Effective Dimensionality Control in Quantitative Finance and Insurance

by

Kai Liu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2017

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Giray Ökten Professor
Supervisor	Ken Seng Tan Professor
Internal Member	Phelim Boyle Professor
Internal Member	Adam Kolkiewicz Professor
Internal-external Member	Ken Vetzal Associate Professor

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

It is well-known that dimension reduction techniques such as the Brownian bridge, principal component analysis, linear transformation could increase the efficiency of Quasi-Monte Carlo (QMC) methods. Caflisch et al. (1997), who introduced two notions of effective dimension known as superposition dimension and truncation dimension, in part explain the overwhelming success of these methods for high-dimensional finance applications.

By exploiting dimension reduction in QMC, we propose a new measure of effective dimension which we denote as the delta dimension. Unlike the previously proposed effective dimensions, it is easy to compute delta dimension, including its dimension distribution. We also propose a new dimension reduction technique known as the directional control (DC) method. By assigning appropriately the direction of importance of the given function, the proposed DC method optimally determines the generating matrix used to simulate the Brownian paths. Because of the flexibility of our proposed method, it can be shown that many of the existing dimension reduction methods are special cases of our proposed DC method. Furthermore, considering the functions with multiple discontinuities or differentiabilities, we propose a severity measure that allows us to identify the relative importance of the various sub-functions, which allows us to dynamically construct the optimal path generation method.

By exploiting dimension reduction techniques in portfolio of insurance contracts, we propose a real-time evaluation model, i.e. Green-mesh, to select smaller number of synthetic representative points. We show that our pre-computed values could be recycled for evaluating incoming contracts. Unlike the general machine learning method, our green-mesh real-time evaluation model only takes a little computing at time 0, and achieves much higher accuracy.

By exploiting dimension reduction techniques in portfolio selection, we propose an Effective Portfolio (EP) model to select smaller number of stocks in portfolio selection and uniquely determine the weights of selected stocks. We propose so called effective portfo-

lio. We show that only certain portion of stocks dominant the whole market in which we define the number of effective stocks as EPD. Unlike randomly selected stocks, our EPD is a counting random variable with corresponding probability mass function, and it can be shown that a better portfolio alpha and beta trade-off based on a sophisticated strategy could be achieved via dimension reduction.

Acknowledgements

Over the past few year, my friends have asked me many times as to when can I graduate, and it is a miracle that I could finally complete my PhD program.

First of all, I would like to thank my only lifetime supervisor, Master and PhD supervisor, Ken Seng Tan, who has finally allowed me to graduate. Wow My boss who appears everywhere in the world except Waterloo. Within the last few years, my first priority every week is to look for my boss, and I learnt a lot from him, e.g. academically travelling everywhere in the world, joining in others work, investing in real estate properties, advising female students on how to improve their studies (joking aside), as a result, my boss has become my best friend and the most important mentor in my life and academics, who pushed these academic buddies to me, Chou-wen Wang, Junichi Imai, Haifei Liu, Ning Zhang and Yan Zeng. I have to put these names here because they really swayed my career decisions from making hundreds of millions on Wallstreet and instead chose the academia life of $100K per year.

I would also like to thank my thesis committee members Prof. Phelim Boyle and Prof. Adam Kolkiewicz, internal external committee Prof. Ken Vetzal and external examiner Prof. Giray Ökten for taking their time and effort to read my thesis and provide insightful comments despite their extremely busy schedules. In addition, there are three people changed my life before age of 20, Prof. Bao-Sheng Zhu who trained me to reach 600 TOEFL score without knowing any English word. Teachers Sheng and Chen, who taught me how to complete in mathematical Olympics. Without them, my life will be totally different, and I would have no chance to study in University of Waterloo, and meet my supervisor and all the other lifelong friends.

In my academic life, I have to give special mention to my academic sister, Wenjun Zhu, who survived with me together in this journey, and my academic brothers Johnny Li, Chengguo Weng, and JianFa Cong, who gave me a lot of advice. Besides these, I have to mention Mu Zhu who adopted me as his research assistant, and gave me a lot of advise in

the last year of my undergraduate, Tony Wirjanto who supported me a lot in my career, Christine Lemiuex who agreed to be on my Masters thesis committee, and Haixin Duan who taught me how to make virus, and hack Bank of America in Tsinghua University, Bin Li, David Landriault, Ruodu Wang who pushed me to be their drink and KTV buddies, Pengfei Li who gave me a lot of advice in my life, Gordon Wilmot as one of the nicest professors who almost became my supervisor if Ken Seng came late, Keith Freeland and Fan Yang who helped me a lot on teaching, Tom Coleman who gave me his own book of computing Jacobian matrix, Harry Panjer who taught my first ACTSC course, and Mary Lou on anything else in this university.

In my growth, I would like to thank my family for their love, accompany and unconditional support. Further thanks go to my buddies Songge (Vincent) He, Yunxuan (Yvonne) Wu, Big Brother and Sister (Hui Chen and Yunyin Shan), Yuntao (Malcolm) Yang, Mingbin (Ben) Feng and Jiayi (Jessica) Li, Qiyi (Gina) Fu, Zhe (Diana) Wang, Qianyao (Christina) Tan, Fu Xing F4 other members (Meng Dai, Qijun Chen, Kai Ma) for their exceptional support.

Finally, I would like to mention Lichen Chen, Yanxin (Curly) Liu, Shu Li, Lu Chen, Fangyuan (Sally) Lin, Fangda Liu, Doncheng Li, Di Xu, Yuchen Mei, Dezhao Han, Garcia Liang, Lu Xin, Kexin Ji, Hengxing Cui, Mingyu Fang, Mengying Deng and her Husband Tianxiang (Tim) Shi, Yanjun Chen, Wei Wei, etc. because we were fighting with this university together.

This is not the end, but I have to stop this etc just like my thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Because of the complexity of derivative securities and the sophistication of financial models, the integrals associated with finance applications typically cannot be evaluated analytically. Consequently Monte Carlo (MC) method, which is first introduced in quantitative finance by Boyle (1977), becomes a popular numerical method. However, MC method attains a convergence rate of $O(1/\sqrt{N})$ where N is the number of simulation trials . MC method is often criticized to be a slow method despite the convergence rate is independent of the dimension. In mid 1990s, several reports have surfaced advocating the use of Quasi-Monte Carlo (QMC) methods, as opposed to the classical MC methods. QMC offers a convergence rate of $O\left(N^{-1}(\log N)^d\right)$ in dimension d. This rate is asymptotically more efficient than the MC. The results in Joy et al. (1996) and Paskov and Traub (1995) showed that QMC yields a much higher accuracy than the Monte Carlo (MC) method, even for several hundred dimensions. As a result of these findings and the theoretically more efficiency (than MC), there is a surge of interest among financial engineers and academicians in using QMC methods to computational finance.

One key area of research focus is to provide a better understanding on why QMC can be effective in finance applications. The analysis of variance (ANOVA) has been widely used in the study of quasi-Monte Carlo (QMC) integration methods, where it is applied to var-

ious notions of the effective dimension of an integrand. Using the ANOVA decomposition of a function, Caflisch et al. (1997) defined two notions of the effective dimension: truncation dimension and superposition dimension, in part explain the overwhelming success of these methods for high-dimensional finance applications. Essentially, the truncation dimension indicates the number of important variables which predominantly capture the evaluation function f, and the superposition dimension measures to what extent the low-order ANOVA terms dominate the function. By exploiting these definitions, Owen (2003) proposed dimension distribution for a square integrable function to provide additional insights to the effective dimension of a function.

However, when the dimension of a function is high, it is computationally burdensome to compute the ANOVA decomposition for an arbitrary function. The first topic of this thesis is to propose an easy-to-build effective dimension and dimension distribution for any arbitrary function, and argue that the conditional tail dimension (CTD) of a dimension random variable is a better measurement than the effective dimension.

Another area of research focus is to seek better QMC-based algorithms for evaluating high-dimensional integrals. We now provide a brief overview connecting MC and QMC methods to pricing European derivative securities. For a detailed exposition of these topics, see Boyle et al. (1997), Glasserman (2004) and Lemieux (2009). For simplicity, we assume that the dynamics of the asset price is governed by the Black-Scholes (BS) model so that the risk-neutral process of the underlying asset S_t at time t, is given by

$$dS_t = rS_t\,dt + \sigma S_t\,dB_t, \tag{1.1}$$

where r is the risk-free interest rate, σ is the volatility and B_t is the standard Brownian motion.

By $h(\mathbf{S}) = h(S_1, \ldots, S_d)$ we denote as the payoff function of a particular derivative security at maturity T years. Note that the payoff function depends on the asset prices $S_j := S_{t_j}$ at equally spaced times $t_j = j\Delta t$ for $j = 1, \ldots, d$ and $\Delta t = T/d$. According to the option

pricing theory, the value of the financial derivative at $t = 0$ is

$$\mathbb{E}\,[e^{-rT} h(\mathbf{S})], \tag{1.2}$$

where $\mathbb{E}\,[\cdot]$ is the expectation under the risk-neutral measure. For example, the price of a European arithmetic Asian option is $\mathbb{E}\,[e^{-rT}\,\max(S_A - K, 0)]$, where S_A is the arithmetic average of the underlying asset prices at times $t_1, \ldots, t_d$ and K is the strike price.

