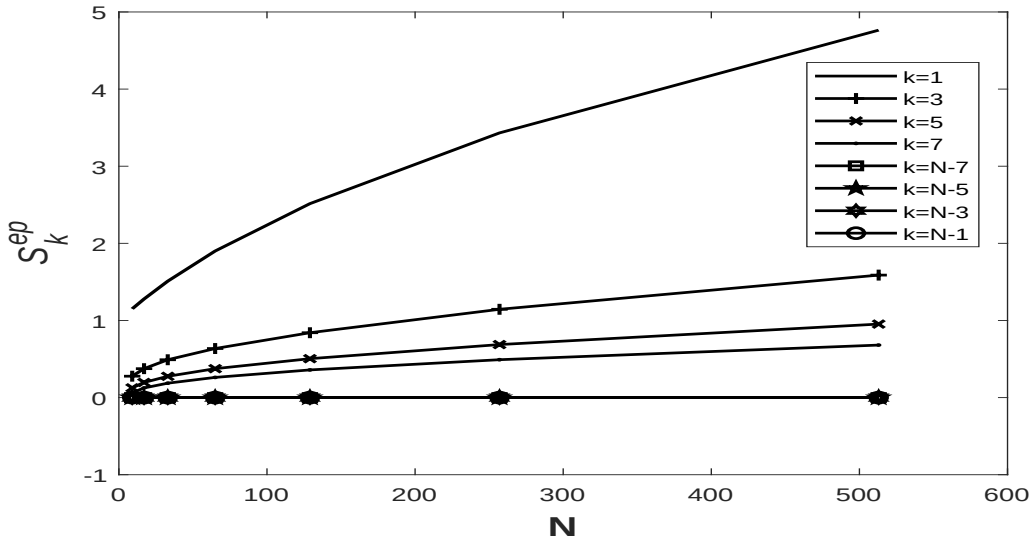


(b)

Figure 4.3 (a) Variations of market exposures M_k^{ep}, and (b) Sharpe ratios S_k^{ep} for the first and the last five odd indexed eigenportfolios (EPs) for exponential correlation model based eigenportfolios, $\mu_k = 1$, $\sigma_k = 1$ and $N = 30$, with respect to ρ.

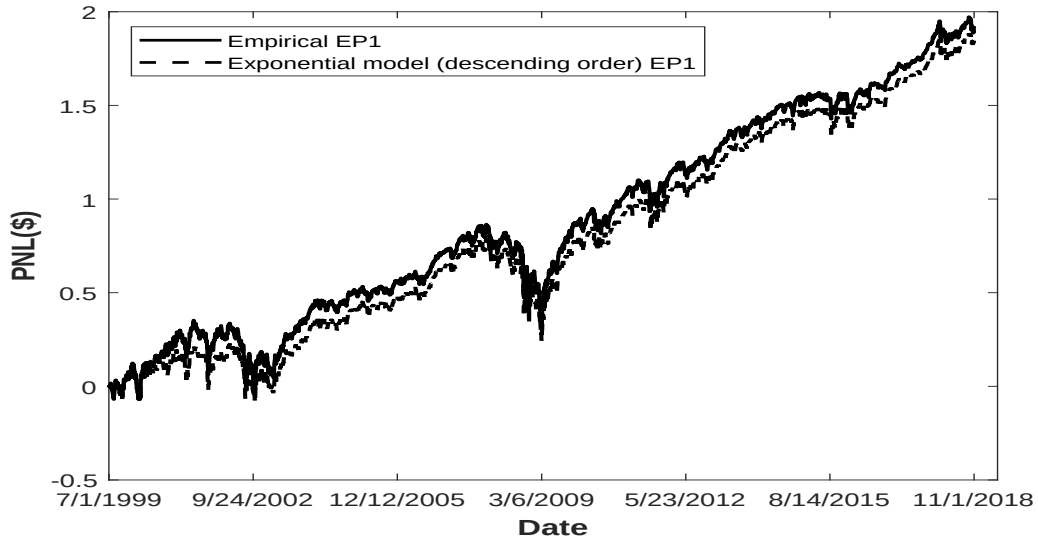


(a)

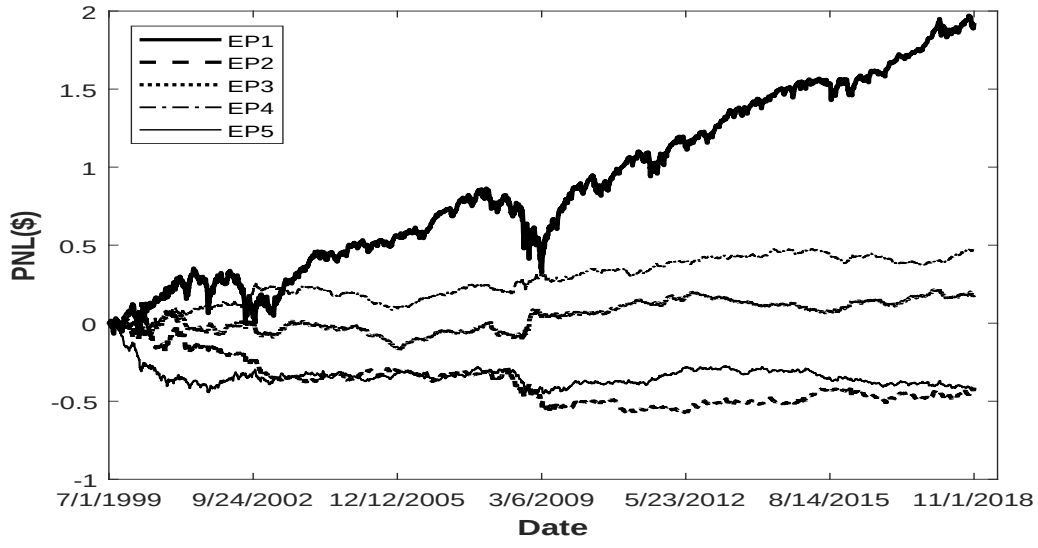


(b)

Figure 4.4 (a) Variations of market exposures M_k^{ep}, and (b) Sharpe ratios S_k^{ep} as a function of portfolio size N for the first and the last four odd indexed eigenportfolios (EPs) generated from the exponential model with $\mu_k = 1$, $\sigma_k = 1$ and $\rho = 0.9$.



(a)



(b)

Figure 4.5 PNL curves for EOD returns, from July 1, 1999 to November 1, 2018, of (a) the first eigenportfolios generated from the empirical and the exponential model based correlation matrices for the 28 stocks in DJIA index, $N = 28$ with $W = 40$, (b) the first five eigenportfolios generated from their empirical correlation matrix where daily rebalancing always maintains the total open positions normalized to \$1.

in Table 4.1 for comparison purposes. Furthermore, Sharpe ratios of the exponential model and empirical correlation based first eigenportfolio design examples for various subintervals of the same basket are tabulated in Table 4.2 to highlight the robustness of the method.

Table 4.1 Mean, Standard Deviation and Annualized Sharpe Ratios for EOD Returns, from July 1, 1999 to November 1, 2018, of the First Five Eigenportfolios Generated from the Empirical and Exponential Model Based Correlation Matrices for the 28 Stocks in DJIA Index, $N = 28$ with $W = 40$.

		k = 1	k = 2	k = 3	k = 4	k = 5
$\mu_{r_k^{ep}}$(bps)	Model	3.80	-0.34	0.66	-0.15	-0.25
	Data	3.95	-0.88	0.34	0.98	-0.86
$\sigma_{r_k^{ep}}$(bps)	Model	108.93	35.44	33.32	29.86	29.22
	Data	112.15	48.65	39.14	36.15	36.24
S_k^{ep} (annual)	Model	0.55	-0.15	0.32	-0.08	-0.13
	Data	0.56	-0.29	0.14	0.43	-0.38

Table 4.2 Mean, Standard Deviation and Annualized Sharpe Ratios for EOD Returns of Several Different Subintervals, of the First Eigenportfolios Generated from the Empirical and Exponential Model Based Correlation Matrices for the 28 Stocks in DJIA Index, $N = 28$ with $W = 40$.

		$\mu_{r_k^{ep}}$(bps)	$\sigma_{r_k^{ep}}$(bps)	S_k^{ep} (annual)
July 1, 1999 - Nov 1, 2018	Model	3.79	108.93	0.55
	Data	3.95	112.15	0.56
Nov 1, 2004 - Nov 1, 2018	Model	4.21	105.53	0.63
	Data	4.16	106.83	0.62
Nov 2, 2009 - Nov 1, 2018	Model	5.07	83.50	0.96
	Data	4.98	84.98	0.93
Nov 3, 2014 - Nov 1, 2018	Model	4.03	77.58	0.82
	Data	4.04	77.59	0.82

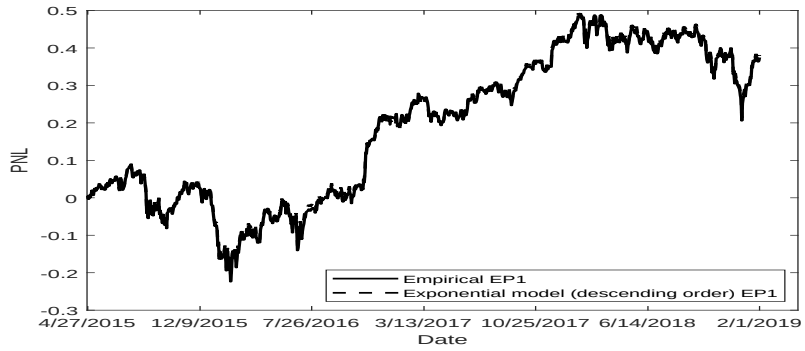
PNL curves for the first eigenportfolios of the three baskets comprised of the most allocated 15 stocks of the sector ETFs XLF (Finance)[3], XLI (Industry)[4] and XLV (Health Care)[5] for EOD returns and with $W = 60$ and the interval from Apr. 27, 2015 to Feb. 1, 2019 for the empirical and exponential correlation model based scenarios are displayed in Figure 4.6. It is shown in the figure that the PNLs and Sharpe ratios of the latter closely mimics the first similar to the observations made with DJIA index case presented above.

In implementation, the measurement windows may be tuned for the given basket in order to reduce approximation errors. Each eigenportfolio is re-balanced daily and no trading cost is considered in these PNLs. This analysis brings new insights to better understand eigenportfolios used in regression based trading algorithms like statistical arbitrage, and in other investment strategies [10, 7, 12, 4].

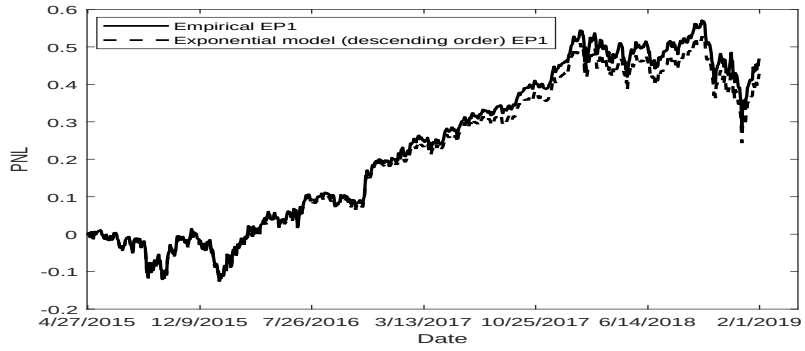
4.4 Performance Comparison of Minimum Variance, Market and Eigen Portfolios

Minimum Variance Portfolio PNL curves of MVP and MP defined in equations (3.16) and (3.23) generated from the empirical correlation matrix and its approximated exponential correlation model for EOD data of 28 stocks in the DJIA index from July 9, 1999 to November 1, 2018 are displayed in Figure 4.7a, and b, respectively. Each portfolio is rebalanced daily and no trading cost is considered in these PNLs. The PNLs of MVP, MP, the first eigenportfolio (EP1) and DIA for the same data set and time duration are shown in Figure 4.7c. Their average returns, volatilities and Sharpe ratios for the exponential model and the empirical correlation based design cases are tabulated in Table 4.3. for comparison purposes. Similarly, mean, standard deviation and annualized Sharpe ratios of MVP, MP, EP1s of the first

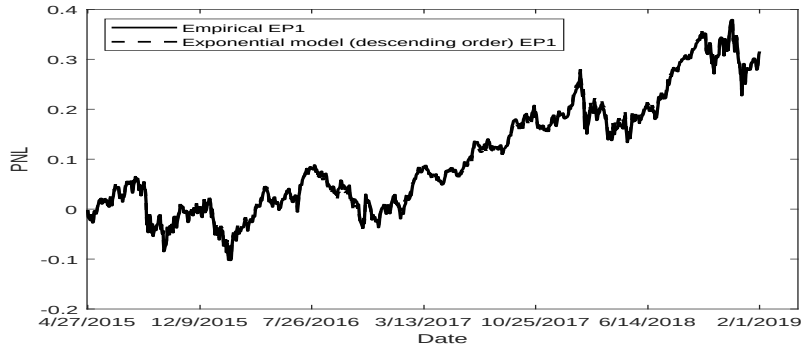
[3]AXP, BAC, BK, BLK, C, CB, CME, GS, JPM, MS, PNC, SCHW, SPGI, USB, WFC
[4]BA, CAT, CSX, DE, GD, GE, HON, LMT, MMM, NOC, NSC, RTN, UNP, UPS, UTX
[5]ABBV, ABT, AMGN, ANTM, BMY, CI, CVS, GILD, JNJ, LLY, MDT, MRK, PFE, TMO, UNH



(a)



(b)



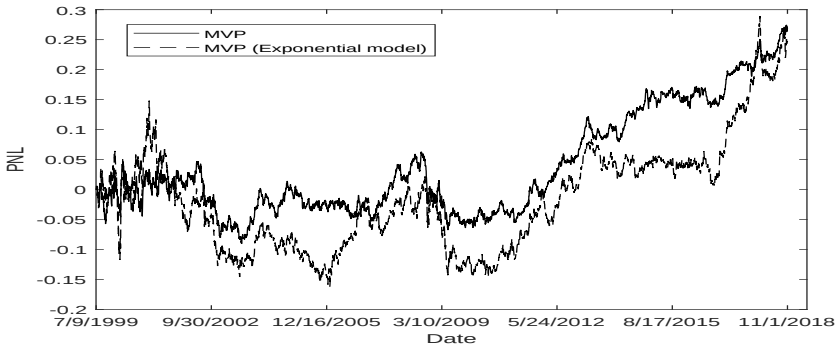
(c)

Figure 4.6 PNL curves from Apr. 27, 2015 to Feb. 1, 2019 for EOD returns of the first eigenportfolios for the 15 most allocated stocks of the three sector ETFs generated from the empirical and the exponential model based correlation matrices, $N = 15$ with $W = 60$, (a) XLF (Finance), (b) XLI (Industrial) and (c) XLV (Health Care) where daily rebalancing has always maintained the total portfolio positions of \$1 (normalized to the sum of absolute values of investment allocation coefficients).