It is easy to verify that the payoff function $h(\mathbf{S})$ can be re-expressed as

$$h(\mathbf{S}) = h(\exp(\mu_1 + \sigma x_1), \ldots, \exp(\mu_d + \sigma x_d)) := g(\boldsymbol{x}), \tag{1.3}$$

where $\mu_j = \log S_0 + (r - \sigma^2/2)\, t_j$, for $j = 1, \cdots, d$, and $\boldsymbol{x} := (x_1, \ldots, x_d)^T \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$; i.e, $\boldsymbol{x}$ is normally distributed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ with its entry given by

$$\boldsymbol{\Sigma}_{ij} = \min(t_i, t_j) = \Delta t\, \min(i, j). \tag{1.4}$$

Note that in (1.3) we have redefined $h(\mathbf{S})$ as $g(\boldsymbol{x})$ to emphasize the explicit role of $\boldsymbol{x}$. Consequently, the time-0 value of the financial derivative can be expressed as a Gaussian integral:

$$V(g) := \mathbb{E}\,[g(\boldsymbol{x})] = \frac{e^{-rT}}{(2\pi)^{d/2}\sqrt{\det \boldsymbol{\Sigma}}} \int_{\mathbb{R}^d} g(\boldsymbol{x})\, \exp\left(-\frac{1}{2}\, \boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x}\right) d\boldsymbol{x}. \tag{1.5}$$

From the point of view of integration, by setting $\boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{z}$, where $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{\Sigma}$, $\boldsymbol{z} = (z_1, \ldots, z_d)^T \sim N_d(\mathbf{0}, I_d)$ and I_d is a $d \times d$ identity matrix, and then imposing the transformation $\boldsymbol{z} = \Phi^{-1}(\boldsymbol{u}) = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d))^T$, the componentwise inverse of the standard normal cumulative distribution function, the Gaussian integral (1.5) is transformed to

$$V(g) = \frac{e^{-rT}}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} G(\boldsymbol{A}\boldsymbol{z})\, \exp\left(-\frac{1}{2}\, \boldsymbol{z}^T \boldsymbol{z}\right) d\boldsymbol{z} = e^{-rT} \int_{[0,1]^d} G(\boldsymbol{A}\,\Phi^{-1}(\boldsymbol{u}))\, d\boldsymbol{u}. \tag{1.6}$$

The change of variables $\boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{z}$ is equivalent to the generation of the Brownian motions

$$(B_1, \ldots, B_d)^T = \boldsymbol{A}\,(z_1, \ldots, z_d)^T, \tag{1.7}$$

where B_j is the Brownian motion at time t_j. Hence we refer to $\boldsymbol{A}$ as the *generating matrix* of the Brownian motion. A key insight to the above transformation is that $\boldsymbol{A}$ can be arbitrary as long as it satisfies $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{\Sigma}$. Consequently, different specification of $\boldsymbol{A}$ yields different methods of generating the Brownian motions. These methods are commonly known as the path generating methods (PGMs) since they relate to the simulation of Brownian paths.

By defining $\boldsymbol{z}_k = \Phi^{-1}(\mathbf{u}_k)$ and $\mathcal{P} := \{\mathbf{u}_i, i = 1, \ldots, N\}$ as a low discrepancy point set over the unit cube $[0,1]^d$, the QMC estimate of (1.6) is given by

$$Q(g, \boldsymbol{A}, \mathcal{P}) = \frac{e^{-rT}}{N}\sum_{k=1}^{N} G(\boldsymbol{A}\,\boldsymbol{z}_k) = \frac{e^{-rT}}{N}\sum_{k=1}^{N} G(\boldsymbol{A}\,\Phi^{-1}(\mathbf{u}_k)),$$

Clearly, the accuracy of the above estimate depends on the point set $\mathcal{P}$, the payoff function g and the generating matrix $\boldsymbol{A}$. While it is known that the MC algorithms based on different PGMs, i.e. different $\boldsymbol{A}$, are equivalent since the mean square error of MC is determined by the variance of the integrand, which is unchanged, different PGMs have significant impact on QMC. This phenomenon arises as the resulting function $G(\boldsymbol{A}\,\Phi^{-1}(\boldsymbol{u}))$ may have different dimension structure and may also induce different smoothness property depending on the chosen $\boldsymbol{A}$. In particular, it is widely believed that techniques that reduce the effective dimension of the function G increases the efficiency of QMC. As such, many dimension reduction techniques such as the Brownian bridge (BB) (Moskowitz and Caflisch 1996), the principal component analysis (PCA) (Acworth et al 1998), the linear transformation (LT) (Imai and Tan 2006), orthogonal transformation on discontinuous function (OT) (Wang and Tan 2012) and fast orthogonal transformation (FOT) method (Leobacher 2012), have been proposed to increase the efficiency of QMC.

Papageorgiou (2002) pointed out that any decomposition $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{\Sigma}$ provides a construction

for a discrete approximation of a Brownian path via $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z}$, where $\boldsymbol{Z}$ is a standard normal vector. In this context, the forward construction corresponds to the Cholesky decomposition of $\boldsymbol{\Sigma}$. However, Wang and Sloan (2011) show an "equivalence principle", which roughly states that every decomposition is equally bad and good for QMC, depending on the function that one wants to integrate. In other words, every decomposition that is good for one payoff function is bad for another.

This thesis designs the QMC-friendly path generation matrix (PGM) explicitly depending on the function of interest under Delta approach for any arbitrary function. The proposed DC method basically comprises of the following two phases:

Step 1. The first phase is a change of coordinates (variables), where we can transform an original function so that the function after this transformation can be handled effectively. The induction of the change of coordinates gives us another degree of freedom that can be used to improve numerical efficiency. In practice, we judiciously choose how to conduct the change of coordinates to reduce the nominal dimension of the function. In order to keep the transformation meaningful for our purpose, the first order of Taylor expansion is utilized. This operation is known as the delta method.

Step 2. In the second phase of the design, we determine an optimal path-generation matrix (PGM) based on the functional covariance matrix, namely, the covariance matrix of the function after the transformation. Note that most of the existing dimension reduction methods are based on a decomposition with respect to a given covariance matrix for the underlying process. All possible path-generation methods, such as Cholesky, BB, PCA and LT could be used in this phase, in response to the structure of the transformed function, that further enhances the numerical efficiency of the QMC. It is worth emphasizing that although the DC approach looks complicated at first sight due to the introduction of the functional covariance matrix, it can facilitate to specify an optimal PGM as will be demonstrated in our numerical demonstration.

Furthermore, we also show that our DC methods can handle multiple non-differentiability and multiple high-dimensionality at the same time and is equipped with enough flexibility to handle complicated functions which cannot be properly handled by many other dimension reduction methods.

When using a PGM to simulate trajectory of stock prices at time periods $1, 2, \ldots, d$, there is an implicit ordering of how stock prices are being simulated. For example, the classical STD method, with its path generation matrix derived from the Cholesky decomposition, simulates the stock prices given S_0 in the order $\{\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_d\}$. On the other hand, the PGM based on BB generates stock prices in the following order $\{\hat{S}_d, \hat{S}_{d/2}, \hat{S}_{3d/4}, \ldots, \hat{S}_1\}$, assuming d is a power of 2. Similarly the LT method constructs the path generation matrix by optimizing its first column, then second column, and iteratively until the d-th column. A natural question to consider is that for these PGMs if there exists a better "ordering"? More specifically, is it always optimal to first simulate the terminal stock price for the BB method? Similarly, for the LT method is it always optimal to optimize the columns starting from the first column? The "optimal" ordering of the BB construction has been addressed in Lin and Wang (2008) who demonstrate that under their prescribed optimality, it is never optimal to first simulate the terminal stock price. In fact they formally establish that the first optimal stock price to be simulated is $\hat{S}_{3d/4}$.

Under the functions with multiple discontinuities or non-differentiabilities, we propose a new measure that allows us to identify the relative importance of the various sub-functions, which is called severity measure. Based on that, we propose a new PGM that explicitly exploits the hierarchical order of functions. It turns out that once the preferred order is determined, the flexibility of the LT method implies that an optimal generation matrix can be determined accordingly. We label the resulting PGM that integrates the LT method with the hierarchical order as the hierarchical liner transformation (HLT) method.

Motivated by the theme of QMC, we extend our dimension reduction on pricing large portfolio of insurance contracts. In the last few decades, variable annuities (VAs) have

become one of the innovative investment-linked insurance products for the retirees. VA is a type of annuity that offers investors the opportunity to generate higher rates of returns by investing in equity and bond subaccounts. Its innovation stems from a variety of embedded guaranteed riders wrapped around a traditional investment-linked insurance product. The appealing features of these guarantees spark considerable growth of the VA markets around the world. According to the Insured Retirement Institute, the VA total sales in the U.S. were $130 billion and $138 billion for 2015 and 2014, respectively. Consequently many insurance companies are managing large VAs portfolio involving hundreds of thousands of policies. This in turn exposes insurance companies to significant financial risks, and hence heightens the need for an effective risk management program (such as the calculation of VA's sensitivity or Greeks to underlying market risk factors) for VAs. Because of the complexity of these products, closed-form pricing formulas exist only in some rather restrictive modelling assumptions and simplified guarantee features. Hence, it is a computational burden to price a portfolio of VA contracts via Monte-Carlo simulation.

Machine learning based methods such as the Kriging method or other spatial interpolation methods were proposed to accomplish this task by many researchers. Although these methods have already achieved a significant reduction in computational time, there are some potential outstanding issues. We identify the following six issues and summon that an efficient large VAs portfolio valuation algorithm should adequately address all of these issues:

1. the complexity of the proposed algorithm,
2. the cost of finding representative VA policies,
3. the cost of initial training set, if any,
4. the cost of estimating the entire large VAs from the representative VA policies,
5. the computer memory constraint,
6. the portability to other large VAs portfolio valuation.

Inevitability of these issues become more pronounced with the size of the VAs portfolios and the representative VA policies. More concretely, if Monte Carlo method were to price a VAs portfolio consisting of 200,000 policies, the time needed is 1042 seconds, as reported in Table 3 of Gan (2013). If one were to implement the method proposed by Gan (2013), the computational time reduces remarkably by about 70 times with 100 representative VA policies. However, if one were to increase the representative VA policies to 2000, the reduction in computational time drops from 70 to 14 times. While we are still able to achieve a 14-fold reduction in computational time, the deterioration of the proposed method is obvious.

Alternatively, we provide another compromised solution attempting to alleviate all of the issues mentioned above. More specifically, our proposed solution is based on the evaluation of low discrepancy synthetic representative points. It has a number of appealing features including its simplicity, ease of implementation, less computer memory, etc. More importantly, the overall computational time is comparatively less and hence our proposed method is a potential real-time solution to the problem of interest. Finally, unlike most other competitive algorithms, our proposed method is portable, i.e. the results can be recycled for other large VAs portfolio valuation.

Furthermore, motivated by the theme of QMC, our fifth topic is to import the idea of dimension reduction into portfolio selection. In about the fourth century, Rabbi Issac bar Aha proposed the following rule for asset allocation: One should always divide his wealth into three parts: a third in land, a third in merchandise, and a third ready to hand. In the literature on asset allocation, there have been considerable advances starting with the pathbreaking work of Markowitz (1952).

In this thesis, we discuss the various models from the portfolio-choice literatures that we consider. DeMiguel et at. (2009) has discussed that no general methods can consistently beat 1/N, so we will instead focus on explaining how to potentially outperform 1/N both theoretically and numerically by choosing smaller number of stocks. Following DeMiguel

et. at. (2009), we use R_t^m to denote the $m-$ vector of excess returns (over the risk-free assets) on m risky assets available for investment at date t. The number of stocks $m = N$ is called as the total number of stocks. The $m-$ dimensional vectors $\boldsymbol{\mu}_t^m$ is used to denote the expected returns on the risky asset in excess of the risk-free rate, and $\boldsymbol{\Sigma}_t^m$ is denoted as the corresponding $m \times m$ variance-covariance matrix of returns, with their sample counterparts given by $\hat{\mu}_t^m$ and $\hat{\Sigma}_t^m$, respectively. Let $\mathbf{1}_m$ denote an $m-$ dimensional vector of ones, $\boldsymbol{I}_m$ indicate the $m \times m$ identity matrix, and x_t^m be the vector of portfolio weights invested in the m risky assets, with $1 - \mathbf{1}_m^T x_t^m$ invested in the risk-free assets. The vector of relative weights in the portfolio with only-risky assets is

$$w_t^m = \frac{x_t^m}{|\mathbf{1}_m^T x_t^m|},$$

where the normalization by the absolute value of the sum of the portfolio weights, $|\mathbf{1}_m^T x_t^m|$, guarantees that the direction of the portfolio position is preserved in the few cases where the sum of the weights on the risky assets is negative.

In general, when $m = N$, we consider an investor whose preferences are fully described by the mean and variance of a chosen portfolio, x_t. At each time t, the decision-maker selects x_t to maximize expected utility:

$$\max_{x_t} x_t^T \mu_t - \frac{\gamma}{2} x_t^T \Sigma_t x_t, \tag{1.8}$$

where γ is interpreted as the investor's risk aversion. The solution of the above optimization is $x_t = (1/\gamma)\Sigma_t^{-1}\mu_t$.

We are trying to study under the framework of $m \ll N$. The fantastic part of our method is to show that given any arbitrary portfolio, we could construct a so called "effective portfolio" with Jensen's alpha > 0 and portfolio beta < 1 via dimension reduction. i.e. by choosing smaller number of stocks. We argue that each selected stock is treated as a factor, defined in a linear multi-factor model, and we show that only small portion of

stocks, denoted as the effective portfolio dimension (EPD), dominant the variance of the whole market. Our proposed portfolio which depends on a predetermined beta value is called effective portfolio (EP). Interestingly, we show that the EP yields a better portfolio alpha and beta tradeoff. Due to the estimation error in the practical situation, we show that our EP could be incorporated with mispricing factors, e.g. momentum and value. Then our EP achieves smaller variance, portfolio alpha > 0 and portfolio beta < 1 with small number of stocks.

The layout of this thesis is described as follows: Chapter 2 introduces dimension distribution and argues that the conditional tail dimension (CTD) is a better measure than the effective dimension. Chapter 3 designs our new QMC-friendly dimension reduction techniques for an arbitrary function, and Chapter 3 discusses the severity of sub-functions and the construction of PGM corresponding to different severities. Chapter 4 extends the dimension reduction technique on the real-time pricing of a large portfolio of insurance contacts. Then Chapter 5 proposes the effective portfolio to improve existing sophisticated strategies under dimension reduction framework.

1.1 Review of Path Generation Method (PGM)

The objective of this section is to provide a brief overview of the existing PGMs, including the standard way of generating the Brownian motions.

1.1.1 Forward or Standard (STD) Construction

The standard construction generates the Brownian motion sequentially as follow: given $B_0 = 0$,

$$B_{t_j} = B_{t_{j-1}} + \sqrt{t_j - t_{j-1}} z_j, z_j \sim N(0,1), j = 1, \ldots, n.$$

This method takes $O(n)$ operations to generate a path. The corresponding generating matrix $\boldsymbol{A}$ is the Cholesky decomposition of the covariance matrix $\boldsymbol{\Sigma}$, which takes the form

$$\boldsymbol{A} = \boldsymbol{A}^{STD} = \begin{pmatrix} \sqrt{t_1} & 0 & \cdots & 0 \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \sqrt{t_1} & \sqrt{t_2 - t_1} & \cdots & \sqrt{t_n - t_{n-1}} \end{pmatrix}.$$

1.1.2 Principal Component Analysis (PCA)

Acworth et al. (1998) proposed the PCA construction, which is based on an eigenvalue decomposition of the covariance matrix $\boldsymbol{\Sigma}$, such that $\boldsymbol{\Sigma} = \boldsymbol{V}^T \boldsymbol{\Lambda} \boldsymbol{V}$, where $\boldsymbol{V}$ is the matrix of its eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix of its eigenvalues. This method maximises the concentration of the total variance of the Brownian path in the first few dimensions. The path is obtained as follows

$$\boldsymbol{A} = \boldsymbol{A}^{\boldsymbol{PCV}} = \sqrt{\boldsymbol{\Lambda}} \boldsymbol{V} = \begin{pmatrix} \sqrt{\lambda_1} v_{11} & ... & \sqrt{\lambda_d} v_{1d} \\ \vdots & \ddots & \vdots \\ \sqrt{\lambda_1} v_{d1} & ... & \sqrt{\lambda_d} v_{dd} \end{pmatrix},$$

where v_{ij} denotes the j-th coordinate of the i-th eigenvector and λ_i denotes the i-th eigenvalue of the covariance matrix $\boldsymbol{\Sigma}$ satisfying $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$. Akesson and Lehoczky (1998) showed that for $i = 1, ..., d$, the i-th eigenvalue and eigenvector of the matrix $\boldsymbol{\Sigma}$ can be written as

$$v_{ij} = \frac{2}{\sqrt{2d+1}} \sin\left(\frac{2i-1}{2d+1} j\pi\right) \tag{1.9}$$

$$\lambda_k = \left(4d \sin^2\left(\frac{2k-1}{2d+1}\right)\frac{\pi}{2}\right)^{-1} \tag{1.10}$$

Since $\boldsymbol{A}$ is a full matrix, the PCA construction requires $O\left(d^2\right)$ operations for the generation of one path instead of $O(d)$ operations which are needed by forward construction. For large d, this often increases the run time of the simulation and limits the practical use of the PCA construction. It has been observed by Scheicher (2007) that a Brownian path using PCA can be computed using $O(d\log(d))$ floating point operations via certain discrete sine transformation. Also see (Leobacher 2011).