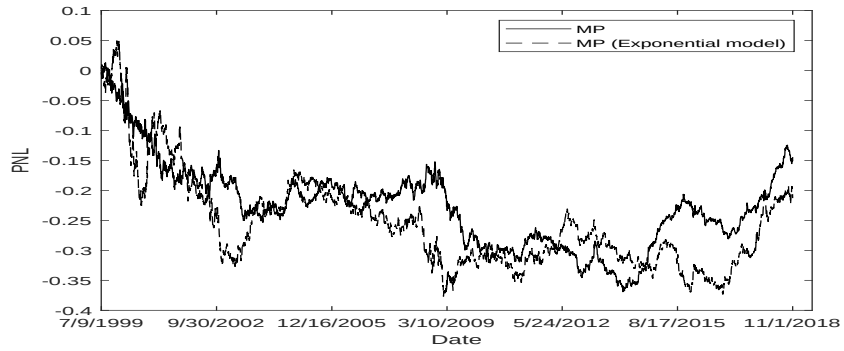
15 most allocated stocks in the sector ETFs XLF, XLI and XLV for the time interval from April 27, 2015 to Feb. 1, 2019 are tabulated in Table 4.4. It is observed from the performance simulations that the exponential model and empirical correlation matrix based eigenportfolios perform closely for all the scenarios presented in the chapter. It is also observed that the first eigenportfolios of the baskets that are considered in this study almost always outperform ETFs, MVP, and MP.

Table 4.3 Mean, Standard Deviation and Annualized Sharpe Ratios of MVP, MP, EP1 and DIA Calculated from the PNLs of Figure 4.7

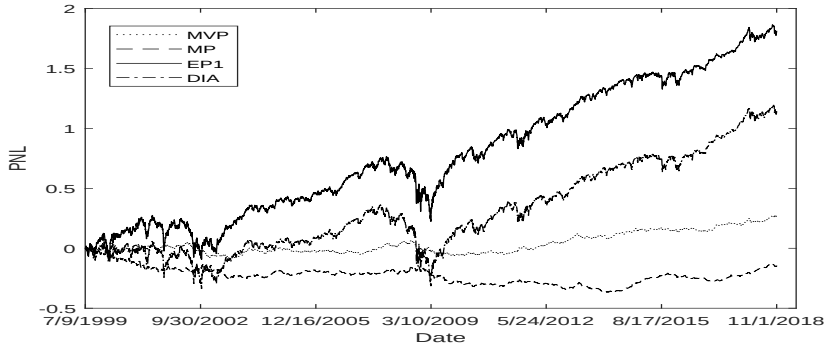
		MVP	MP	EP1	DIA
μ(bps)	Model	0.51	-0.44	3.85	NA
	Data	0.55	-0.30	3.73	2.35
σ(bps)	Model	46.22	40.46	109.02	NA
	Data	32.48	30.30	112.42	114.48
S (annual)	Model	0.18	-0.17	0.56	NA
	Data	0.27	-0.15	0.53	0.33



(a)



(b)



(c)

Figure 4.7 Profit and Loss (PNL) curves of (a) MVP, and (b) MP generated from the empirical and the exponential model based correlation matrices of the 28 stocks in DJIA index, for EOD returns from July 9, 1999 to November 1, 2018 with $W = 45$. (c) PNL curves of MVP, MP, EP1 derived from the empirical correlation matrix, and DIA for the same time interval and $W = 45$. The portfolios were rebalanced daily, and the total positions for each day are maintained as \$1.

Table 4.4 Mean, Standard Deviation and Annualized Sharpe Ratios of the MVP, MP, EP1 of the First 15 Most Allocated Stocks of the Sector ETFs (a) XLF, (b) XLI and (c) XLV.

(a)

		MVP	MP	EP1	XLF
μ(bps)	Model	0.15	-0.67	4.03	NA
	Data	1.30	0.77	3.96	3.28
σ(bps)	Model	42.50	38.17	119.37	NA
	Data	42.97	25.78	122.54	97.44
S (annual)	Model	0.06	-0.28	0.54	NA
	Data	0.48	0.47	0.51	0.54

(b)

		MVP	MP	EP1	XLI
μ(bps)	Model	2.01	2.62	4.52	NA
	Data	2.00	2.27	4.95	3.14
σ(bps)	Model	42.14	36.45	98.61	NA
	Data	40.12	29.81	99.73	99.58
S (annual)	Model	0.76	1.14	0.73	NA
	Data	0.79	1.21	0.79	0.50

(c)

		MVP	MP	EP1	XLV
μ(bps)	Model	2.00	-0.29	3.34	NA
	Data	1.21	-1.17	3.33	2.52
σ(bps)	Model	47.45	37.87	92.73	NA
	Data	44.19	34.78	93.62	95.88
S (annual)	Model	0.67	-0.12	0.57	NA
	Data	0.43	-0.53	0.56	0.42

4.5 On Equivalence of Minimum Variance Portfolio and First Eigenportfolio

Let's assume that the correlation matrix $\mathbf{R}$ has the following structure

$$\mathbf{R} = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} \tag{4.15}$$

and, Σ has equal diagonal elements as

$$\Sigma = \begin{bmatrix} \sigma & 0 & \cdots & 0 \\ 0 & \sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma \end{bmatrix} \tag{4.16}$$

Then, it can be shown that the resulting MVP and one of the eigenportfolios of $\mathbf{R}$ are identical.

With such condition, the MVP is calculated as

$$\mathbf{q}_{min} = \frac{\mathbf{C}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1}} = \frac{\Sigma^{-1}\mathbf{R}^{-1}\left(\Sigma^T\right)^{-1}\mathbf{1}}{\mathbf{1}^T\Sigma^{-1}\mathbf{R}^{-1}\left(\Sigma^T\right)^{-1}\mathbf{1}} = \frac{\sigma^{-2}\mathbf{R}^{-1}\mathbf{1}}{\mathbf{1}^T\sigma^{-2}\mathbf{R}^{-1}\mathbf{1}} = \frac{\mathbf{R}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}} \tag{4.17}$$

Then, $\mathbf{R}^{-1} = \frac{1}{det(\mathbf{P})} adj\,(\mathbf{R})$ where $adj\,(\mathbf{R}) = \begin{Bmatrix} c_{11} & c_{12} & \cdots & c_{1N} \\ c_{21} & c_{22} & \cdots & c_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{N1} & c_{N2} & \cdots & c_{NN} \end{Bmatrix}$. If $\mathbf{R}$ has the special structure where $c_{ii} = d$, and $c_{ij} = c$ for $i \neq j$, such that

$$\mathbf{R}^{-1} = \frac{1}{det\,(\mathbf{R})} \begin{bmatrix} d & c & c & \cdots & c \\ c & d & c & \cdots & c \\ c & c & d & \cdots & c \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c & c & c & \cdots & d \end{bmatrix} \tag{4.18}$$

then, it is easy to derive that MVP has equal weight $q_i = \frac{1}{N}$ for each assets i.

The first eigenvector $\boldsymbol{\phi_1}$ of $\mathbf{R}$ is derived by solving the optimization problem

$$\begin{aligned} \max \quad & \boldsymbol{\phi}_k^T \mathbf{R} \boldsymbol{\phi}_k \\ \mathbf{s.t.} \quad & \boldsymbol{\phi}_k^T \boldsymbol{\phi}_n = \delta_{k-n}; k, n = 1, 2, ..., N \end{aligned} \tag{4.19}$$

Then, we can express the Lagrangian multiplier as

$$L\,(\boldsymbol{\phi}_1, \lambda) = \boldsymbol{\phi}_1^T \mathbf{R} \boldsymbol{\phi}_1 + \lambda \left(1 - \boldsymbol{\phi}_1^T \boldsymbol{\phi}_1\right) \tag{4.20}$$

By taking the partial derivative we will have

$$\mathbf{R}\boldsymbol{\phi}_1 - \lambda \boldsymbol{\phi}_1 = 0 \tag{4.21}$$

$$\mathbf{R}\boldsymbol{\phi}_1 = \lambda \boldsymbol{\phi}_1 \tag{4.22}$$

Hence, the solution of $\boldsymbol{\phi}_1 = [\phi_{11}, \phi_{12}, ..., \phi_{1N}]^{\mathrm{T}}$ is the first eigenvector of the correlation matrix $\mathbf{R}$ and λ is the corresponding eigenvalue.

$$\begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{11} \\ \phi_{12} \\ \vdots \\ \phi_{1N} \end{bmatrix} = \begin{bmatrix} \lambda\phi_{11} \\ \lambda\phi_{12} \\ \vdots \\ \lambda\phi_{1N} \end{bmatrix} \tag{4.23}$$

$$\begin{aligned} \phi_{11} + \rho\left(\phi_{12} + \cdots + \phi_{1N}\right) &= \lambda\phi_{11} \\ \phi_{12} + \rho\left(\phi_{11} + \cdots + \phi_{1N}\right) &= \lambda\phi_{12} \\ &\vdots \\ \phi_{1N} + \rho\left(\phi_{11} + \cdots + \phi_{1N-1}\right) &= \lambda\phi_{1N} \end{aligned} \tag{4.24}$$

form (4.24) we can have $(1-\rho)(\phi_{11} - \phi_{12}) = \lambda(\phi_{11} - \phi_{12})$, assume that $\phi_{11} \neq \phi_{12} \neq ... \neq \phi_{1N}$, then

$$\lambda = 1 - \rho. \tag{4.25}$$

If $\phi_{11} = \phi_{12} = \cdots = \phi_{1N} = \sqrt{\frac{1}{N}}$, then

$$\lambda = 1 + \rho(N-1) \tag{4.26}$$

while since $1 + \rho(N-1) \geq 1 - \rho$ always true when $\rho \geq 0$, so $\lambda_1 = 1 + \rho(N-1)$ is the largest eigenvalue, which makes $\phi_{11} = \phi_{12} = \cdots = \phi_{1N} = \sqrt{\frac{1}{N}}$ the first eigenvector, the rest eigenvalues are all identical $\lambda_{rest} = 1 - \rho$. When $\rho < 0$, then $1 + \rho(N-1) < 1 - \rho$, so that $\lambda_1 = 1 + \rho(N-1)$ is the smallest eigenvalue, which makes $\phi_{11} = \phi_{12} = \cdots = \phi_{1N} = \sqrt{\frac{1}{N}}$ the last eigenvector. Therefore, after normalization the first (last) eigenportfolio is identical with the MVP.

4.6 Chapter Summary

In this chapter, the empirical correlations of asset returns in a basket with exponential function is modeled. These exponential correlations populate a Toeplitz matrix with closed-form expressions for its eigenvalues and eigenvectors. They are utilized to design model based eigenportfolios. The eigenportfolios designed by using empirical correlation matrix generated from market data are also investigated. The performances of the two eigenportfolio sets are simulated by using real mark data. This chapter also demonstrates that the exponential approximation to empirical correlations provide a good model to design eigenportfolios and to evaluate their performance. The performances of the minimum variance, the market and eigenportfolios for EOD returns of U.S. equities along with the index and sector ETFs are compared. It is concluded that the first eigenportfolio (EP1) provides almost always the best performance among all portfolios considered in this dissertation.

CHAPTER 5

SPARSITY OF EIGENPORTFOLIOS

In order to reduce volatility, large size portfolios with built-in diversity are commonly used in practice. On the other hand, the portfolio maintenance (re-balancing) becomes more costly when portfolio size is large, e.g., a few hundred asset portfolio. Therefore, calculated small adjustments of some asset positions are judiciously ignored in the implementation during the periodic re-balancing process by employing a method to sparse large size portfolios. A sparsing technique for Markowitz (mean-variance) portfolio [26] was proposed in [9] a penalty (regularization term) which is proportional by employing L1-norm based lasso regression [36].

The work in this chapter is a continuation of the subspace sparsing framework proposed in [43]. It is based on the rate-distortion theory and employs zero-zone (mid-tread) pdf-optimized (Lloyd-Max) quantizer created for the histogram of an eigenvector or eigenmatrix and the desired level of sparsity in the subspace [28, 23, 2]. We focus on sparsing the eigenportfolios of stocks in S&P500 index by using this method and their resulting performance.

5.1 Sparsity in Subspace Methods

The energy compaction that is achieved through the unevenness of transform coefficient variances, and their pairwise correlations are the performance metrics derived from the rate-distortion theory to evaluate orthogonal sets (orthogonal subspace methods). The energy compaction measure emphasizes the spectral (frequency domain) features of a subspace representation and the foundation of transform coding, that is the industry standard for image and video compression standards [19, 2]. In contrast, sparse representation aims to replace insignificant components of the basis vectors, cardinality reduction, that define an orthogonal

subspace. Hence, it highlights the signal domain (time domain) characteristics of subspace representation. We investigate both the time and the frequency domain sparsities of subspace methods where both the explained variance and sparsity are quantified.

Most of the sparse representation techniques reported in the literature are based on various subspace optimization methods where sparsity is imposed in the design. More recently, the quantization of basis vector components of a subspace was experimented as a more efficient alternative to the existing sparsing approaches. The following subsections focus on the quantization of orthogonal eigen subspace vectors and matrices that are used as the capital allocation coefficients of the resulting eigenportfolios [2, 43].

5.1.1 Quantization of Subspace

Let's focus on the quantization of forward transform matrix (vectors) in this section. The quantized (sparsed) version rather than the original transform matrix is employed for signal representation. The motivation for such vector sparsing (cardinality reduction) is to replace insignificant vector components with zero. Thus, one may reduce computational and implementation cost of orthogonal set based subspace applications spanning from image compression to eigenportfolios.