1.1.3 Brownian Bridge (BB)

The BB method, which was first proposed by Moskowitz and Caflisch (1996), simulates the Brownian motion by first generating the final value B_d, and then samples the intermediate values $B_{\lfloor d/2\rfloor}$ conditional on the values of B_d and B_0. After that, this method recursively fills the intermediate values on $\left(B_0, B_{\lfloor d/2\rfloor}\right)$ and $\left(B_{\lfloor d/2\rfloor}, B_d\right)$, where $\lfloor x\rfloor$ denotes the greatest integer less than or equal to x. In particular, if d is a power of 2, then BB generates the Brownian motion as follows:

$$\begin{aligned} B_d &= \sqrt{T}Z_1, \\ B_{d/2} &= \frac{1}{2}(B_0 + B_d) + \sqrt{T/4}Z_2, \\ B_{d/4} &= \frac{1}{2}\left(B_0 + B_{d/2}\right) + \sqrt{T/8}Z_3, \\ &\cdots \\ B_{d-1} &= \frac{1}{2}\left(B_{d-2} - B_d\right) + \sqrt{T/2d}Z_d, \end{aligned} \tag{1.11}$$

where Z_i are independent standard normal random variables for $i = 1, ..., d$. BB construction corresponds to a certain matrix $\boldsymbol{A_{BB}}$ such that $\boldsymbol{A_{BB}}\boldsymbol{A_{BB}}^T = \boldsymbol{\Sigma}$. For example, if $T = 4$ and $\Delta t = 1$, then the decomposition matrix corresponds to the BB method is

$$\boldsymbol{A_{BB}} = \begin{pmatrix} 1/2 & 1/2 & \sqrt{2}/2 & 0 \\ 1 & 1 & 0 & 0 \\ 3/2 & 1/2 & 0 & \sqrt{2}/2 \\ 2 & 0 & 0 & 0 \end{pmatrix}.$$

More generally, suppose we are interested in constructing a discrete Brownian path $(B_{t_1}, ..., B_{t_d})$ with covariance matrix $\mathbf{\Sigma}$.

Algorithm 1.1.1. *Suppose the elements of $(B_{t_1}, ..., B_{t_d})$ should be computed in the order of $B_{t_{\pi(1)}}$, $B_{t_{\pi(2)}}$, ... , $B_{t_{\pi(d)}}$ for some permutation π of d Brownian paths. Consequently, in computing $B_{t_{\pi(j)}}$, we need to take into account the previously computed elements. Fortunately at most two of those are of relevance, the one next to $\pi(j)$ on the left and the one next to $\pi(j)$ on the right. Now, for every $j \in \{1, ..., n\}$, $L(j) := \{k : k < \pi(j) \text{ and } \pi^{-1}(k) < j\}$, and $R(j) := \{k : k > \pi(j) \text{ and } \pi^{-1}(k) < j\}$. That is, L contains all the indices k that are smaller than $\pi(j)$ for which B_{t_k} has already been constructed and R contains all the indices k that are greater than $\pi(j)$ for which B_{t_k} has already been constructed. Then, if we let*

$$l(j) := \begin{cases} 0 & if \;\; L(j) = \phi \\ \max L(j) & if \;\; L(j) \neq \phi \end{cases},$$

$$r(j) := \begin{cases} \infty & if \;\; R(j) = \phi \\ \min R(j) & if \;\; R(j) \neq \phi \end{cases},$$

and set $B_{t_0} = 0$, we have

$$B_{t_{\pi(j)}} := \begin{cases} B_{t_{l(j)}} + \sqrt{t_{\pi(j)} - t_{l(j)}} Z_j & if \;\; r(j) = \infty \\ \frac{t_{r(j)} - t_{\pi(j)}}{t_{r(j)} - t_{l(j)}} B_{t_{l(j)}} + \frac{t_{\pi(j)} - t_{l(j)}}{t_{r(j)} - t_{l(j)}} B_{t_{r(j)}} + \sqrt{\frac{(t_{\pi(j)} - t_{l(j)})(t_{r(j)} - t_{\pi(j)})}{t_{r(j)} - t_{l(j)}}} Z_j & if \;\; r(j) < \infty \end{cases}.$$

It is straightforward to check that the vector $B_{t_1}, ..., B_{t_n}$ constructed in this way has again covariance matrix $\mathbf{\Sigma}$ with $\mathbf{\Sigma}_{ij} = (\min(t_j, t_k))$ for $i, j = 1, \cdots, d$. It is obvious that the

Brownian bridge construction uses $O(n)$ floating point operations. In particular, the classical BB construction are constructed in the order B_d, $B_{d/2}$, $B_{d/4}$, $B_{3d/4}$, $\cdots$.

The above description of the BB gives one implementation of generating the discrete Brownian motions. The optimal permutation of the Brownian bridge construction in the sense of explained variability is proved by Lin and Wang (2008) according to the following three theorems:

Theorem 1.1.1. *In a permutation-based construction of the Brownian motion $B_1, ..., B_d$, the optimal first step π_1 is the integer nearest to $(6d + 3)/8$.*

Theorem 1.1.2. *In a permutation-based construction of the Brownian motion $B_1, ..., B_d$, assume that B_q has been sampled (for some q with $0 \leq q \leq d - 1$), while $B_{q+1}, ..., B_d$ have not yet been generated. Among $B_{q+1}, ..., B_d$ conditional on the past value B_q, the local optimal new step q_{new} is the integer nearest to $(6d + 2q + 3)/8$.*

Theorem 1.1.3. *In a permutation-based construction of the Brownian motion $B_1, ..., B_d$, assume that B_{q_1} and B_{q_2} have been sampled (for some q_1, q_2 with $0 \leq q_1 < q_2 \leq d$), while $B_{q_1+1}, ..., B_{q_2-1}$ have not yet been generated. Among $B_{q_1+1}, ..., B_{q_2-1}$ conditional on the past value B_{q_1} and the future value B_{q_2}, the optimal new step is the integer nearest to $(q_1 + q_2)/2$.*

See Lin and Wang (2008) for the optimal permutations when $d = 2, 4, 8, 16, 32, 64, 128, 256$.

1.1.4 Orthogonal Transformation (OT)

Wang and Tan (2012) proposed an OT method for generating Brownian paths under Barrier options.

Theorem 1.1.4. *(Wang and Tan) Let $\boldsymbol{C}$ be a $d \times d$ positive definite matrix and let $\boldsymbol{A_0}$ be a fixed decomposition matrix such that $\boldsymbol{A_0}\boldsymbol{A_0}^T = \boldsymbol{C}$. Suppose that the indicator function $\Lambda(\boldsymbol{x})$ has the form*

$$\Lambda(\boldsymbol{x}) = I_{h(\boldsymbol{q}^T\boldsymbol{x})}(\boldsymbol{x}),\ \boldsymbol{x} = (x_1, ..., x_d)^T \sim N_d(\boldsymbol{0}, \boldsymbol{C}), \tag{1.12}$$

where $\boldsymbol{q} = (q_1, ..., q_d)^T$ *is a vector of constants and* $h(\cdot)$ *is a function defined in* R*. If* U *is a* $d \times d$ *orthogonal matrix, whose first column* $\boldsymbol{U_1}$ *is given by*

$$\boldsymbol{U_1} = \frac{1}{D}\boldsymbol{A_0}^T\boldsymbol{q}, \tag{1.13}$$

where $D := \sqrt{\boldsymbol{q}^T\boldsymbol{C}\boldsymbol{q}}$*, the remaining columns are arbitrary as long as they satisfy the orthogonality conditions, then by the transformation* $\boldsymbol{x} = \boldsymbol{A_0}\boldsymbol{U}\boldsymbol{z}$*, the function* $h(\boldsymbol{q}^T\boldsymbol{x})$ *involved in the indicator function* $\Lambda(\boldsymbol{x})$ *is transformed to a function depending only on the first component of* $\boldsymbol{z}$*:*

$$h(\boldsymbol{q}^T\boldsymbol{x}) = h(Dz_1).$$

Consequently, the indicator function $\Lambda(\boldsymbol{x})$ *is transformed to*

$$\Lambda(\boldsymbol{x}) = I_{h(Dz_1)<H}(\boldsymbol{z}),\ \ \boldsymbol{z} = (z_1, ..., z_d)^T \sim N_d(\boldsymbol{0}, (I)).$$

If $h(\cdot)$ *is strictly increasing on* R *and if* $I_{h(Dz_1)<H}(\boldsymbol{z})$ *is further transformed by the inverse normal transformation* $\boldsymbol{z} = \Phi^{-1}(\boldsymbol{u})$*, then the indicator function* $\Lambda(\boldsymbol{x})$ *is transformed to a one-dimensional function:*

$$\Lambda(\boldsymbol{x}) = I_{u_1<c}(\boldsymbol{u}),\ \ \boldsymbol{u} = (u_1, ..., u_d)^T \sim U(0,1)^d,$$

where $c = \Phi(D^{-1}h^{-1}(H))$ *is a constant. The discontinuities of the indicator function* $I_{u_1<c}(\boldsymbol{u})$ *are aligned with the coordinate axes.*

Proof. See Appendix *D*. □

Theorem 1.1.4 offers a new PGM for simulating Brownian motion. Wang and Tan (2012) referred to this new PGM as the orthogonal transformation (OT). The OT method is simple to implement by constructing an orthogonal matrix $\boldsymbol{U}$ according to Theorem 1.1.4 and taking the generating matrix to be $\boldsymbol{A} = \boldsymbol{A_0}\boldsymbol{U}$ for some fixed $\boldsymbol{A_0}$ satisfying $\boldsymbol{A_0}\boldsymbol{A_0}^T = \boldsymbol{C}$.

Whenever a function involves an indicator function of the form (1.12), Theorem 1.1.4 guarantees that the discontinuities are aligned with the coordinate axes. When a function involves an indicator function which is not exactly the form (1.12), but is "close"to this form in some sense, Theorem 1.1.4 is still useful in finding a good PGM as illustrated by Wang and Tan (2012). The value of the simple situation in Theorem 1.1.4 is to give insight of a function.

From a practical point of view, Wang and Tan (2012) mentioned two issues in Theorem 1.1.4.

1. Theorem 1.1.4 gives us only the first column of $\boldsymbol{U}$. Other columns of $\boldsymbol{U}$ are found by the Gram-Schmidt method or modified Gram-Schmidt algorithm (Wang and Sloan 2010). The determination of other columns also leaves room for further optimization of the generating matrix by taking into account the knowledge of the function.

2. Another issue is the choice of the initial decomposition matrix $\boldsymbol{A_0}$. If the underlying integrand is, say, a product of an indicator function of the form (1.12) with another function $G_0(\boldsymbol{x})$, i.e. $G(\boldsymbol{x}) = G_0(\boldsymbol{x})\Lambda(\boldsymbol{x})$, then the choice of $\boldsymbol{A_0}$ could have impact on the practical performance of QMC methods, since Theorem 1.1.4 only focuses on the indicator function $\Lambda(\boldsymbol{x})$. If there is an indication that the function $G_0(\boldsymbol{x})$ is PCA-friendly, then we may choose the initial decomposition matrix $\boldsymbol{A_0}$ to be $\boldsymbol{A_{PCA}}$. If no priori information is available, then $\boldsymbol{A_0}$ is taken to be the Cholesky decomposition of $\boldsymbol{C}$.

1.1.5 Linear Transformation

To derive the optimal orthogonal matrix $\boldsymbol{A}$ in general functions, Imai and Tan (2006) proposed to approximate the objective function by linearizing the function g and then maximizing the variance contribution according the first-order Taylor expansion to a gen-

eral function g at around an arbitrary point $\boldsymbol{\epsilon} = \hat{\boldsymbol{\epsilon}} + \Delta\boldsymbol{\epsilon}$:

$$g(\boldsymbol{\epsilon}) \approx g\left(\hat{\boldsymbol{\epsilon}}\right) + \sum_{k=1}^{d} \frac{\partial g}{\partial \epsilon_k}|_{\boldsymbol{\epsilon}=\hat{\boldsymbol{\epsilon}}} \Delta\epsilon_k, \tag{1.14}$$

and the approximated function is linear in the normal random variables $\Delta\boldsymbol{\epsilon}$ with

$$\left(\frac{\partial g}{\partial \epsilon_k}|_{\boldsymbol{\epsilon}=\hat{\boldsymbol{\epsilon}}}\right)^2 \tag{1.15}$$

to be the variability contributed by the k^{th} component. The optimization of $\boldsymbol{A}$ is given by

$$\max_{\boldsymbol{A_{.k}} \in R^d} \left(\frac{\partial g(\boldsymbol{A\epsilon})}{\partial \epsilon_k}|_{\boldsymbol{\epsilon}=\hat{\boldsymbol{\epsilon}_k}}\right)^2$$

subject to $\|\boldsymbol{A_{.k}} = 1\|$ and $\langle \boldsymbol{A_{.j}}, \boldsymbol{A_{.k}} \rangle = 0$, for $j = 1, ..., k-1$.

Furthermore, Imai and Tan (2009) extends the LT method into a more general way. By addressing the general problem of estimating $E(g(\boldsymbol{X}))$ where $\boldsymbol{X} = (X_1, \cdots, X_d)^T$ is a vector of d iid random variables with arbitrary probability density function (pdf) $f(x)$ and cumulative distribution function (cdf) $F(x)$. Note that the distribution of $\boldsymbol{x}$ may not be normally distributed. Then we have

$$E[g(\boldsymbol{X})] = \int_{\Omega} g(\boldsymbol{x}) f(x_1) \cdots f(x_d) dx_1 \cdots dx_d,$$

where Ω is the domain of $\boldsymbol{X}$. By substituting $y_i = F(x_i)$, $i = 1, \cdots, d$, the above integration reduces to the following problem:

$$E[g(\boldsymbol{X})] = \int_{[0,1]^d} g(F^{-1}(y_1), \cdots, F^{-1}(y_d)) dy_1 \cdots dy_d.$$

Now consider the transformation $Z = \Phi^{-1}(Y)$ where Φ represents the cdf of the standard

normal distribution. Then $E[g(\boldsymbol{X})]$ can be expressed as follows:

$$\int \cdots \int_{-\infty}^{\infty} g(F^{-1}(\Phi(z_1)), \cdots, F^{-1}(\Phi(z_d)))\phi(z_1)\cdots\phi(z_d)dz_1\cdots dz_d = g(F^{-1}(\Phi(Z_1)), \cdots, F^{-1}(\Phi(Z_d))),$$

where ϕ is the pdf of the standard normal, and $\boldsymbol{Z} = (Z_1, \cdots, Z_d)^T$ is a vector of independent standard normal random variable. This implies that another consistent estimator of $E[g(\boldsymbol{X})]$ can be obtained via $E[g(F^{-1}(\Phi(\boldsymbol{A}_{1.}\boldsymbol{Z})), \cdots, F^{-1}(\Phi(\boldsymbol{A}_{d.}\boldsymbol{Z})))]$, for any orthogonal matrix $\boldsymbol{A}$ where $\boldsymbol{A}_{j.}$ corresponds to the j-the row of $\boldsymbol{A}$. This is referred as the generalized LT method (GLT).

Chapter 2

Art of Dimension Distribution

Key contributions of this chapter

1. Propose a new measure of effective dimension known as the delta dimension.
2. Compared to the superposition and truncation dimensions, the delta dimension is simple and it is easy to compute.
3. Propose two other measures of quantifying effective dimensions known as the tail dimension and the conditional tail dimension.
4. Some justifications are provided to demonstrate the importance of the conditional tail dimension in determining the efficiency of QMC.

2.1 Introduction

There is ample numerical evidence attesting to the competitive advantage of quasi-Monte Carlo (QMC) methods in finance applications, even if these problems are of several hundred dimensions. While the success of the QMC methods in finance applications cannot be fully explained by the Koksma-Hlawka error bound, (see Niederreiter, 1992), it is widely believed that it is attributed to the notion of effective dimension, as argued by Caflisch et al. (1997). By exploiting the analysis of variance (ANOVA) decomposition of a function,

Caflisch et al. (1997) proposed two versions of effective dimensions known as the superposition dimension and truncation dimension. As we will review in Section 2.3.2, the downside of these measures is that it is computationally burdensome to compute them for an arbitrary function, hence hindering their usefulness. Motivated by this limitation, the key contribution of this chapter is to propose a new measure of effective dimension, which we denote as the *delta dimension*. The proposed delta dimension has the advantage of its simplicity and that it can easily be calculated for any arbitrary function. Furthermore, based on the dimension distribution of a dimension random variable, we create other measures known as the *tail dimension* and the *conditional tail dimension*. Some numerical evidences are provided to demonstrate that the efficiency of QMC is linked to the conditional tail dimension. The contributions of this chapter can be summarized as follows: (1) Propose a new measure of effective dimension known as the delta dimension. (2) Compared to the superposition and truncation dimensions, the delta dimension is simple and it is easy to compute. (3) Propose two other measures of quantifying effective dimensions known as the tail dimension and the conditional tail dimension. (4) Some justifications are provided to demonstrate the importance of the conditional tail dimension in determining the efficiency of QMC.