We quantize orthogonal set (transform matrix) by using a quantizer $Q_{\Phi}\{\cdot\}$ [43]

$$\hat{\Phi} = Q_{\Phi}\{\Phi\}, \tag{5.1}$$

such that the error matrix of the quantized set is written as

$$\tilde{\Phi} = \Phi - \hat{\Phi}, \tag{5.2}$$

It is noted that this quantization compromises the orthogonality property of the set. The levels of non-orthogonality and the sparsity are coupled in this problem. The

transform coefficients for the quantized set are calculated as

$$\hat{\boldsymbol{\theta}}_{FT} = \hat{\Phi}\mathbf{x}, \tag{5.3}$$

Therefore, the quantization error embedded in transform coefficients due to the use of $\hat{\Phi}$ rather than the (original) matrix Φ is shown as

$$\tilde{\boldsymbol{\theta}}_{FT} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{FT} = \Phi\mathbf{x} - \hat{\Phi}\mathbf{x} = \left(\Phi - \hat{\Phi}\right)\mathbf{x} = \tilde{\Phi}\mathbf{x}, \tag{5.4}$$

Then, the reconstructed signal by using the original inverse transform matrix $\Phi^{-1} = \Phi^{\mathrm{T}}$ is obtained as

$$\hat{\mathbf{x}}_{FT} = \Phi^{T}\hat{\boldsymbol{\theta}}_{FT}, \tag{5.5}$$

The reconstruction error is given in the expression

$$\tilde{\mathbf{x}}_{FT} = \mathbf{x} - \hat{\mathbf{x}}_{FT} = \Phi^{\mathrm{T}}\boldsymbol{\theta} - \Phi^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{FT} = \Phi^{\mathrm{T}}\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{FT}\right) = \Phi^{\mathrm{T}}\tilde{\boldsymbol{\theta}}_{FT}. \tag{5.6}$$

The mse of coefficients is equal to the mse of reconstructed signal due to the orthogonality of the original set as

$$E\left\{\tilde{\mathbf{x}}_{FT}^{\mathrm{T}}\tilde{\mathbf{x}}_{FT}\right\} = E\left\{\left(\Phi^{\mathrm{T}}\tilde{\boldsymbol{\theta}}_{FT}\right)^{\mathrm{T}}\left(\Phi^{\mathrm{T}}\tilde{\boldsymbol{\theta}}_{FT}\right)\right\} = E\left\{\tilde{\boldsymbol{\theta}}_{FT}^{\mathrm{T}}\underbrace{\Phi\Phi^{T}}_{I}\tilde{\boldsymbol{\theta}}_{FT}\right\} = E\left\{\tilde{\boldsymbol{\theta}}_{FT}^{\mathrm{T}}\tilde{\boldsymbol{\theta}}_{FT}\right\}, \tag{5.7}$$

In summary,

$$E\left\{\tilde{\mathbf{x}}_{FT}^{\mathrm{T}}\tilde{\mathbf{x}}_{FT}\right\} = E\left\{\tilde{\boldsymbol{\theta}}_{FT}^{\mathrm{T}}\tilde{\boldsymbol{\theta}}_{FT}\right\} = E\left\{\mathbf{x}^{T}\tilde{\Phi}^{\mathrm{T}}\tilde{\Phi}\mathbf{x}\right\}. \tag{5.8}$$

One can easily exchange the roles of the forward and the inverse transform matrices for the applications like eigenportfolio design where the sparsity of the representation (inverse) set is desired.

5.2 The Eigensubspace Distribution for Exponential Correlation Matrix and Rate Distortion Comparisons

In previous chapter, the eigenportfolios are designed based on the empirical correlation matrix approximated by a exponential correlation model. It is showed that, this model provide a similar performance with the one generated by market data. Consequently, it is naturally to study the coefficients distribution of such a eigensubspace.

5.2.1 Estimation of Probability Density Function of KLT Kernel for Exponential Correlation Matrix

Suppose a sinusoidal function is defined as

$$v = \mathrm{A}\sin(y) \tag{5.9}$$

if y is a random variable with uniform distribution $Y \sim U(-\pi, \pi)$, then the probability density function of v can be easily derived as

$$f_V(v) = \frac{1}{\pi\sqrt{(\mathrm{A}+v)(\mathrm{A}-v)}} \tag{5.10}$$

which is an arcsine distribution.

An efficient root finding method for explicit solutions of transcendental equations was proposed in [38], from which the roots $\{0 \leq \omega_k \leq \pi\}$ can be found, by using equations (2.16), (2.17) and (2.18), the KLT kernel $\mathbf{A}_{KLT}$ can be calculated, the distribution of the components for entire $\mathbf{A}_{KLT}$ matrix are displayed in Figure 5.1, this distribution is similar to an arcsine distribution with some discrepancies. In fact, the amount of positive components in $\mathbf{A}_{KLT}$ is larger than the amount of negative components. It would be easily noticed by plotting the distribution for each column of $\mathbf{A}_{KLT}$ for large dimensionality as Figure 5.4 shows, the reason behind this difference are revealed as follows.

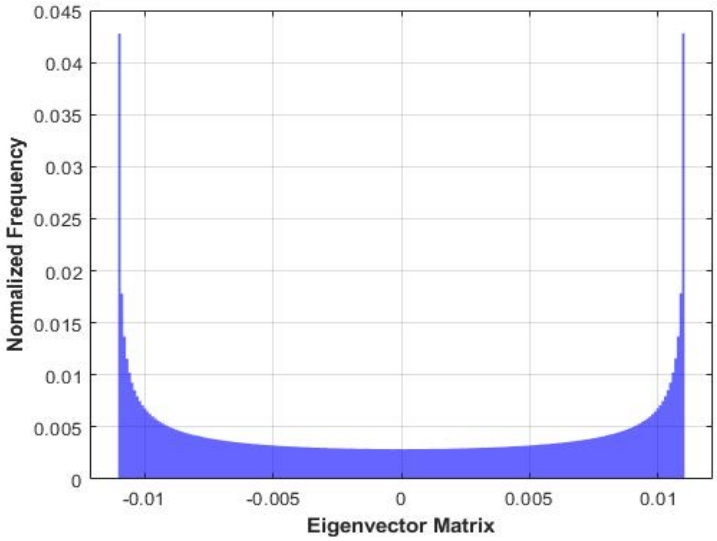


Figure 5.1 Distribution of eigenvector matrix for exponential correlation matrix with $\rho = 0.9$ and $N = 16384$.

If the argument of the sinusoidal function in (2.18) is defined as

$$y_k = \omega_k \left(n - \frac{N-1}{2} \right) + \frac{(k+1)\pi}{2} \tag{5.11}$$

then y_k can be easily calculated.

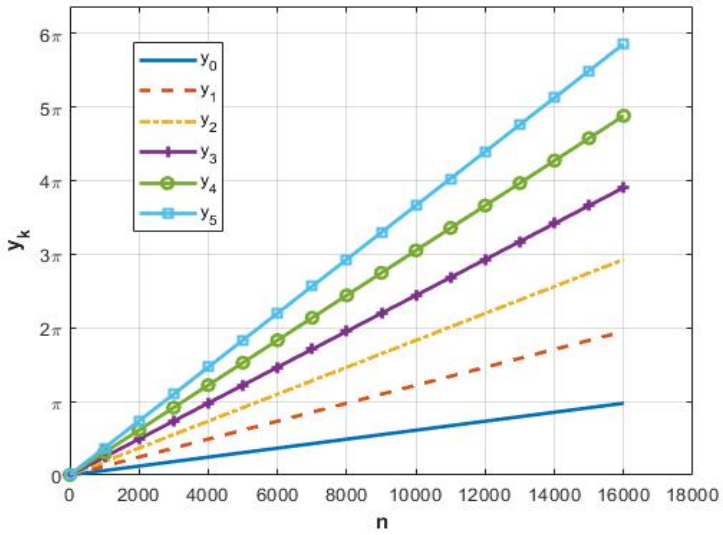


Figure 5.2 Plots of y_k for exponential correlation model with $\rho = 0.9$ and $N = 16384$.

From Figure 5.2, the range of each y_k for large N, can be approximated as

$$y_k \in [0, (k+1)\pi] \quad k = 0, 1, 2, ...N-1, \tag{5.12}$$

each column of $\mathbf{A}_{KLT}$ is an eigenvector. Figure 5.3 shows the plot of PC1 to PC6 of the eigenvector matrix. Notice that for even index PC, k in equation (5.12) is an odd number, the components plots are mirror symmetric of $n = \frac{N}{2}$, which mean the components with positive and negative value have the same amount; for odd index PC and k is an even number, the components plots are centrosymmetric of $n = \frac{N}{2}$. Therefore, the components with positive and negative value have a different amount.

Now, in order to derive the probability density function (pdf) of the distribution for columns in $\mathbf{A}_{KLT}$, assume that $y_k^{'}$ is a random variable and it follows an uniform distribution

$$\text{distribution of } y_k^{'} \text{ is } Y_k^{'} \sim U\left(0, (k+1)\pi\right) \tag{5.13}$$

The pdf of $v_k = c_k \sin(y_k^{'})$ where c_k as a constant, can be derived as a function of k

$$f_{V_k}(v_k) = \begin{cases} \frac{1}{\pi\sqrt{(c_k+v_k)(c_k-v_k)}} & -c_k \le v_k \le c_k \qquad k \text{ is odd} \\ \begin{cases} \frac{k+2}{(k+1)\pi\sqrt{(c_k+v_k)(c_k-v_k)}} & 0 \le v_k \le c_k \\ \frac{k}{(k+1)\pi\sqrt{(c_k+v_k)(c_k-v_k)}} & -c_k \le v_k < 0 \end{cases} & k \text{ is even} \end{cases} \qquad k = 0, 1, 2, ...N-1 \tag{5.14}$$

Proof: The derivation of (5.14) can be started with some simple problems, consider a random variable $y_0^{'}$ has a distribution of $Y_0^{'} \sim U(0, \pi)$, and a sinusoidal function is defined as

$$v_0 = c_0 \sin(y_0^{'}) \tag{5.15}$$

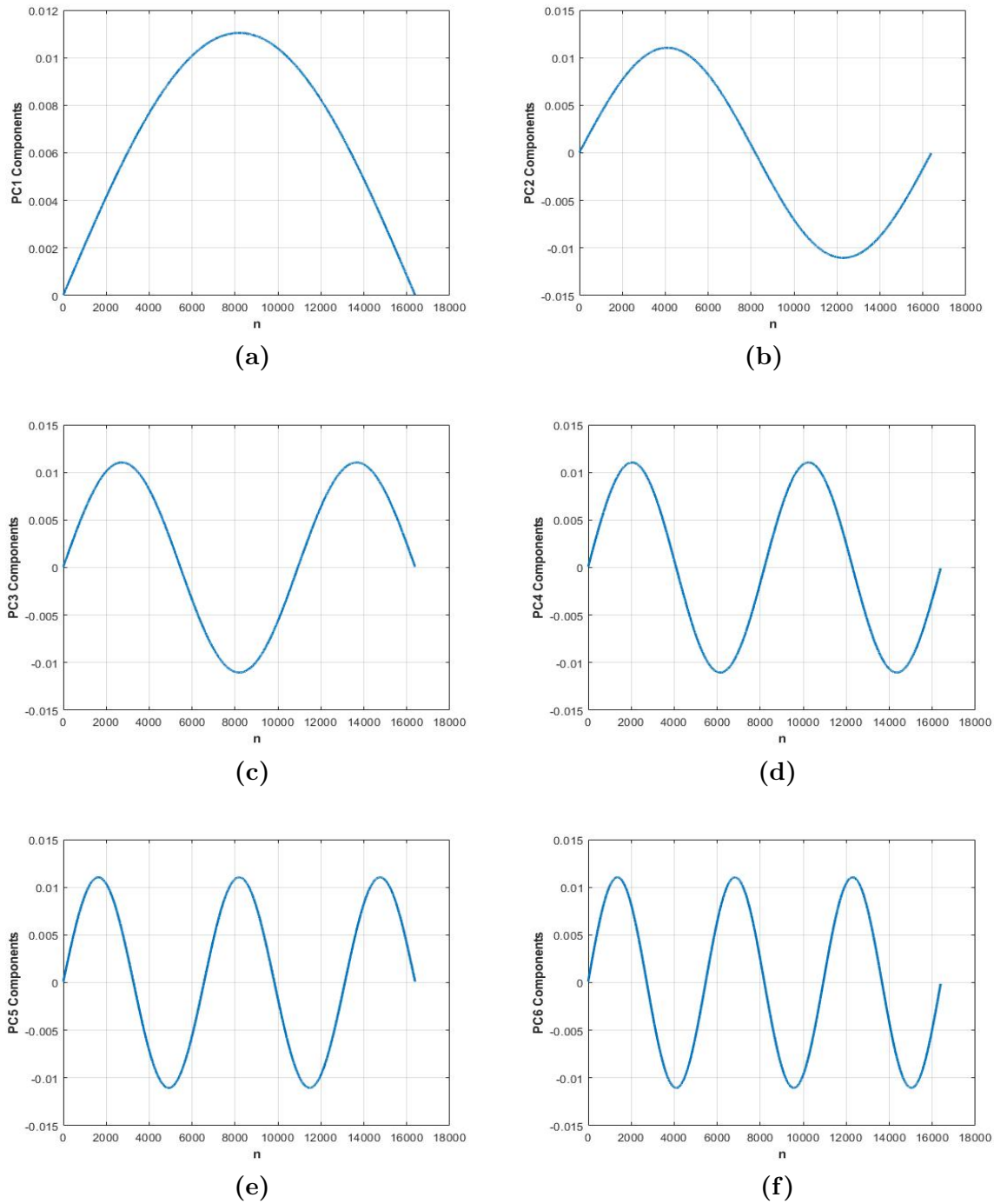


Figure 5.3 Plots of PC1 (a), PC2 (b), PC3 (c), PC4 (d), PC5 (e), PC6 (f) for exponential correlation matrix, with $\rho = 0.9$ and $N = 16384$.