The rest of this chapter is organized as follows. Next section provides a brief overview of the sampling methods based on the Monte Carlo (MC) and Quasi-Monte Carlo (QMC) for estimating high-dimensional integrals. Section 2.3 then discusses the decomposition of a function, including the ANOVA decomposition. Based on the ANOVA decomposition, Section 2.4 formally defines the effective dimensions of Caflisch et al. (1997) as well as the dimension distributions of Owen (2003). Finally, Section 2.5 presents our proposed new measures of quantifying effective dimension. A numerical example is provided to conclude the chapter to highlight the advantage of our proposed measures.

2.2 Sampling Methods for Estimating High-Dimensional Integrals

High-dimensional integrals appear in various mathematical models in finance. In many cases, the arising integrals cannot be calculated analytically and hence are often resorted to numerical methods. The curse of dimensionality for the classical multivariate quadrature methods states that their efficiency decay exponentially with the dimension of the problem; hence their applicability to estimating high-dimensional problems in finance is very limited. In this section, we briefly review these methods as well as other sampling methods based on Monte Carlo and quasi-Monte Carlo methods.

2.2.1 Classical Multivariate Quadrature Methods

We start with the numerical methods for the computation of high-dimensional integrals

$$I(f) := \int_{[0,1]^d} f(\boldsymbol{x}) d\boldsymbol{x} \tag{2.1}$$

over the unit cube. Note that any rectangular integration domains $[a_1, b_1] \times ... \times [a_d, b_d]$ can be mapped to the unit cube by a simple linear transformation via

$$\int_a^b f(y)dy = (b-a) \int_0^1 f(a + (b-a)x)dx.$$

We also consider numerical methods for high-dimensional integrals of the form

$$I_\varphi(f) := \int_{R^d} f(\boldsymbol{x})\varphi_d(\boldsymbol{x}) d\boldsymbol{x}$$

over the d-dimensional Euclidean space with the Gaussian weight function φ_d.

All numerical quadrature methods approximate the d-dimensional integral $I(f)$ by a weighted sum of N function evaluations:

$$Q_N(f) := \sum_{i=1}^{N} w_i f(\boldsymbol{x}^{\boldsymbol{i}}), \tag{2.2}$$

with weights $w_i \in \mathbb{R}$ and nodes $\boldsymbol{x}^{\boldsymbol{i}} = (x_1^i, ..., x_d^i) \in \Omega^d$. Here, Ω is either $[0,1]$ or $\mathbb{R}$. Depending on the choice of the weights and nodes, different classes of methods with varying properties are obtained, for examples, QMC, MC, polynomial-based, and sparse grid methods. In this thesis, we are only interested in QMC and MC methods, with particular focus in the high-dimensional finance applications.

2.2.2 Crude Monte Carlo (MC) Methods

The MC technique is possibly the most popular way of estimating the high dimensional integral (2.1). It involves sampling points from a pseudo-random sequence so that the MC estimate that corresponds to the estimator (2.2) is an equally weighted rule of the following:

$$\hat{Q}_N^{MC}(f) := \frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{x}^{\boldsymbol{i}}). \tag{2.3}$$

where $\boldsymbol{x}^{\boldsymbol{i}} \in [0,1]^d$ and N denotes the number of random points used to estimate (2.1). The law of large numbers ensures that the above MC estimate $\hat{Q}_N^{MC}(f)$ converges to $I(f)$ as $N \to \infty$. If f is square-integrable and let σ^2 be the variance of f, then the MC estimate $\hat{Q}_N^{MC}(f)$ converges to $I(f)$ at a rate of $\sigma/\sqrt{N}$, which is independent of the dimension d. In practice, σ is typically estimated from the sample standard deviation as

$$\hat{\sigma}_N(f) = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} [f(\boldsymbol{x}^{\boldsymbol{i}}) - \hat{Q}_N^{MC}]^2}.$$

The Central Limit Theorem can then be used to construct the confidence interval of the MC estimate to gauge its efficiency.

While the MC method has some appealing features, including it is a flexible method with a wide range of application, and its convergence rate is independent of dimension, nevertheless this method is often criticized for its slow rate of convergence. Consequently ways of enhancing the efficiency of the crude MC method have been proposed. Among them the variance reduction techniques such as the control variate, importance sampling, etc. have been proposed. A comprehensive account of these methods can be found in standard textbook on MC methods. An excellent source of reference is Glasserman (2004). The quasi-Monte Carlo methods, which we will discuss in the next subsection, is another sampling-based method alternate to MC in an attempt to outperform the MC methods.

2.2.3 Quasi-Monte Carlo (QMC) Methods

As in the MC methods, the QMC methods are another type of equal-weight sampling methods for estimating high-dimensional integrals (2.1). A critical difference between MC and QMC methods lies in the properties of the points $\boldsymbol{x^i} \in [0,1]^d, i = 1, \ldots, N$ that are used to evaluate (2.1). The former method relies on a sequence of points that are randomly, uniformly, and independently generated from $[0,1]^d$. In contrast, the latter method uses a specially constructed sequence known as the low discrepancy sequence. The low discrepancy sequences are deterministic and are known to have greater uniformity than the traditional pseudo-random sequences. As we have mentioned in the last chapter, the irregularity or the uniformity of a set of points is measured by the discrepancy. Some popular low discrepancy sequences are attributed to Halton (1960), Faure (1982), and Sobol (1967).

It follows from the Koksma-Hlawka inequality that the worst case error bound of QMC is of order $O\left(N^{-1}(\log N)^d\right)$ for integrands of bounded variation. This rate is asymptotically better than the probabilistic $O(N^{-1/2})$ error bound of MC. Unlike the MC methods, the convergence rate of QMC depends on the dimension d, as reflected in the factor $(\log N)^d$.

In practical applications, QMC methods involving randomized low discrepancy sequences,

as opposed to the traditional deterministic low discrepancy sequences, are typically implemented. As we have discussed briefly that the traditional deterministic low discrepancy sequences can be randomized via methods such as random shifts or random permutations of digits. The advantage of using the randomized QMC permits an easy way of quantifying the precision of the QMC estimate. See for example Tan and Boyle (2000), Ökten and Eastman (2004) and Xu and Ökten (2015). For a survey on the randomized QMC, see Lecuyer and Lemieux (2002).

2.3 Dimension-wise Decomposition

In this section we briefly review two possible ways of dimension-wise decomposition. The first method is proposed by Griebel and Holtz (2010) and the second method is based on the ANOVA decomposition. The latter decomposition has been exploited in quantifying the effective dimensions of an arbitrary functrion.

2.3.1 Griebel and Holtz (2010)

Griebel and Holtz (2010) provide a general dimension-wise decomposition of an arbitrary function. Let $\Omega \subseteq \mathbb{R}$ be a set and let

$$d\mu(\boldsymbol{x}) = \prod_{j=1}^{d} d\mu_j(x_j) \tag{2.4}$$

denote a d-dimensional product measure defined on Borel subsets of Ω^d. Here, $\boldsymbol{x} = (x_1, ..., x_d)^T$ and μ_j for $j = 1, ..., d$ are probability measures on Borel subsets of Ω. Let $V^{(d)}$ be the Hilbert space of all square integrable functions $f : \Omega^d \to \mathbb{R}$ with the inner product

$$\langle f, g \rangle := \int_{\Omega^d} f(\boldsymbol{x}) g(\boldsymbol{x}) d\mu(\boldsymbol{x}).$$

For a given set $u \subseteq \mathcal{D}$, where $\mathcal{D} := \{1, ..., d\}$ denotes the set of coordinate indices and $|u|$ denotes the cardinality of u, the projections $P_u : V^{(d)} \to V^{(|u|)}$ can be defined as

$$P_u f(\boldsymbol{x_u}) := \int_{\Omega^{d-|u|}} f(\boldsymbol{x}) d\mu_{D\setminus u}(\boldsymbol{x}),$$

where $\boldsymbol{x_u}$ denotes the $|u|-$dimensional vector containing those components of x whose indices belong to the set u, and $\mathcal{D}\setminus u$ denotes its complementary set in $\mathcal{D}$ so that $d\mu_{D\setminus u}(\boldsymbol{x}) := \prod_{j\notin u} d\mu_j(x_j)$. In the special case with $\boldsymbol{u} = \emptyset$, we have

$$P_\emptyset f(\boldsymbol{x_\emptyset}) := \int_{\Omega^d} f(\boldsymbol{x}) d\mu(\boldsymbol{x}) =: I(f).$$

The projections define a dimension-wise decomposition of $f \in V^{(d)}$ into a finite sum according to

$$f(x_1, ..., x_d) = f_\emptyset + \sum_{i=1}^{d} f_i(x_i) + \sum_{i,j=1, i<j}^{d} f_{i,j}(x_i, x_j) + ... + f_{1,...,d}(x_1, ..., x_d),$$

which is often written in a more compact notation as follows:

$$f(\boldsymbol{x}) = \sum_{u\subseteq\mathcal{D}} f_u(\boldsymbol{x_u}).$$

The 2^d terms f_u describe the dependence of the function f on the dimensions $j \in u$ with respect to the measure μ. They could be defined recursively by

$$f_u(\boldsymbol{x_u}) := P_u f(\boldsymbol{x_u}) - \sum_{v\subset u} f_v(\boldsymbol{x_v}), \tag{2.5}$$

or equivalently given by

$$f_u(\boldsymbol{x_u}) = \sum_{v\subseteq u} (-1)^{|u|-|v|} P_v f(\boldsymbol{x_v}).$$

The resulting decomposition is unique for a fixed measure μ and orthogonal in the sense that

$$\langle f_u, f_v \rangle = 0 \qquad \text{for } u \neq v.$$

2.3.2 ANOVA Decomposition

ANOVA decomposition is a way of decomposing a function into a sum of orthogonal functions. By assuming $f(\boldsymbol{x})$ is a square integrable function and denoting $[0,1]^{|u|}$ as the $|u|$-dimensional unit cube involving the coordinates in u, we can write $f(\boldsymbol{x})$ as the sum of its 2^d ANOVA terms:

$$f(\boldsymbol{x}) = \sum_{u \subseteq \mathcal{D}} f_u(\boldsymbol{x}),$$

i.e.

$$f(\boldsymbol{x}) = f_\emptyset + \sum_{i=1}^{d} f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) + \cdots + f_{1,2,\cdots,d}(x_1, \cdots, x_d). \tag{2.6}$$

The ANOVA terms $f_u(\boldsymbol{x})$ are defined recursively by

$$f_u(\boldsymbol{x}) = \int_{[0,1]^{d-|u|}} f(\boldsymbol{x}) d\boldsymbol{x}_{D \setminus u} - \sum_{v \subset u} f_v(\boldsymbol{x})$$

with $f_\emptyset = \int_{[0,1]^d} f(\boldsymbol{x}) dx = I(f)$. The ANOVA term $f_u(\boldsymbol{x})$ is the part of the function depending only on the variables x_j with $j \in u$. The ANOVA terms enjoy some interesting properties:

1. $\int_0^1 f_u(\boldsymbol{x}) dx_j = 0$ *for* $j \in u$.

2. The ANOVA decomposition is orthogonal whenever $u \neq v$. i.e. $\int_{[0,1]^d} f_u(\boldsymbol{x}) f_v(\boldsymbol{x}) d\boldsymbol{x} = 0$.

3. By letting $\sigma^2(f) := \int_{[0,1]^d} (f(\boldsymbol{x}) - I(f))^2 dx = \int_{[0,1]^d} f^2(\boldsymbol{x}) d\boldsymbol{x} - [I(f)]^2$ be the variance of the function f, then the orthogonality implies that

$$\sigma^2(f) = \sum_{u \subseteq \mathcal{D}} \sigma_u^2(f),$$

where $\sigma_u^2(f) = \int_{[0,1]^d} [f_u(\boldsymbol{x})]^2 d\boldsymbol{x}$ for $|u| > 0$ is the variance of f_u, and $\sigma_\emptyset^2(f) = 0$.

In Sobol' and Kucherenko (2005), the ratio $\sigma_u^2(f)/\sigma^2(f)$ is known as the global sensitivity index and it has been advocated to measure the relative importance of the term f_u with respect to the function f.

2.3.3 Special Case I

Griebel (2006) and Griebel and Holtz (2010) argue that ANOVA decomposition is closely related to the multivariate Taylor expansion. Consider an infinitely differentiable, real-valued function $f(\boldsymbol{x})$ that depends on d independent variables $\boldsymbol{x} = (x_1, \cdots, x_d)^T \in \mathbb{R}^d$ or $[0,1]^d$. The Taylor expansion of $f(\boldsymbol{x})$ can be expressed by

$$\begin{aligned} f(\boldsymbol{x}) &= f(\boldsymbol{\mu}) + \sum_{j=1}^{\infty} \frac{1}{j!} \sum_{i=1}^{d} \frac{\partial^j f(\boldsymbol{\mu})}{\partial x_i^j} (x_i - \mu_i)^j \\ &+ \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \frac{1}{j_1! j_2!} \sum_{i_1 < i_2} \frac{\partial^{j_1+j_2} f(\boldsymbol{\mu})}{\partial x_{i_1}^{j_1} \partial x_{i_2}^{j_2}} (x_{i_1} - \mu_{i_1})^{j_1} (x_{i_2} - \mu_{i_2})^{j_2} \\ &+ \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} \sum_{j_3=1}^{\infty} \frac{1}{j_1! j_2! j_3!} \sum_{i_1 < i_2 < i_3} \frac{\partial^{j_1+j_2+j_3} f(\boldsymbol{\mu})}{\partial x_{i_1}^{j_1} \partial x_{i_2}^{j_2} \partial x_{i_3}^{j_3}} (x_{i_1} - \mu_{i_1})^{j_1} (x_{i_2} - \mu_{i_2})^{j_2} (x_{i_3} - \mu_{i_3})^{j_3} \\ &+ \cdots \\ &+ \sum_{j_1=1}^{\infty} \cdots \sum_{j_d=1}^{\infty} \frac{1}{j_1! j_2! \cdots j_d!} \sum_{1<2<\cdots<d} \frac{\partial^{j_1+j_2+\cdots+j_d} f(\boldsymbol{\mu})}{\partial x_1^{j_1} \cdots \partial x_d^{j_d}} (x_1 - \mu_1)^{j_1} \cdots (x_d - \mu_d)^{j_d}, \end{aligned} \tag{2.7}$$

where $\mu_i = \mathbb{E}(x_i)$ and $\frac{\partial^j f(\boldsymbol{\mu})}{\partial x_i^j}$ means the j-th derivative with respect to x_i. It is easy to see that the first additive term in (2.7) is related to f defined in the ANOVA decomposition

(2.6), the second additive term in (2.7) is related to $\sum_{i=1}^{d} f_i(x_i)$ defined in the ANOVA decomposition (2.6), $\cdots$, and the last additive term in the formula (2.7) is related to $f_{1,2,\cdots,d}(x_1, \cdots, x_d)$ defined in the ANOVA decomposition (2.6).

2.3.4 Special Case II

Lemma 2.3.1. *If a function f has only the first order partial derivative with respect to any x_i in (2.7), for $i = 1, \cdots, d$, then the ANOVA decomposition of $f(\boldsymbol{x})$, which depends on d independent variables $\boldsymbol{x}$, is in the form*

$$\begin{aligned} f_{\emptyset} &= f(\boldsymbol{\mu}) \\ f_{\{i\}} &= \frac{\partial f}{\partial x_i}(\boldsymbol{\mu})(x_i - \mu_i) \\ f_{\{i,j\}} &= \frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{\mu})(x_i - \mu_i)(x_j - \mu_j) \\ &\cdots \\ f_{\{1,2,\cdots,d\}} &= \frac{\partial^d f(\boldsymbol{\mu})}{\partial x_1 \cdots \partial x_d} \prod_{i=1}^{d} (x_i - \mu_i) \end{aligned}$$

Proof: The Taylor expansion of f with only the first order partial derivative with respect to any x_i can be written as

$$\begin{aligned} f(\boldsymbol{x}) &= f(\boldsymbol{\mu}) + \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(\boldsymbol{\mu})(x_i - \mu_i) \\ &+ \sum_{i_1 < i_2} \frac{\partial^2 f}{\partial x_{i_1} \partial x_{i_2}}(\boldsymbol{\mu})(x_{i_1} - \mu_{i_1})(x_{i_2} - \mu_{i_2}) \\ &+ \sum_{i_1 < i_2 < i_3} \frac{\partial^3 f(\boldsymbol{\mu})}{\partial x_{i_1} \partial x_{i_2} \partial x_{i_3}}(x_{i_1} - \mu_{i_1})(x_{i_2} - \mu_{i_2})(x_{i_3} - \mu_{i_3}) \\ &+ \cdots \\ &+ \frac{\partial^d f(\boldsymbol{\mu})}{\partial x_1 \cdots \partial x_d} \prod_{i=1}^{d} (x_i - \mu_i). \end{aligned}$$

Since we have $\int (x_i - \mu_i) dx_i = 0$, for any $i = 1, \cdots, d$, assume that we have independent marginal distributions of x_i, it is obvious to see that each additive term in the Taylor expansion is orthogonal to others, and the Taylor expansion is in the sum of all ANOVA decomposition terms. i.e. $f(\boldsymbol{x}) = f_{\emptyset} + \sum_i f_{\{i\}} + \cdots + f_{\{1,2,\cdots,d\}}$. □

To conclude this section, we remark that in general it is computationally burdensome to compute the terms in the ANOVA decomposition.

2.4 Truncation and Superposition Dimensions

By exploiting the properties of ANOVA decomposition, Caflisch et al. (1997) provide two formal definitions of effective dimensions, known as the truncation dimension and the superposition dimension.