the probability density function (pdf) of $y_0^{'}$ is

$$f_{Y_0^{'}}(y_0^{'}) = \begin{cases} \frac{1}{\pi} & 0 \leq y_0^{'} \leq \pi \\ 0 & else \end{cases} \tag{5.16}$$

then the pdf of v_0 can be calculated as follow, for a function $v = g(y^{'})$, $y^{'}$ has the probability density function (pdf) $f_{Y^{'}}(y^{'})$, then the pdf of v can be calculated as [22]

$$f_V(v) = \sum_n f_{Y^{'}}(y^{'}) \left| \frac{dy^{'}}{dv} \right\|_{y^{'}=y_n^{'}} \tag{5.17}$$

note that n represent there are n solutions for equation $v = g(y^{'})$, so for equation (5.15), $n = 2$ with $0 \leq y_0^{'} \leq \pi$, $0 \leq v_0 \leq c_0$, then the pdf of v_0 can be calculated as

$$f_{V_0}(v_0) = \frac{2}{\pi} \left| \frac{dy_0^{'}}{dv_0} \right| = \frac{2}{\pi} \left| \frac{d(arcsin(\frac{v_0}{c_0}))}{dv_0} \right| = \frac{2}{\pi\sqrt{((c_0 + v_0)(c_0 - v_0))}}, \tag{5.18}$$

with $0 \leq v_0 \leq c_0$. However, if a random variable $y_1^{'}$ has a distribution of $Y_1^{'} \sim U(0, 2\pi)$, and a sinusoidal function is defined as

$$v_1 = c_1 \sin(y_1^{'}) \tag{5.19}$$

the probability density function (pdf) of $y_1^{'}$ is

$$f_{Y_1^{'}}(y_1^{'}) = \begin{cases} \frac{1}{2\pi} & 0 \leq y_1^{'} \leq 2\pi \\ 0 & else \end{cases} \tag{5.20}$$

for equation (5.19), $n = 2$ with $0 \leq y_1^{'} \leq 2\pi$, $-c_1 \leq v_1 \leq c_1$, then the pdf of v_1 can be calculated as

$$f_{V_1}(v_1) = \frac{2}{2\pi} \left| \frac{dy_1^{'}}{dv_1} \right| = \frac{1}{\pi} \left| \frac{d(arcsin(\frac{v_1}{c_1}))}{dv_1} \right| = \frac{1}{\pi\sqrt{((c_1 + v_1)(c_1 - v_1))}}, \tag{5.21}$$

with $-c_1 \leq v_1 \leq c_1$. Now the problem can be extended, suppose a random variable $y_k^{'}$, k is an even number $k = 0, 2, 4, ...$, has a distribution of $Y_k^{'} \sim U(0, (k+1)\pi)$, and

a sinusoidal function is defined as

$$v_k = c_k \sin(y_k^{'}) \tag{5.22}$$

the probability density function (pdf) of $y_k^{'}$ is

$$f_{Y_k^{'}}(y_k^{'}) = \begin{cases} \frac{1}{(k+1)\pi} & 0 \le y_k^{'} \le (k+1)\pi \\ 0 & else \end{cases} \tag{5.23}$$

there are $k+2$ solutions of $v_k = c_k \sin(y_k^{'})$ for $0 \le v_k \le c_k$ and k solutions for $-c_k \le v_k < 0$, so the pdf of v_k can be derived as

$$f_{V_k}(v_k) = \begin{cases} \frac{k+2}{(k+1)\pi} \left| \frac{dy_k^{'}}{dv_k} \right| = \frac{k+2}{(k+1)\pi} \left| \frac{d(arcsin(\frac{v_k}{c_k}))}{dv_k} \right| = \frac{k+2}{(k+1)\pi\sqrt{(c_k+v_k)(c_k-v_k)}} & 0 \le v_k \le c_k \\ \frac{k}{(k+1)\pi} \left| \frac{dy_k^{'}}{dv_k} \right| = \frac{k}{(k+1)\pi} \left| \frac{d(arcsin(\frac{v_k}{c_k}))}{dv_k} \right| = \frac{k}{(k+1)\pi\sqrt{(c_k+v_k)(c_k-v_k)}} & -c_k \le v_k < 0 \end{cases} \quad k \text{ is even.} \tag{5.24}$$

Another situation, k is an odd number $k = 1, 3, 5, ...$, has a distribution of $Y_k^{'} \sim U(0, (k+1)\pi)$, and a sinusoidal function is defined as

$$v_k = c_k \sin(y_k^{'}) \tag{5.25}$$

the probability density function (pdf) of $y_k^{'}$ is

$$f_{Y_k^{'}}(y_k^{'}) = \begin{cases} \frac{1}{(k+1)\pi} & 0 \le y_k^{'} \le (k+1)\pi \\ 0 & else \end{cases} \tag{5.26}$$

there are $k+1$ solutions of $v_k = c_k \sin(y_k)$ for $0 \le v_k \le c_k$ and $k+1$ solutions for $-c_k \le v_k < 0$, so the pdf of v_k can be derived as

$$f_{V_k}(v_k) = \begin{cases} \frac{k+1}{(k+1)\pi} \left| \frac{dy_k^{'}}{dv_k} \right| = \frac{1}{\pi} \left| \frac{d(arcsin(\frac{v_k}{c_k}))}{dv_k} \right| = \frac{1}{\pi\sqrt{(c_k+v_k)(c_k-v_k)}} & 0 \le v_k \le c_k \\ \frac{k+1}{(k+1)\pi} \left| \frac{dy_k^{'}}{dv_k} \right| = \frac{1}{\pi} \left| \frac{d(arcsin(\frac{v_k}{c_k}))}{dv_k} \right| = \frac{1}{\pi\sqrt{(c_k+v_k)(c_k-v_k)}} & -c_k \le v_k < 0 \end{cases} \quad k \text{ is odd,} \tag{5.27}$$

so that

$$f_{V_k}(v_k) = \frac{1}{\pi\sqrt{(c_k + v_k)(c_k - v_k)}} \quad k \text{ is odd}, \tag{5.28}$$

combine equations (5.24) and (5.28)

$$f_{V_k}(v_k) = \begin{cases} \frac{1}{\pi\sqrt{(c_k+v_k)(c_k-v_k)}} & -c_k \le v_k \le c_k \qquad k \text{ is odd} \\ \begin{cases} \frac{k+2}{(k+1)\pi\sqrt{(c_k+v_k)(c_k-v_k)}} & 0 \le v_k \le c_k \\ \frac{k}{(k+1)\pi\sqrt{(c_k+v_k)(c_k-v_k)}} & -c_k \le v_k < 0 \end{cases} & k = 0, 1, 2, ...N-1 \\ & k \text{ is even} \end{cases}$$

which is the same with equation (5.14).

Figure 5.4 displays the normalized histograms of the first 6 eigenvectors components (PC1 to PC6 loading coefficients) for exponential Toeplitz matrix. The value of N is selected large enough to generate proper histograms. The intervals τ_k of the adjacent bins in the histogram need to be chosen appropriately such that the number of bins of each histogram can be set as $\frac{\max(v_k)\text{-}\min(v_k)}{\tau_k}$. The dashed lines in each normalized histogram show the probability which is calculated by integrating $f_{V_k}(v_k)$ in equation (5.14) for each bin interval. It shows that the histogram of each column of the $\mathbf{A}_{KLT}$ can be illustrated by equation (5.14) explicitly, and the the components with positive and negative value have a discrepancy when k is an even number. This discrepancy can derive by taking the difference of $f_{V_k}(v_k)$ for even number k as

$$\begin{aligned} \Delta_k(v_k) &= \frac{k+2}{(k+1)\pi\sqrt{(c_k + v_k)(c_k - v_k)}} - \frac{k}{(k+1)\pi\sqrt{(c_k + v_k)(c_k - v_k)}} \\ &= \frac{2}{(k+1)\pi\sqrt{(c_k + v_k)(c_k - v_k)}} \quad 0 \le v_k \le c_k \textit{ and } k \text{ is even} \end{aligned} \tag{5.29}$$

5.3 Rate-distortion of Lloyd-Max pdf-optimized quantizer

By applying the pdf-optimized quantizer on the arc-sine and others such as Gaussian, Gamma, Laplacian and Uniform distribution, the rate-distortion curve can be derived

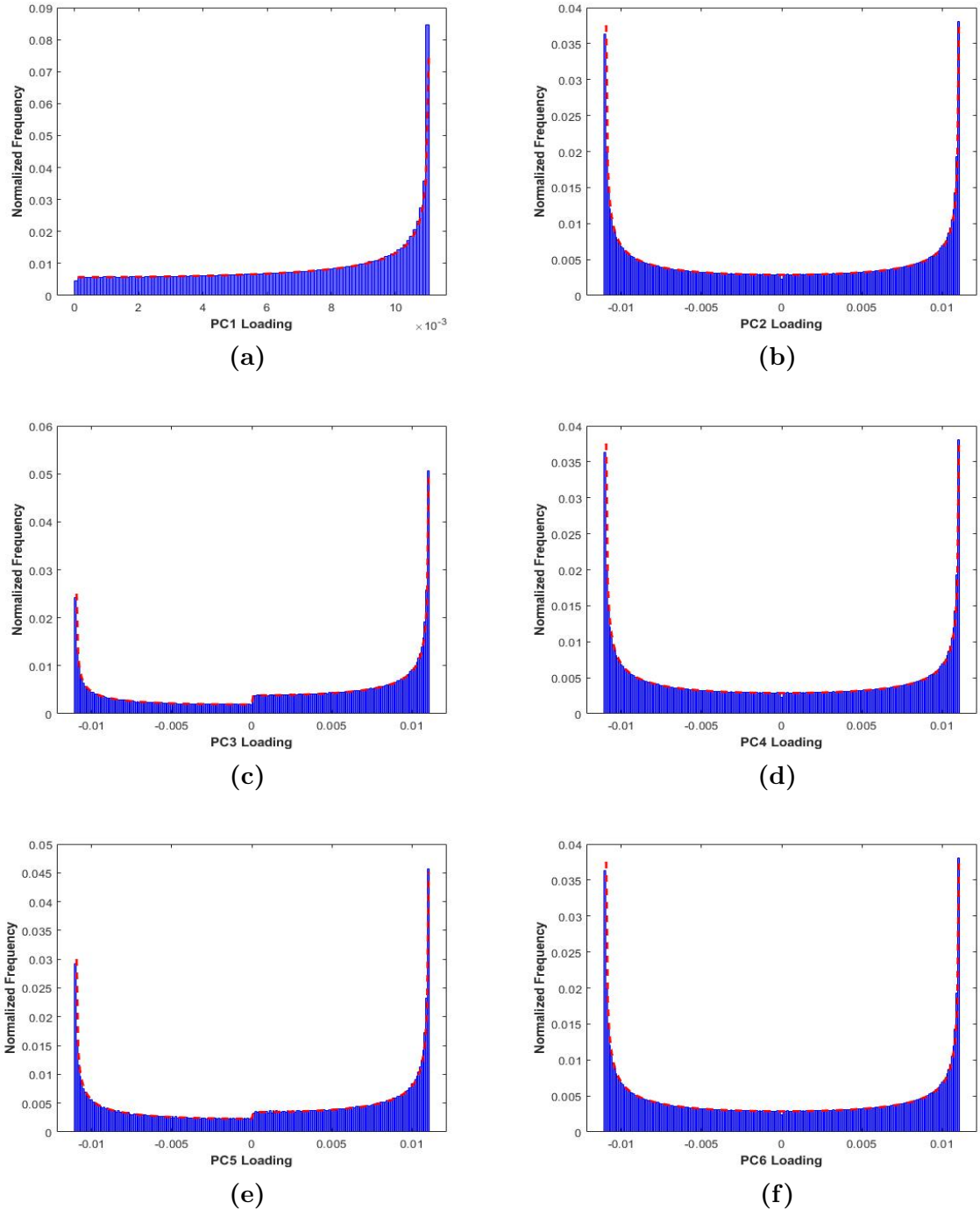


Figure 5.4 Normalized histograms of principle components for exponential correlation matrix with $\rho = 0.9$ and $N = 16384$. The dashed lines in each histogram show the probability that is calculated by integrating (5.14) for each bin interval.; $k = 0$ for (a), $k = 1$ for (b), $k = 2$ for (c), $k = 3$ for (d), $k = 4$ for (e) and $k = 5$ for (f)

according to equation (2.28) and showed in Figure 5.5. Noted that the input sources are all zero-mean unit variance [19].

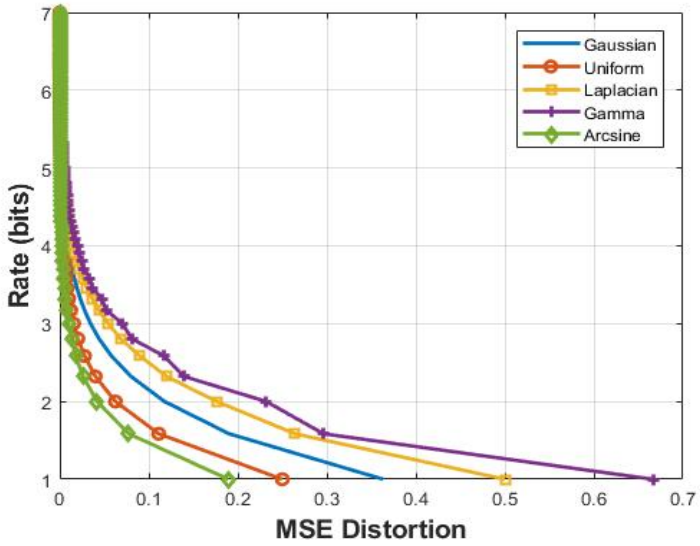


Figure 5.5 Entropy Rate-distortion curve of pdf-optimized quantizer for Gaussian, Uniform, Laplacian, Gamma and Arc-sine sources.

For the purpose of comparison, an approximations has been made for the high-rate situations, which the quantization intervals $[x_k, x_{k+1}]$ can be assumed as same size. The distortion-rate function for pdf-optimized quantizers with fixed-length for such an assumption is derived in [29]

$$\begin{aligned} D(R) &\cong \varepsilon^2 \sigma_x^2 2^{-2R} \\ \varepsilon^2 &= \frac{1}{12\sigma_x^2}\left(\int_{-\infty}^{\infty} \sqrt[3]{f(x)}dx\right), \end{aligned} \tag{5.30}$$

where σ_x^2 is the variance of input random source X, the rate is calculated as $R = log_2(L)$, where L is the number of output level defined by the quantizer, the factor ε^2 in (5.30) for different pdf type of the input source has been calculated and showed in Table 5.1, so that the rate-distortion curve for high-rate approximation can be derived as well. As Figures 5.5 and 5.6 shows, for a given rate, the Arc-sine pdf source has the minimum distortion.

Table 5.1 Values of ε^2 for Fixed-length PDF-Optimized Quantizers for Gaussian, Uniform, Laplacian, Gamma and Arc-sine Sources.

Source	**Uniform**	**Gaussian**	**Laplacian**	**Gamma**	**Arc-sine**
Value of ε^2	1	2.721	4.500	5.622	0.597

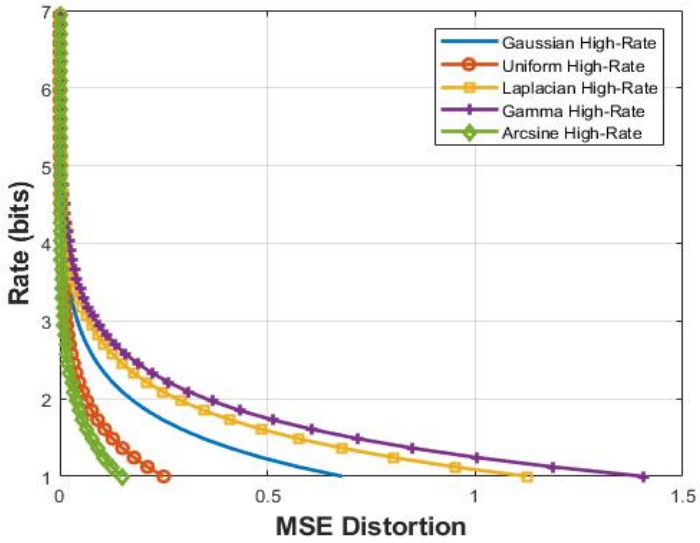


Figure 5.6 High-rate approximation rate-distortion curve for pdf-optimized quantizer, for Gaussian, Uniform, Laplacian, Gamma and Arc-sine sources.

5.4 Sparsity of Eigenportfolios

The rate-distortion theory based sparsity method was detailed in [43], and it is used to generate sparse eigenportfolios in this study. The midtread (zero-zone) quantizer type is employed to quantize each basis function (components of each vector) or the entire basis set of a transform to achieve a sparse representation. It is noted that only the center bin (zero-zone) of the mid-tread quantizer around zero is used in sparsity applications. The size of this zero-zone is adjusted to achieve the desired level of sparsity.

The calculated eigenvectors of empirical correlation matrix may have components (capital allocation coefficients) with small values. The maintenance of portfolios with large number of assets becomes burdensome and costly. It is a common practice to

avoid investing in assets of a portfolio with small capital allocations in the overall investment [9, 3]. A design framework to sparse portfolios was proposed in [43], and utilized in this study.

The eigenportfolios are generated based on eigenvectors of empirical correlation matrix obtained from past returns of a pre-selected basket of assets for a given time window. We used the returns of stocks in S&P500 index for the period of Dec. 1, 2015 to Dec. 1, 2017 for various sparsity levels in order to validate the merit of the sparsity method.

There are 492 tickers of S&P500 index that continue to exist during that time period. Therefore, we used their end of day (EOD) simple returns, $\mathbf{r}(n)$, to calculate the empirical correlation matrix at time n, $\mathbf{R}_E(n)$, as follows [4],[5]

$$\mathbf{r}\,(n) = [r_k\,(n)]\,;\ k = 1, 2, ...492 \tag{5.31}$$

and, the resulting empirical correlation matrix $\mathbf{R}_E(n)$ is calculated from equations (4.3) and (4.4). The time window of $W = 60$ is used. Note that the returns are normalized to zero mean and unit variance, and $\mathbf{R}_E\,(n)$ is real, symmetric and positive definite matrix. Then, the eigenmatrix $\mathbf{A}_{KLT}$ of $\mathbf{R}_E\,(n)$ that satisfies the eigenmatrix decomposition property as given below is obtained [2, 5, 43]

Now, we focus on the elements of $\mathbf{A}_{KLT}$ with small values. The histogram of the elements of an eigenmatrix $\mathbf{A}_{KLT}$ is displayed in Figure 5.7.

Then, design a mid-tread (zero-zone) pdf-optimized quantizer for this histogram to replace the small valued elements of $\mathbf{A}_{KLT}$ by zero. Adjusting the zero-zone of the quantizer by simultaneously adding the neighboring pairs of intervals on both sides to achieve the desired sparsity level. The zero-zone of the quantizer is used for this application.

Each eigenvector may have different histogram, and one can design a separate quantizer for every one of them in particular for portfolios with large number of

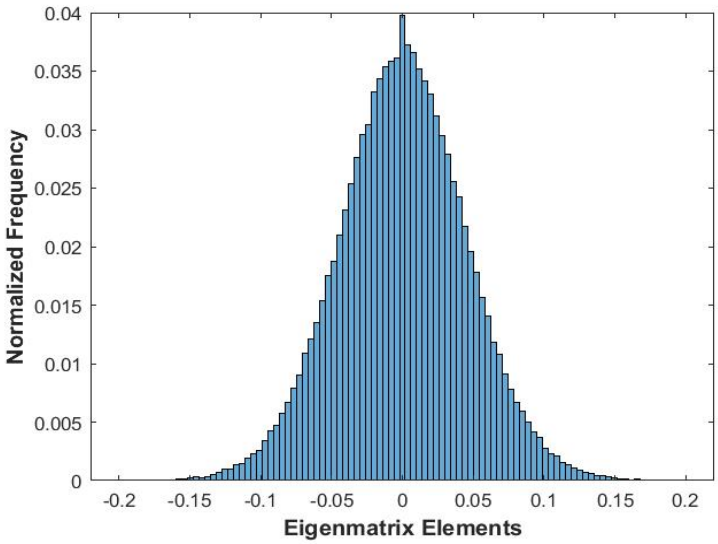


Figure 5.7 Normalized histogram of eigenmatrix elements for empirical correlation matrix of end of day (EOD) returns for 492 stocks in S&P 500 index with $W = 60$-day measurement window starting on December 1st, 2015, ending on February 26, 2016, and widow shifts for another 252 days, starting on February 26, 2016, ending on February 24, 2017.

assets. The explained variance (eigenvalue) of the eigenvectors $\boldsymbol{\phi}_k(n)$ is investigated in order to identify significant ones and evaluate their eigenportfolio performance. The k^{th} eigenvalue at time n, $\lambda_k(n)$, is equivalent to the variance of the k^{th} transform coefficients, σ_k^2, and calculated as $\left\{\lambda_k(n) = \boldsymbol{\phi}_k^{\mathrm{T}}(n)\,\mathbf{R}_E(n)\,\boldsymbol{\phi}_k(n)\right\}\ \forall k$ [2]. Figure 5.8 illustrates the the cumulative explained variance as a function of the number of eigenvalues for various sparsity levels. Similarly, the explained variances (eigenvalues) of the sparsed eigenvectors at time n, $\left\{\hat{\boldsymbol{\phi}}_k(n)\right\}$, are calculated as $\left\{\hat{\lambda}_k(n) = \hat{\boldsymbol{\phi}}_k^{\mathrm{T}}(n)\,\mathbf{R}_E(n)\,\hat{\boldsymbol{\phi}}_k(n)\right\}\ \forall k$.

Note that first 59 out of 492 eigenvalues (eigenvectors or principle components) explained 99% of the total variance (of the random vector process) in this experiment. A specific pdf-optimized quantizer for each one of the first 25 eigenvectors is designed, that explain 80% of the total variance.

The relationship between the sparsity and the resulting variance loss for these 25 eigenvectors are displayed in Figure 5.9. The variance loss of the k^{th} eigenvector

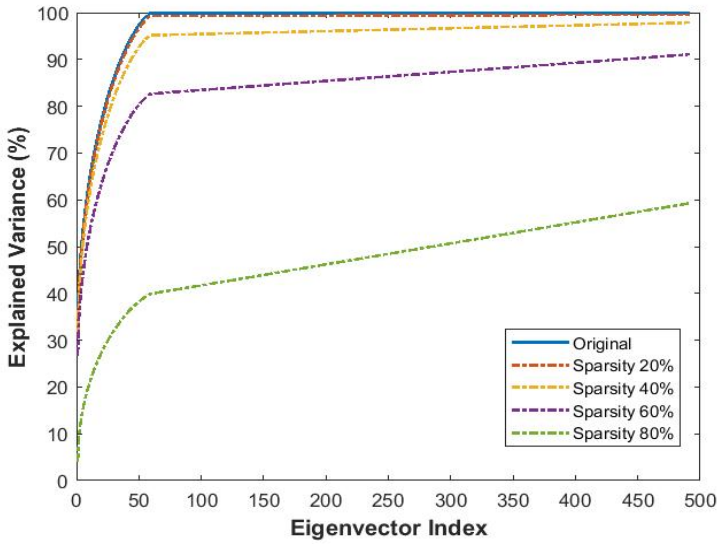


Figure 5.8 Cumulative explained variances as a function of the number of eigenvectors included in representation for different levels of sparsity.

due to the sparsity is defined as

$$E\left\{VL_k\left(n\right)\right\} = \frac{1}{N}\sum_{n=0}^{N-1} VL_k\left(n\right) \tag{5.32}$$

where $\left\{VL_k\left(n\right) = \left(1 - \frac{\hat{\lambda}_k(n)}{\lambda_k(n)}\right) \times 100\right\} \forall k$.

The \$1 normalized investment is assumed in each eigenportfolio with long and short positions, in general, and no transaction cost is considered in these experiments. Then, the Profit and Loss (PNL) curve is calculated [44]. It is observed from Figure 5.10 and Table 5.2 that the sparse eigenportfolios with significant reduction in transaction cost perform similar to the original eigenportfolios for the stocks of S&P500 index stocks.

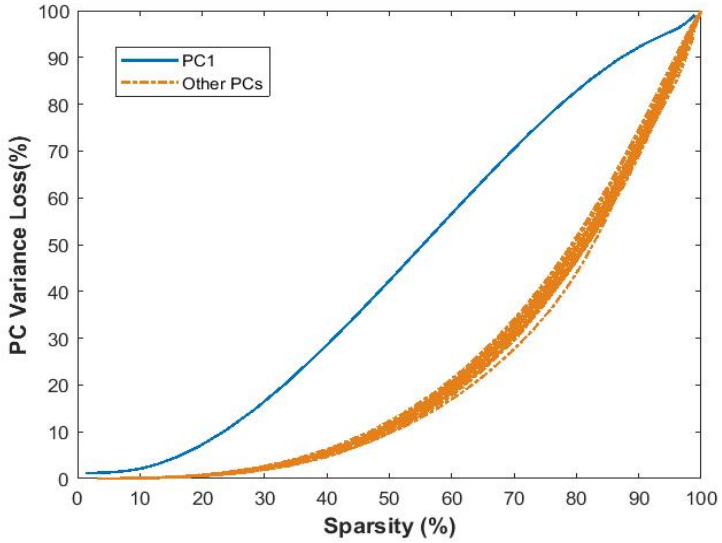


Figure 5.9 Variance loss of first 25 eigenvectors (principle components) as a function of sparsity level.

Table 5.2 Annualized Sharpe Ratios of Each Eigenportfolios at Different Level of Sparsity, Using S&P 500 EOD Returns, Starting on February 26, 2016, Ending on February 24, 2017.

Sharpe Ratio	EP 1	EP 2	EP 3	EP 4	EP 5	EP 6
Original	1.88	-0.77	0.93	1.3	1.29	-0.42
20% Sparsity	1.81	-0.78	0.94	1.27	1.29	-0.41
40% Sparsity	1.77	-0.78	0.96	1.36	1.4	-0.38
60% Sparsity	1.66	-0.74	0.88	1.44	1.57	-0.46
80% Sparsity	1.64	-0.37	0.93	1.48	1.48	-0.78

5.5 Chapter Summary

In this chapter, the probability density function of each eigenvector for the exponential correlation matrix is derived. The rate-distortions of applying the pdf-optimized quantizers on various types of distribution sources are measured, it is shown that the arc-sine distribution source has the optimal rate-distortion curve.

Since portfolio managers and investors desire to have smaller number of positions to open and rebalance. Therefore, one needs to develop a methodology

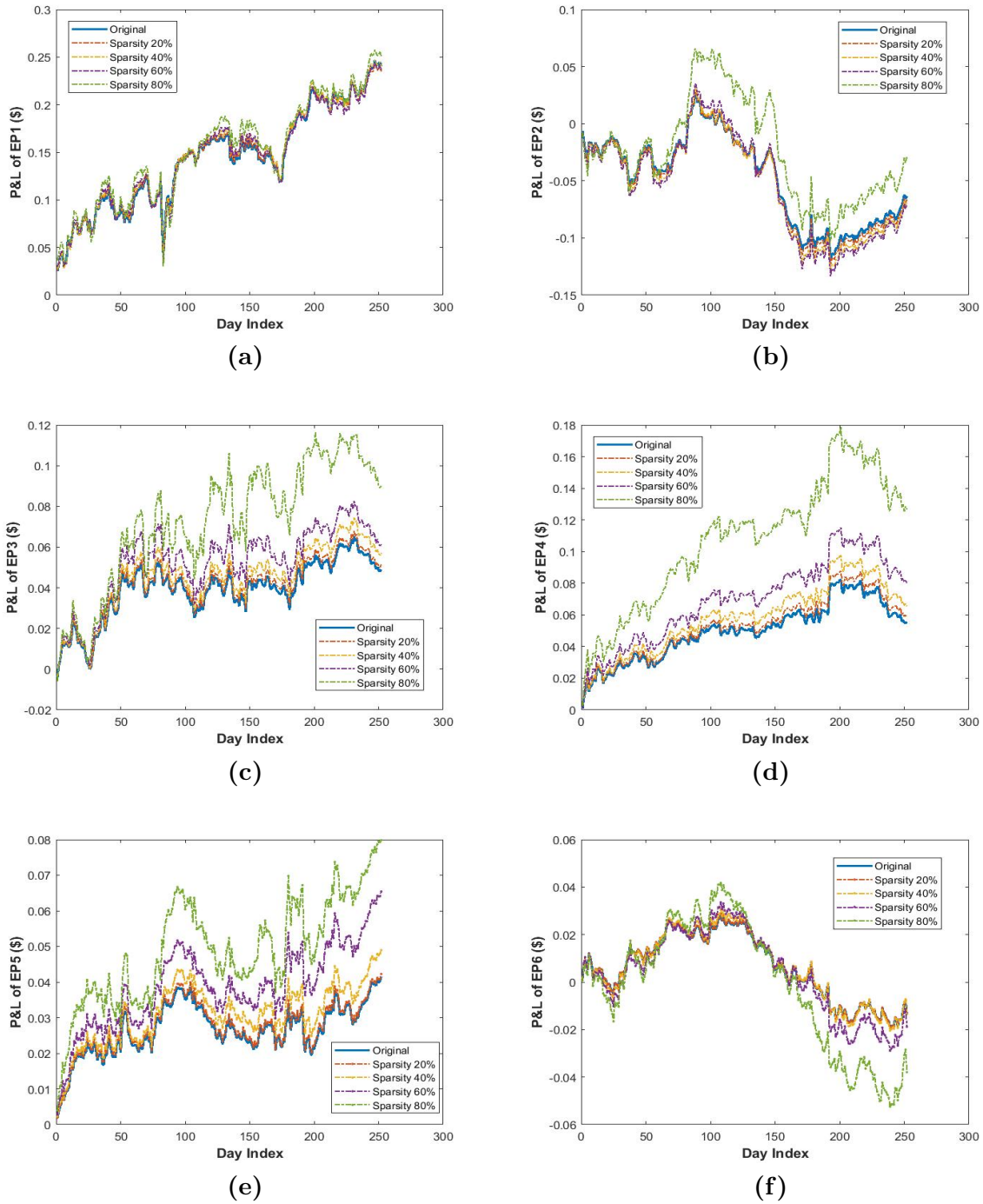


Figure 5.10 PNL curves of the original and sparsed eigenportfolios for EOD returns of stocks in S&P500 index between February 26, 2016 and February 24, 2017.

to define a threshold where an investment allocation is deemed insignificant. This problem becomes important for very large size portfolios, e.g., Russell 2000, VTI, VGTSX) and it can be formulated under the rate-distortion theory and a solution by using the mid-tread (zero-zone) pdf-optimized quantizer to sparse orthogonal subspaces was proposed in the literature. Noted that the optimal quantizers are tuned for different portfolios with desired sparsity levels. Usually, the execution related concerns, i.e. hard to find ticker, lot size, and trading cost, are known in advance, and implemented in the adjustment of the zero-zone pdf-optimized quantizer, accordingly. In this chapter, the performance of such quantizers to sparse eigenportfolios of stock returns in S&P500 Index is investigated. It is shown that the method is simple to implement and it provides high levels of sparsity without causing PNL loss. Consequently, transaction cost of maintaining a large size portfolio is reduced by employing such an efficient sparsity method.

CHAPTER 6

PARTITIONS OF EIGENSUBSPACE: DIMENSION VS CORRELATION

The modern portfolio theory (MPT) provides a framework for portfolio optimization where one can design an optimal portfolio with the minimum variance (MVP) or market portfolio (MP) with minimized portfolio risk for the desired expected return [27, 26]. However, MVP and MP are not necessarily the optimal portfolios for *out-of-sample* data since they depend on the inverse of covariance (correlation) matrix of asset returns that is not always robust for noisy market data. Therefore, the portfolio solutions could be totally different for the *in-sample* and *out-of-sample cases*, especially for large size covariance matrices.

Hierarchical risk parity (HRP) portfolio design avoids using the inverse of covariance matrix. It employs a clustering (partitioning) technique on the correlation matrix and creates N-asset portfolio by using the hierarchical structure of clusters. It is reported that HRP with the built-in hierarchical clustering feature may yield more stable correlation structures [25].

The eigenportfolios (EPs) are designed by using the eigenvectors of the empirical correlation matrix, and portfolios returns are pairwise decorrelated [7, 6, 42]. Hence, eigenportfolio returns are also used as the independent variables of regression for asset returns in some trading strategies including statistical arbitrage [4, 7]. It was reported in the literature that the first eigenportfolio (EP1) outperforms MVP and MP [15, 17, 11, 42].

In this chapter, a correlation based set partitioning algorithm is proposed. The impact of correlation and subspace dimension on eigenportfolio performance is emphasized. Then, these points is incorporated in the design of size-N portfolios that are comprised of properly weighted eigenportfolios of asset partitions. It is

shown that the proposed portfolio type significantly outperforms MVP, MP, IVP, and HRP based on profit and loss (PNL) curve and Sharpe ratio (SR) for all backtesting cases considered in the chapter. Moreover, both of its out-of-sample PNL and SR performances are displayed over 20% better than the DJIA index ETF DIA for the *survivorship* bias free case in the period from June 8, 2009 to Aug 30, 2019.

6.1 Impact of Dimension and Correlation for Energy Compaction

Subspace performance improves when dimension N gets larger. Performance is also better when correlations in $\mathbf{R}$ get higher [2] for the given N as shown in Figure 6.1

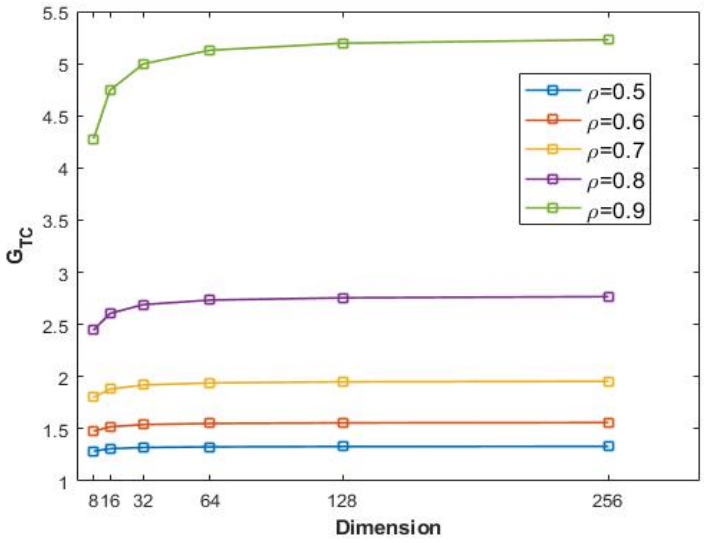


Figure 6.1 G_{TC} performance as a function of dimension N for eigensubspace of exponential correlation matrix $\mathbf{R}$ with $\rho = 0.5, 0.6, 0.7, 0.8, 0.9$.

Thus, correlation based partitioning of the group (set) Ω of N assets into K subgroups $\{\Omega_p\}$ $1 \leq p \leq K$ with N_{Ω_p} assets each, $1 \leq N_{\Omega_p} \leq N$ where $\Omega = \Omega_1 \cup \Omega_2 \cup .. \cup \Omega_K$; $\Omega_k \cap \Omega_l = 0 \, \forall k, l \, k \neq l$ and $N = \sum_{p=1}^{K} N_{\Omega_p}$, facilitates higher intra-subgroup (partition) correlations that may yield improved overall performance with combined subspaces of smaller dimensions than the original one with N dimensions. This is the theoretical reasoning to judiciously partition large size asset groups into their subgroups (partitions) with improved correlations. Hence, one

needs to simultaneously assess the performance trade-off between higher correlation and reduced dimension for a given problem [2].

6.2 Correlation Based Set Partitioning Algorithm

In this section, the procedure to partition a set of N assets into subsets based on pairwise correlations of normalized returns that populate their $N \times N$ empirical correlation matrix is introduced, $\mathbf{R} \triangleq [R(k,l) = \rho_{k,l}]$ as calculated by using equation (4.4). Due to the symmetrical nature of $\mathbf{R}$, there are $\frac{N(N-1)}{2}$ unique pairs of assets with their correlation coefficients. The algorithm employs a pre-defined threshold ρ_{TH} to avoid price co-movements of low correlation, and it is described as follows.

Step 1: The unique $\frac{N(N-1)}{2}$ correlation coefficients $R(k,l) = \rho_{k,l}$ $k \neq l$ are placed in the one-dimensional array $\boldsymbol{\rho}$ by following the descending order.

Step 2: The components of $\boldsymbol{\rho}$ are thresholded and $\rho(k) > \rho_{TH}$ $k = 1, 2, ..., \frac{N(N-1)}{2}$ populate the one-dimensional array of size L, $\boldsymbol{\rho}_d = [\rho_{m_l,n_l}(l)]$ $l = 1, 2, ..., L \leq \frac{N(N-1)}{2}$, in the descending order where (m_l, n_l) label the two assets related to this correlation coefficient.

Step 3: For $l = 1$, examine the components of $\boldsymbol{\rho}_d = [\rho_{m_l,n_l}(l)]$ in order to partition the set of assets involved. Start the process by placing (m_1, n_1) into the first partition $\Omega_1 = \{m_1, n_1\}$.

Step 4: For $l = 2$, the two assets (m_2, n_2) are regrouped as follows.

4.a: If none of them is a member of Ω_1, then, create a new partition $\Omega_2 = \{m_2, n_2\}$.

4.b: If one of them is in the set Ω_1, focus on the other asset to check if it is qualified to be in Ω_1 by calculating the average of its correlation coefficients with the current members of Ω_1 and threshold it with ρ_{TH}. If it is larger than ρ_{TH}, include

that asset in Ω_1. Otherwise, it will be re-examined during the assessment of the remaining vector components $\boldsymbol{\rho}_d = [\rho_{m_l,n_l}(l)]\ l = 3, 4, ..., L$.

Step 5: For $l = 3$, (m_3, n_3) of $\rho_{m_3,n_3}(3)$ are regrouped as follows.

5.a: If both of (m_3, n_3) belong to any existing partition, then, go to the next component of $\boldsymbol{\rho}_d$.

5.b: If none of (m_3, n_3) belongs to any of the existing partitions, then, create a new partition comprised of them.

5.c: If one of (m_3, n_3) is a member of an existing partition, then, check the other one if it is qualified to belong to the same partition by following the procedure given in 4.b.

Step 6: Repeat 5 for $l = 4, 5, ..., L$. There might be some assets that are not included in any existing partition. Then, each one of them is considered as an additional new partition.

Next, the merit of the proposed algorithm is showed to partition a set of assets into multiple subsets of assets based on co-movements (correlations) of their normalized returns. Let's consider a basket of five assets with normalized returns r_1, r_2, r_3, r_4, r_5 and their empirical correlation matrix as shown in Figure 6.2. Let's also assume $\rho_{TH} = 0.5$ in this example.

Step 3 yields the array $\boldsymbol{\rho}_d = [0.716, 0.703, 0.682]$ with $\rho_{1,3} = 0.716$ as its largest component. Hence, $\Omega_1 = \{1, 3\}$. In Step 4, $\rho_{2,4} = 0.703$ is identified as the second largest component of $\boldsymbol{\rho}_d$, and the new partition $\Omega_2 = \{2, 4\}$ is created. In Step 5, $\rho_{2,5} = 0.682$ is processed as the third largest component of $\boldsymbol{\rho}_d$. Since asset $2 \in \Omega_2$ but $5 \notin \Omega_2$,it is assessed if asset-5 is qualified to be included in Ω_2 according to Step 5.c. It is concluded that all of the five assets are partitioned in the subsets $\Omega_1 = \{1, 3\}$ and $\Omega_2 = \{2, 4, 5\}$. The partitioned version of the original empirical correlation matrix, $\mathbf{R}_P$, is also displayed as a heat-map in Figure 6.3.

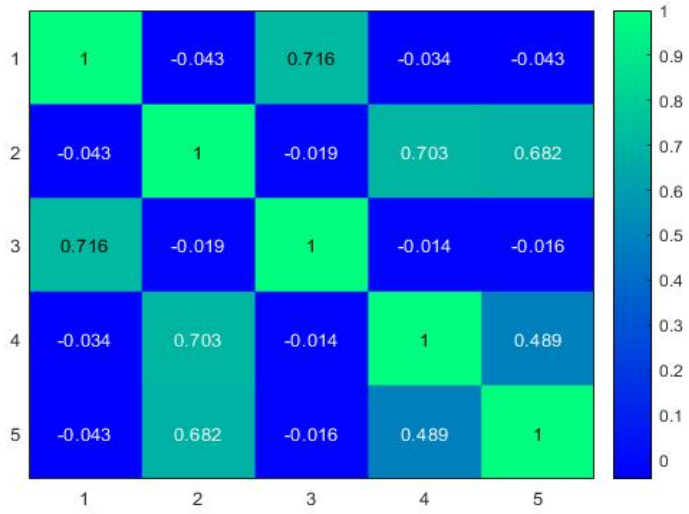


Figure 6.2 The empirical correlation matrix, $\mathbf{R} \triangleq [R(k,l) = \rho_{k,l}\, k, l = 1, 2, .., 5]$, of a basket displayed as a heat-map.

The original empirical correlation matrix $\mathbf{R}$ of Figure 6.2 with $N = 5$ is remapped into partitioned matrix $\mathbf{R}_P$ with the diagonal block sub-matrices $\mathbf{R}_P^{\Omega_1}$ and $\mathbf{R}_P^{\Omega_2}$ of interest with sizes $N_{\Omega_1} = 2$ and $N_{\Omega_2} = 3$, respectively. They represent the empirical correlation matrices of the resulting two partitions $\Omega_1 = \{1, 3\}$ and $\Omega_2 = \{2, 4, 5\}$ with the original asset indices. Therefore, $\mathbf{R}_P^{\Omega_1}$ and $\mathbf{R}_P^{\Omega_2}$ are independently eigendecomposed in order to design their eigenportfolios as explained in next section. The eigenportfolios of partitions are weighted and used to build size $N = N_{\Omega_1} + N_{\Omega_2}$ portfolios as described in the following section of the chapter.

6.3 Design of N-Asset Portfolio by Combining Eigenportfolios of Partitions

The traditional eigenportfolio design considers N-assets of the entire set Ω as explained in Chapter 4. On the other hand, it was argued that partitioning Ω into subgroups $\{\Omega_k\}\, k = 1, 2, .., K$ with higher intra-subgroup correlations and creating their eigenportfolios independently may improve the overall subspace performance. Now, let's investigate these two portfolio design approaches and compare their performances for various scenarios with market data.

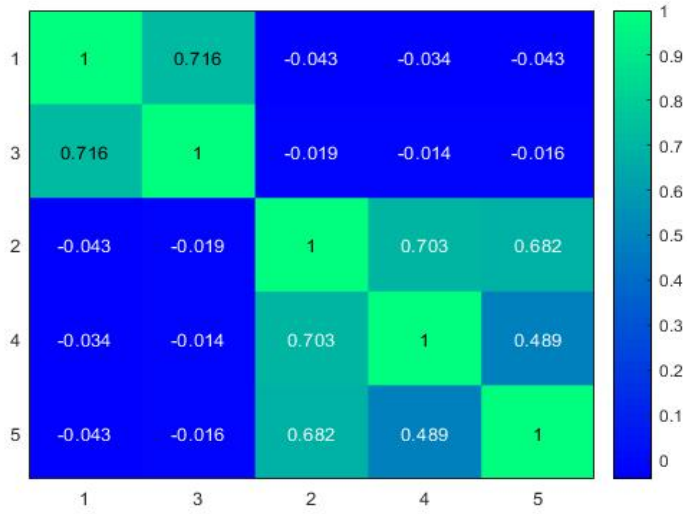


Figure 6.3 The partitioned empirical correlation matrix $\mathbf{R}_P$ with its diagonal block sub-matrices $\mathbf{R}_P^{\Omega_1}$ and $\mathbf{R}_P^{\Omega_2}$ of sizes $N_{\Omega_1} = 2$ and $N_{\Omega_2} = 3$, respectively, shown as a heat-map.

The eigen-decomposition of the empirical correlation matrix $\mathbf{R}_P^{\Omega_k}$ of the k^{th}partition Ω_k with size N_{Ω_1} is given as

$$\mathbf{\Lambda}^{N_{\Omega_k}} = [\mathbf{A}_{KLT}^{N_{\Omega_k}}]^T \mathbf{R}_P^{\Omega_k} [\mathbf{A}_{KLT}^{N_{\Omega_k}}] \; k = 1, 2, ..., K \tag{6.1}$$

Then, $\mathbf{A}_{KLT}^{N_{\Omega_k}}$ is normalized per equation (4.5) as

$$\widetilde{\mathbf{A}}_{KLT}^{N_{\Omega_k}} = [\Sigma^{N_{\Omega_k}}]^{-1} [\mathbf{A}_{KLT}^{N_{\Omega_k}}] [\mathbf{\Psi}^{N_{\Omega_k}}]^{-1}, \; \forall k \tag{6.2}$$

where the m^{th} column vector $\widetilde{\phi}_m^{N_{\Omega_k}}$ of $\widetilde{\mathbf{A}}_{KLT}^{N_{\Omega_k}}$ is used to design the m^{th} eigenportfolio of partition Ω_k. The returns of these N_{Ω_k} eigenportfolios are still perfectly decorrelated although the orthogonality of capital allocation vectors, $\{\widetilde{\phi}_m^{N_{\Omega_k}}\}$, is compromised due to the normalizations in equation (6.2). This process is repeated for all partitions, $k = 1, 2, ..., K$.

Now, one can build an N-asset portfolio by populating its capital allocation vector with the weighted eigenportfolios of desired partitions. As an example, we

can express the capital allocation vector of a size-N portfolio by combining the first eigenportfolios of all assets as

$$[\widetilde{\phi}^{N}_{combined}]^{T} = [w_1(\widetilde{\phi}_1^{N_{\Omega_1}})^{T}, w_2(\widetilde{\phi}_1^{N_{\Omega_2}})^{T}, ..., w_K(\widetilde{\phi}_1^{N_{\Omega_K}})^{T}] \tag{6.3}$$

where $N = \sum_{k=1}^{K} N_{\Omega_k}$,and the weighting coefficients $\{w_k\}$ are defined based on an established method. Three of such methods are discussed in next section. Note that the total positions of the capital allocation vector $\widetilde{\phi}^{N}_{combined}$ is also normalized before implementing the size-N portfolio.

6.3.1 Capital Allocation Among Eigenportfolios of Asset Partitions

The three methods to define capital allocations among eigenportfolios of asset partitions that are included in the creation of an N-asset portfolio as explained below. The methods by using the generation of the first size-N portfolio is described as an example. They can be generalized for various combinations of the partition portfolios.

a. Partition Size Based Weights The capital allocation coefficients $\{w_k\}$ of partitions are calculated according to the numbers of assets in partitions, $w_k = \frac{N_{\Omega_k}}{N}$ where $\sum_{k=1}^{K} w_k = 1$. Then, w_k is allocated among the tickers of the relevant eigenportfolio of the k^{th} partition. It is labeled as $\text{EP}_{partition\,A}$ in performance comparisons presented in the chapter.

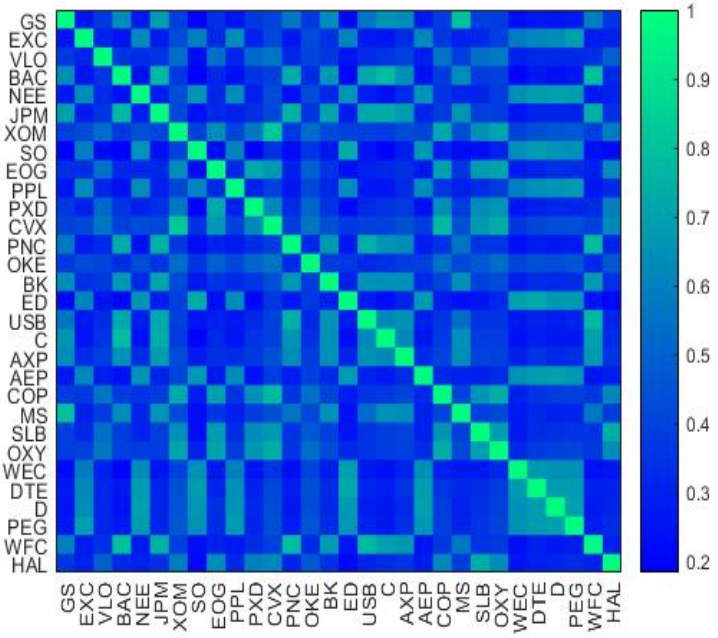
b. Inverse Volatility Based Weights The weighting coefficients are calculated based on the inverse volatilities of partitions $\{w_k = \frac{\sigma_{\Omega_k}^{-1}}{\sum_{k=1}^{K}\sigma_{\Omega_k}^{-1}}\}\ k = 1, 2, ..., K$, where $\sigma_{\Omega_k} = \left((\widetilde{\phi}_1^{N_{\Omega_k}})^{T}\mathbf{C}_P^{\Omega_k}(\widetilde{\phi}_1^{N_{\Omega_k}})\right)^{\frac{1}{2}}$, and $\mathbf{C}_P^{\Omega_k}$ is the covariance matrix of the k^{th} partition Ω_k. It is called as $\text{EP}_{partition\,B}$ in performance comparisons.

c. Eigenportfolio of Partition Eigenportfolios First, the returns of the first eigenportfolios of all partitions are derived, $\{\widetilde{\phi}_1^{N_{\Omega_k}}\}$ $k = 1, 2, ..., K$. Then, the eigendecomposition of their empirical correlation matrix is performed to generate the resulting super eigenportfolios, the columns of $\widetilde{\mathbf{A}}_{KLT}^{\Omega}$, as shown in equation (4.5). The coefficients of the first eigenportfolio are used as the weighting coefficients, $w_k = \widetilde{\phi}_1^{\Omega}(k)$ in order to create the size-N portfolio. It is cited as $\text{EP}_{partition\,C}$ in the chapter.

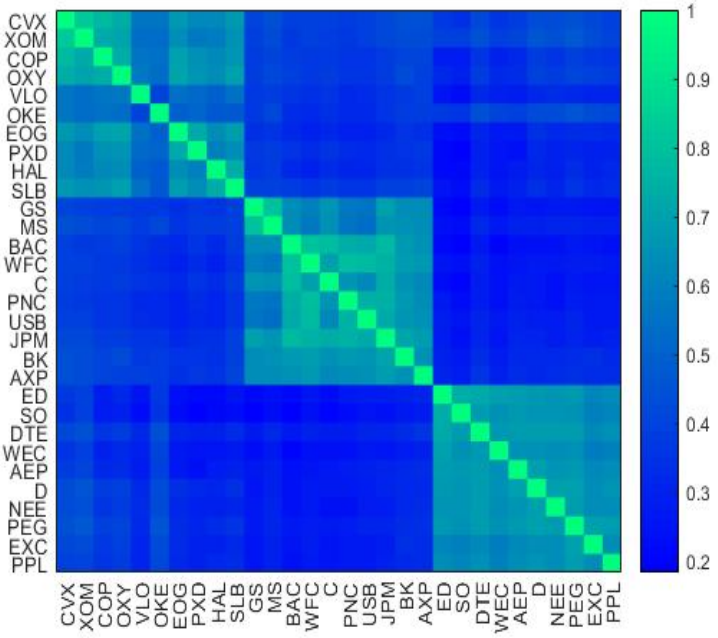
6.4 Performance Comparisons

In order to show the impact of the partitioning algorithm on the performance, we picked the most allocated 10 stocks of the three sector ETFs, XLE (Energy)[1], XLF (Finance)[2] and XLU (Utilities)[3]. We used their EOD return data from Aug 30, 1999 to Aug 30, 2019 for performance simulations. The empirical correlation matrix of size 30×30 is shown in Figure 6.4 as a heat-map. The PNL curves of minimum variance portfolio (MVP), market portfolio (MP) [4], inverse variance portfolio (IVP), Hierarchical Risk Parity (HRP) portfolio [25], first eigenportfolio (EP1) and size-N portfolios generated from partitions' first eigenportfolios by using the three different weighting methods, $\text{EP}_{partition\,A}$, $\text{EP}_{partition\,B}$, $\text{EP}_{partition\,C}$, described in previous section. for the out-of-sample market data from Aug 29, 2000 to Aug 30, 2019 with $W = 252$ and $\rho_{TH} = 0.45$ are displayed in Figure 6.5. Their annualized returns, volatilities, Sharpe ratios and PNLs are tabulated in Table 6.1(a) for comparison purposes. Similarly, we also tabulated their performance for various time intervals of the same basket in Table 6.1(b)(c)(d) to show the robustness of the portfolio performances. These results highlight the superiority of the proposed size-N portfolio over the others under the same test conditions.

[1]COP, CVX, EOG, HAL, OKE, OXY, PXD, SLB, VLO, XOM
[2]AXP, BAC, BK, C, GS, JPM, MS, PNC, USB, WFC
[3]AEP, D, DTE, ED, EXC, NEE, PEG, PPL, SO, WEC



(a)



(b)

Figure 6.4 Empirical correlation matrix of the EOD asset returns for the 10 most allocated stocks of the three sector ETFs, XLE (Energy), XLF (Finance), XLU (Utilities), $N = 30$, from Aug. 30, 1999 to Aug. 30, 2019. (a) Original empirical correlation matrix, (b) partitioned empirical correlation matrix with $W = 252$ and $\rho_{TH} = 0.45$.

Next, the EOD returns of 28 DJIA stocks [4] from Aug 29, 2000 to Aug 30, 2019 are used as the market data for further performance comparisons. We generated the PNLs and tabulated their annualized returns, volatilities, Sharpe ratios and PNLs in Table 6.2 for $W = 252$ and $\rho_{TH} = 0.40$. These results also show the merit of the proposed size-N portfolio type.

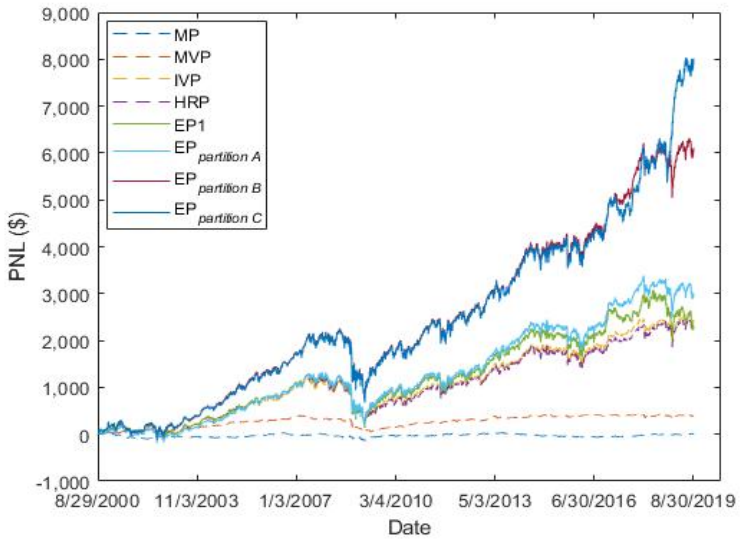


Figure 6.5 PNL curves of MVP, MP, IVP, HRP, EP1 and $\text{EP}_{partition\,A}$, $\text{EP}_{partition\,B}$, $\text{EP}_{partition\,C}$ for EOD returns of the most allocated 10 stocks in sector ETFs XLE, XLF and XLU from Aug. 29, 2000 to Aug. 30, 2019 with $W = 252$ and $\rho_{TH} = 0.45$. All portfolios are self-funded (available capital is always positioned) with initial investment of \$1,000, rebalanced daily without any trading cost involved.

Finally, the performance of the proposed size-N portfolios with the index ETF DIA are compared. It is noted that the DJIA has been changing its index tickers in time. In order to eliminate the built-in survivorship bias, the dataset of the simulations is updated to perfectly mimic DJIA components for the interval from June 8, 2009 to Aug 30, 2019. The PNL curves for all portfolios considered in the comparisons are generated with the parameters $W = 252$ and $\rho_{TH} = 0.40$ as displayed in Figure 6.6. Their annualized returns, volatilities, Sharpe ratios and PNLs are compared in Table 6.3. These results state that the size-30 portfolios designed by

[4] AAPL, AXP, BA, CAT, CSCO, CVX, DIS, DWDP, GS, HD, IBM, INTC, JNJ, JPM, KO, MCD, MMM, MRK, MSFT, NKE, PFE, PG, TRV, UNH, UTX, VZ, WMT, XOM.

the proposed method outperform DIA as well as the traditional first eigenportfolios (EP1) over 20% in their PNLs and Sharpe ratios.

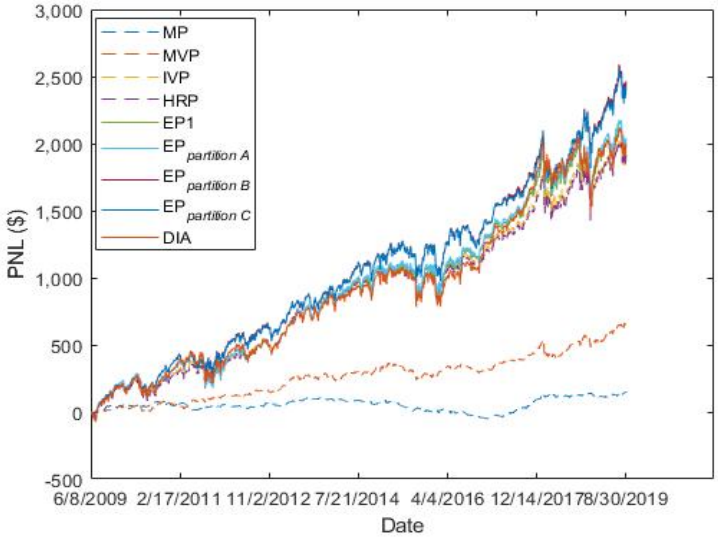


Figure 6.6 PNL curves of MVP, MP, IVP, HRP, EP1 and $\text{EP}_{partition\,A}$, $\text{EP}_{partition\,B}$, $\text{EP}_{partition\,C}$ and DIA for EOD returns of 30 stocks in the index DJIA from June 8, 2009 to Aug. 30, 2019 with the parameters $W = 252$ and $\rho_{TH} = 0.4$. All portfolios are self-funded (available capital is always positioned), with initial investment of \$1,000, rebalanced daily without any trading cost involved.

The amount of capital re-allocated for daily rebalancing of $\text{EP}_{partition\,C}$ is investigated in order to assess the impact of trading cost on the overall portfolio performance. Therefore, the histogram of capital re-allocations for 30 stocks of DJIA index (without survivorship bias) is created. On the average, 6.91% of portfolio value is traded every day for rebalancing. It means that 0.87% of the capital is spent for the trading cost with the assumption of rebalancing cost $5bps/day$. This number is 0.34% for the 30-stock portfolio case (10 most allocated stocks of the three sector ETFs). These results show that the trading cost does not have significant impact on the overall performance of the proposed portfolio type.

Furthermore, the probabilities for a stock to be a member of a certain size partition are calculated for DJIA index components in the interval from June 8, 2009 to Aug 30, 2019 as displayed in the histogram in Figure 6.7. It summarizes

the significance of the proposed set partitioning algorithm on the dimensions of the resulting eigenportfolios with various sizes. It is noted that the traditional first eigenportfolio is always comprised of the entire set of the index components. Hence, the first tracks the variations of correlations more closely than the latter for portfolio design.

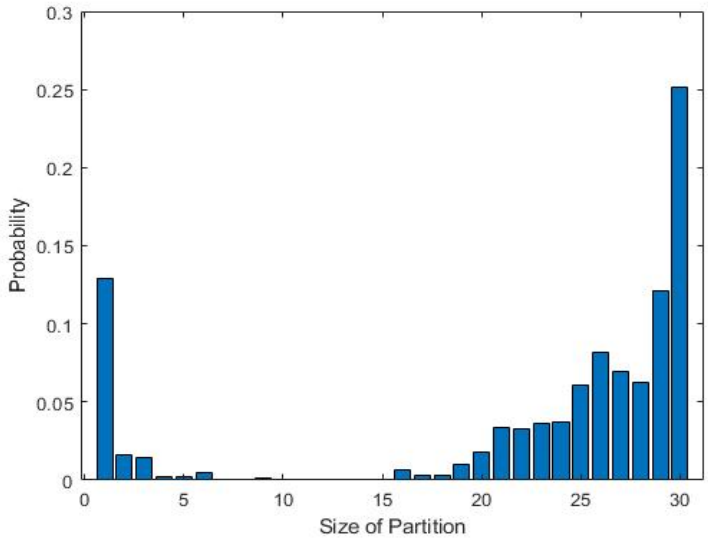


Figure 6.7 The probabilities for a stock to be a member of a certain size partition for DJIA index components in the interval from June 8, 2009 to Aug 30, 2019 with the parameters $W = 252$ and $\rho_{TH} = 0.4$.

Table 6.1 Annualized Returns, Volatilities, and Sharpe Ratios, and PNLs of MP, MVP, IVP, HRP, EP1 and $\text{EP}_{partition\,A}$, $\text{EP}_{partition\,B}$, $\text{EP}_{partition\,C}$ with $W = 252$, $\rho_{TH} = 0.45$ Calculated from PNL Charts of Figure 6.5 are Shown in (a). The Performances for the Same Dataset with Different Time Intervals are Tabulated in (b) from Aug 30, 2005 to Aug 30, 2019, (c) from Aug 31, 2010 to Aug 30, 2019, and (d) from Aug 31, 2015 to Aug 30, 2019. All Portfolios are Self-funded (Available Capital is Always Invested in Stocks) with Initial Investment of $1,000, Rebalanced Daily and No Trading Cost is Considered.

(a)

	MP	MVP	IVP	HRP	EP1	$\text{EP}_{partition\,A}$	$\text{EP}_{partition\,B}$	$\text{EP}_{partition\,C}$
μ (% annualized)	0.12	1.91	8.00	7.80	8.20	9.19	12.25	13.55
σ (% annualized)	4.29	5.57	17.20	17.14	19.34	19.39	19.55	19.76
SR (annualized)	0.03	0.34	0.47	0.45	0.42	0.47	0.63	0.69
PNL ($)	5	395	2,446	2,322	2,327	3,005	6,103	8,017

(b)

	MP	MVP	IVP	HRP	EP1	$\text{EP}_{partition\,A}$	$\text{EP}_{partition\,B}$	$\text{EP}_{partition\,C}$
μ (% annualized)	0.37	0.78	6.85	6.40	6.84	8.27	10.39	12.22
σ (% annualized)	4.10	5.24	17.57	17.43	20.17	20.10	19.89	20.11
SR (annualized)	0.09	0.15	0.39	0.37	0.34	0.41	0.52	0.61
PNL ($)	41	95	1,100	979	958	1,397	2,243	3,160

(c)

	MP	MVP	IVP	HRP	EP1	$\text{EP}_{partition\,A}$	$\text{EP}_{partition\,B}$	$\text{EP}_{partition\,C}$
μ (% annualized)	0.45	1.74	8.46	8.32	7.85	9.49	11.75	14.55
σ (% annualized)	3.19	4.62	12.83	12.88	15.20	14.50	14.29	14.76
SR (annualized)	0.14	0.38	0.66	0.65	0.52	0.65	0.82	0.99
PNL ($)	36	158	986	961	826	1,135	1,624	2,353

(d)

	MP	MVP	IVP	HRP	EP1	$\text{EP}_{partition\,A}$	$\text{EP}_{partition\,B}$	$\text{EP}_{partition\,C}$
μ (% annualized)	1.28	0.45	6.90	6.70	4.24	7.31	10.42	16.90
σ (% annualized)	2.92	4.72	11.74	11.95	14.97	13.30	12.15	13.32
SR (annualized)	0.44	0.10	0.59	0.56	0.28	0.55	0.86	1.27
PNL ($)	51	14	282	271	133	293	473	897

Table 6.2 Mean Returns, Standard Deviations, Annualized Sharpe Ratios and PNLs of MVP, MP, IVP, EP1 and $\text{EP}_{partition\,A}$, $\text{EP}_{partition\,B}$, $\text{EP}_{partition\,C}$ Calculated for EOD Return of 28 DJIA Stocks from Aug. 29, 2000 to Aug. 30, 2019 with $W = 252$ and $\rho_{TH} = 0.40$. All Portfolios are Self-funded (Available Capital is Always Positioned), with Initial Investment of \$1,000, Rebalanced Daily without Any Trading Cost Involved.

	MP	MVP	IVP	HRP	EP1	$\text{EP}_{partition\,A}$	$\text{EP}_{partition\,B}$	$\text{EP}_{partition\,C}$
μ (% annualized)	1.17	3.36	8.59	8.84	9.10	9.39	10.75	10.98
σ (% annualized)	5.08	6.75	15.88	15.60	17.44	17.28	17.11	17.08
SR (annualized)	0.23	0.50	0.54	0.57	0.52	0.54	0.63	0.64
PNL (\$)	219	813	3,020	3,246	3,216	3,471	4,833	5,099

Table 6.3 Annualized Returns, Volatilities, and Sharpe Ratios, and PNLs for MP, MVP, IVP, HRP, EP1 and $\text{EP}_{partition\,A}$, $\text{EP}_{partition\,B}$, $\text{EP}_{partition\,C}$ Along with DIA Calculated from Figure 6.6.

	MP	MVP	IVP	HRP	EP1	$\text{EP}_{partition\,A}$	$\text{EP}_{partition\,B}$	$\text{EP}_{partition\,C}$	DIA
μ (% annualized)	1.38	5.17	11.31	11.27	11.69	11.85	13.08	12.99	11.77
σ (% annualized)	3.88	5.58	12.82	12.55	13.90	13.88	13.50	13.49	14.22
SR (annualized)	0.36	0.93	0.88	0.90	0.84	0.85	0.97	0.96	0.83
PNL (\$)	143	670	1,921	1,921	1,993	2,042	2,468	2,438	2,005

6.5 Chapter Summary

In this chapter, a set partitioning algorithm is proposed to define asset partitions. The eigenportfolios of partitions are independently created. Then, these smaller size partition eigenportfolios with their calculated weights (allocation of total capital among eigenportfolios) are used to hierarchically design N-asset portfolios. The performance of the proposed portfolios are compared with the other N-asset portfolio types including the traditional eigenportfolios, MVP, MP, IVP, HRP and the DJIA index ETF DIA by generating their PNLs and the resulting Sharpe ratios for U.S. equities market data of various ticker sets and time intervals. It is concluded that N-

asset portfolios designed by using the proposed method consistently and significantly outperform the rest for all backtesting cases considered in the dissertation.

CHAPTER 7

CONCLUSIONS

The orthonormal subspaces and transforms have been one of the pillars of signal analysis and synthesis methods due to their tractability and ease of implementation. The eigensubspace, also known as KLT or PCA, is defined by the set of optimal basis functions (eigenvectors) and eigenvalues for the given random vector process with built-in perfect decorrelation and maximized energy compaction proerties. These optimal basis functions are derived from empirical correlation matrix of normalized returns for a basket of U.S. equities. They are used to design eigenportfolios for such a basket. The design of eigenportfolios and their performance analysis is the research focus of this dissertation. The findings and contributions of the study are summarized next.

7.1 Contributions of the Dissertation

1. Empirical correlations of returns for U.S. equities are approximated with an exponential function. Hence, exponential correlations populate a Toeplitz matrix where closed-form solutions for eigenvalues and eigenvectors exist. This framework is developed to model and generate eigenportfolios for a basket of stocks by using their price data from the market. It is shown that the model based eigenportfolios perform better or comparable to their traditionally designed counterparts. It is shown that they also consistently outperform the minimum variance portfolio (MVP), the market portfolio (MP), and the relevant sector ETF.

2. The pdf-optimized quantizers are designed and used to sparse eigenportfolios where the trading cost of their maintenance is significantly reduced. The theory of the sparsing procedure and its merit are investigated in the dissertation.

3. It is shown that lower dimensional eigensubspace with higher correlations provides better energy compaction performance than larger dimensional eigensubspace with lower correlation. A correlation based eigensubspace partitioning algorithm is developed and the trade-offs between correlation and dimension are investigated for U.S. equities. A novel method to design advanced portfolios

by judiciously combining eigenportfolios of partitions is proposed. It is shown that these new portfolios outperform the other known portfolio types including the traditional eigenportfolios, MVP, MP, IVP, HRP and index ETF for U.S. equities.

REFERENCES

[1] N. Ahmed, T. Natarajan, and K. R. Rao. Discrete cosine transfom. *IEEE Transactions on Computers*, 23(1):90–93, January 1974.

[2] A. N. Akansu and R. A. Haddad. *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets.* New York, NY, USA: Academic Press, Inc., 1992.

[3] A. N. Akansu, S. R. Kulkarni, and D. M. Malioutov. *Financial Signal Processing and Machine Learning.* Hoboken, New Jersey, USA: Wiley-IEEE Press, 2016.

[4] A. N. Akansu and M. U. Torun. *A Primer for Financial Engineering: Financial Signal Processing and Electronic Trading.* New York, NY, USA: Academic Press, Inc., 2015.

[5] A. N. Akansu and M.U. Torun. Toeplitz approximation to empirical correlation matrix of asset returns: A signal processing perspective. *IEEE Journal of Selected Topics in Signal Processing*, 6(4):319–326, Aug. 2012.

[6] A. N. Akansu and A. Xiong. Design of eigenportfolios for us equities using exponential correlation model. In *IEEE 53rd Annual Conference on Information Sciences and Systems (CISS)*, 2019.

[7] M. Avellaneda and J. H. Lee. Statistical arbitrage in the US equities market. *Quantitative Finance*, 10(7):761–782, 2010.

[8] C. M. Bishop. *Pattern Recognition and Machine Learning.* New York, NY, USA: Springer, 2006.

[9] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris. Sparse and stable markowitz portfolios. *Working Paper Series, European Central Bank*, (936):1–23, Sept 2008.

[10] G. Chamberlain and M. Rothschild. Arbitrage, factor structure and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304, 1983.

[11] R. Clarke, H. de Silva, and S. Thorley. Minimum-variance portfolio composition. *The Journal of Portfolio Management*, 37(2):31–45, 2010.

[12] S. Darolles, P. Duvaut, and E. Jay. *Multi-factor Models and Signal Processing Techniques: Application to Quantitative Finance.* Wiley-ISTE, Hoboken, NJ, USA, 2013.

[13] W. B. Davenport and W. L. Root. *An Introduction to the Theory of Random Signals and Noise.* New York, NY, USA: McGraw-Hill, 1958.

[14] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning: Data mining, Inference, and Prediction*. New York, NY, USA: Springer, 2001.

[15] Robert A. Haugen and Nardin L. Baker. The efficient market inefficiency of capitalization-weighted stock portfolios. *The Journal of Portfolio Management*, 17(3):35–40, 1991.

[16] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.

[17] Ravi Jagannathan and Tongshu Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683, 2003.

[18] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. New York, NY, USA: Springer, 2013.

[19] N. S. Jayant and P. Noll. *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Englewood Cliffs, NJ, USA: Prentice-Hall Professional Technical Reference, 1984.

[20] I. T. Jolliffee. *Principle Component Analysis, 2nd Ed.* New York, NY, USA: Springer-Verlag, New York, USA, 2002.

[21] K. Karhunen. Uber lineare methoden in der wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.*, 37:1–79, 1947.

[22] A. Leon-Garcia. *Probability, Statistics, and Random Processes for Electrical Engineering*. Upper Saddle River, NJ, USA: Prentice Hall, 2008.

[23] S. Lloyd. Least squares quantization in pcm. *Transactions on Information Theory*, 28(2):129–137, Mar 1982.

[24] M. Loeve. *Probability Theory*. Princeton, NJ, USA: D Van Nostrand, 1955.

[25] M. López de Prado. Building diversified portfolios that outperform out-of-sample. *Journal of Portfolio Management*, 42(4):59–69, 2016.

[26] H. M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. New York, NY, USA: Wiley, 1959.

[27] H. N. Markowitz. The optimization of a quadratic function subject to linear constraints. *Naval Research Logistics Quarterly, III*, pages 111–133, 1956.

[28] J. Max. Quantizing for minimum distortion. *IRE Transactions on Information Theory*, 6(1):7–12, Mar 1960.

[29] P. F. Panter and W. Dite. Quantization distortion in pulse count modulation with nonuniform spacing of levels. *Proceedings of the IRE*, 49:44–48, Jan 1951.

[30] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. New York, NY, USA: McGraw-Hill, 1965.

[31] A. Paulraj, B. Ottersten, R. Roy, A. Swindlehurst, G. Xu, and T. Kailath. Subspace methods for directions of arrival estimation. *Handbook of Statistics*, 10, 1993.

[32] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11):559–572, 1901.

[33] W. Ray and R. Driver. Further decomposition of the Karhunen-Loeve series representation of a stationary random process. *IEEE Transactions on Information Theory*, 16(6):663–668, September 1970.

[34] T. Roncalli. *Introduction to Risk Parity and Budgeting*. Chapman & Hal l/ CRC, London, UK, 2013.

[35] W. F. Sharpe. Mutual fund performance. *Journal of Business*, 39:119–138, 1966.

[36] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58:267–288, 1996.

[37] M. U. Torun and A. N. Akansu. On Toeplitz approximation to empirical correlation matrix of financial asset returns. In *IEEE 46th Annual Conference on Information Sciences and Systems (CISS)*, 2012.

[38] M. U. Torun and A. N. Akansu. An efficient method to derive explicit KLT kernel for first-order autoregressive discrete process. *IEEE Transactions on Signal Processing*, 61(15):3944–3953, Aug 2013.

[39] M. U. Torun, A. N. Akansu, and M. Avellaneda. Portfolio risk in multiple frequencies. *IEEE Signal Processing Magazine, Special Issue on Signal Processing for Financial Applications*, 28(5):61–71, Sep. 2011.

[40] R. S. Tsay. *Analysis of Financial Time Series*. Hoboken, New Jersey, USA: John Wiley & Sons, Inc., 2005.

[41] J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Oxford University Press, Oxford, UK, 1965.

[42] A. Xiong and A. N. Akansu. Performance comparison of minimum variance, market and eigen portfolios for us equities. In *IEEE 53rd Annual Conference on Information Sciences and Systems (CISS)*, 2019.

[43] O. Yilmaz and A. N. Akansu. Quantization of eigen subspace for sparse representation. *IEEE Transactions on Signal Processing*, 63(14):3616–3625, July 2015.

[44] O. Yilmaz and A. N. Akansu. Performance analysis of eigenportfolios for AR(1) process. In *Proc. of IEEE 50th Annual Conference on Information Sciences and Systems (CISS)*, 2016.