Definition 2.4.1. *The effective dimension of f in the truncation sense (or truncation dimension) is the smallest integer d_T such that*

$$\sum_{u \subseteq \{1,\ldots,d_T\}, u \neq \emptyset} \sigma_u^2(f) \geq \alpha \sigma^2(f), \tag{2.8}$$

where $\alpha \in (0, 1)$.

In the above definition, the parameter α is arbitrary but it is often set to some values close to one, such as 0.99. Roughly speaking, the truncation dimension represents that the first d_T variables are "important" variables of the function f. For large d, it is no longer possible to compute all 2^d ANOVA terms, but the truncation dimension can be computed recursively by

$$D_u(f) = \int_{[0,1]^{2d-|u|}} f(\boldsymbol{x}) f(\boldsymbol{x}_u, \boldsymbol{y}_{\mathcal{D}\setminus u}) d\boldsymbol{x} d\boldsymbol{y}_{\mathcal{D}\setminus u} - (I(f))^2, \tag{2.9}$$

where $D_u(f)$ is computed by $(2d - |u|)$-dimensional integral with $\boldsymbol{x} = (\boldsymbol{x}_u, \boldsymbol{x}_{\mathcal{D}\setminus u})$ and $\boldsymbol{y} = (\boldsymbol{y}_u, \boldsymbol{y}_{\mathcal{D}\setminus u})$. In practice the estimation of the truncation dimension is accomplished by

estimating the integrals via MC methods. This requires computing at most $d+1$ integrals with each integral of dimension up to $2d-1$.

The algorithm of computing the truncation dimension proposed by Barth et al. (2011) is given as follow:

Algorithm 2.4.1. *(truncation dimension):*

Step 1: Compute $I(f)$ *and* $\sigma^2(f)$

Step 2: For $i = 1, , , d$

Compute $D_{1,\ldots,i}(f)$ *by* (2.9).

If $D_{1,\ldots,i}(f) > \alpha\sigma^2(f)$, *then*

return i

End If

End For

Definition 2.4.2. *For* $\alpha \in (0,1)$, *the effective dimension of* f *in the superposition sense (or superposition dimension) is the smallest integer* d_S *such that*

$$\sum_{|u|\leq d_S, u\neq\emptyset} \sigma_u^2(f) \geq \alpha\sigma^2(f). \tag{2.10}$$

The superposition dimension roughly represents the highest order of important interactions between variables. The algorithm of computing the superposition dimension proposed by Barth et al. (2011) is given as follow:

Algorithm 2.4.2. *(superposition dimension):*

Step 1: Compute $I(f)$ *and* $\sigma^2(f)$, $D\left(f^{tot}\right) = 0$, $\sigma_{tot}^2 = 0$

Step 2: For $i = 1, \ldots, d$

Compute $\sigma_i^2(f) = D_i(f)$ *by* (2.9).

$\sigma_{tot}^2 = \sigma_{tot}^2 + \sigma_i^2(f)$

End For

Step 3: If $\sigma_{tot}^2 > \alpha\sigma^2(f)$

return $d_S = 1$

End If

Step 4: *For* $i = 1, ..., d$

For $j = i + 1, ..., d$

Compute $D_{i,j}(f)$ *by* (2.9).

$\sigma^2_{i,j}(f)$ = $D_{i,j}(f)$ - $\sigma^2(f_i)$ - $\sigma^2(f_j)$

$\sigma^2_{tot} = \sigma^2_{tot} + \sigma^2(f_{i,j})$

End For

End For

Step 5: *If* $\sigma^2_{tot} > \alpha\sigma^2(f)$

return $d_S = 2$

Else

return $d_S \geq 3$

End If

We remark that computing the superposition dimension is computational burdensome. It requires to compute $\sigma^2_u(f)$, for $u \in \mathcal{D}$, calculated from (2.9). This implies that if $d_S = 1$, we need to compute $\binom{d}{1}$ values; if $d_S = 2$, we need to compute $\binom{d}{1} + \binom{d}{2}$ values, and etc. Hence for large d and $d_S \geq 3$, it can be computationally intensive. This also explains why the algorithm provided above only calculates the variance contributions of the order-1 and order-2 terms of the ANOVA decomposition. Hence the algorithm works only if $d_S \leq 2$.

For each notion of effective dimension, it is also possible to define its corresponding dimension distribution, as discussed in Owen (2003). More specifically, assuming that the dimension of any arbitrary function f is a random variable, Owen (2003) gives the following two definitions of dimension distributions of f as follows:

Definition 2.4.3. *Let* D_S *denote the dimension random variable under superposition sense. The dimension distribution of* f *(in the superposition sense) is the probability dis-*

tribution of $|U|$ *where* $Pr(U = u) = \sigma_u^2/\sigma^2$. *It has probability mass function*

$$Pr(D_S = i) = \sum_{|u|=i} \frac{\sigma_u^2}{\sigma^2},$$

for $i = 1, ..., d$.

Definition 2.4.4. *Let* D_T *denote the dimension random variable under truncation sense. The dimension distribution of* f *(in the truncation sense) is the probability distribution of* $\max\{j|j \in U\}$ *where* $Pr(U = u) = \sigma_u^2/\sigma^2$. *It has probability mass function*

$$Pr(D_T = i) = \sum_{\max\{j|j\in u\}=i} \frac{\sigma_u^2}{\sigma^2},$$

for $i = 1, \cdots, d$.

Remark 2.4.1. *Since the set* $\{u : \max\{j|j \in u\} = i\} \subseteq \{u : |u| = i\}$, $Pr(D_T \leq i) \leq Pr(D_S \leq i)$, *for* $i = 1, \cdots, d$.

Armed with the probability mass functions of the dimension random variables, it is possible to obtain a number of other interesting information. For example, their cumulative mass functions (CMFs) are easily derived. We will use $F_{D_S}(i)$ and $F_{D_T}(i)$ to denote the resulting cumulative mass function for D_S and D_T, respectively. The mean dimensions corresponding to the expectations of $|U|$ and $max_{j\in U}$ can similarly be calculated, as defined as follows: The mean dimension of f in the superposition sense is defined as

$$m_S = \frac{\sum_{|u|>0} \sigma_u^2 |u|}{\sum_{|u|>0} \sigma_u^2},$$

and the mean dimension of f in the truncation sense is defined as

$$m_T = \frac{\sum_{|u|>0} \sigma_u^2 \max\{j|j \in u\}}{\sum_{|u|>0} \sigma_u^2}.$$

2.5 Delta Effective Dimension

The objective of this section is to propose some other plausible measures of effective dimension. The key contribution of this chapter is also collected in this section. Our proposed delta dimension is formally defined based on $f(\boldsymbol{x})$ and $\boldsymbol{A}$ where $\boldsymbol{x} \sim N_d(\boldsymbol{0}, \boldsymbol{\Sigma})$ and assuming that $\boldsymbol{A}$ is a decomposed matrix of $\boldsymbol{\Sigma}$; i.e. $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{\Sigma}$. The proposed delta dimension is defined as follows.

Definition 2.5.1. *(Delta Dimension) For $\alpha \in (0,1)$, consider a function $f(\boldsymbol{x}) = f(\boldsymbol{A}\boldsymbol{z})$ where $\boldsymbol{z}$ is a vector of d dimensional standard normal random variables, then the delta dimension of f, denoted as d_D, is the smallest integer k such that*

$$\begin{aligned} d_D &= \min\{k : Var(f(\hat{x}_1, \cdots, \hat{x}_k)) \geq \alpha Var(f)\} \\ &= \min\{k : Var(f(\boldsymbol{A}_{\cdot 1:k}\boldsymbol{z}_{1:k})) \geq \alpha Var(f))\}, \end{aligned}$$

where $(\hat{x}_1, \cdots, \hat{x}_k) = \boldsymbol{A}_{\cdot 1:k}\boldsymbol{z}_{\boldsymbol{1:k}}$, $(\boldsymbol{A})_{\cdot 1:k}$ is the first k columns of $\boldsymbol{A}$, and $\boldsymbol{z}_{1:k}$ is the first k elements of $\boldsymbol{z}$ for $k = 1, 2, ..., d$. Note: the ordering in this definition is important as this will have a direct impact on the "tail dimension", which will be discussed later.

The term $Var(f(\boldsymbol{A}_{\cdot 1:k}\boldsymbol{z}_{1:k}))$ corresponds to the variance as captured by the first k components of $\boldsymbol{x}$. Note that since $f(\boldsymbol{A}_{\cdot 1:k}\boldsymbol{z}_{1:k})$ may not be orthogonal to $f - f(\boldsymbol{A}_{\cdot 1:k}\boldsymbol{z}_{1:k})$, this means $f(\boldsymbol{A}_{\cdot 1:k}\boldsymbol{z}_{1:k})$ may not be the ANOVA decomposition of f and hence d_D will be different from d_T and d_S. **However, if f follows the definition in Special Case II, then $f(\boldsymbol{A}_{\cdot 1:k}\boldsymbol{z}_{1:k})$ and $f - f(\boldsymbol{A}_{\cdot 1:k}\boldsymbol{z}_{1:k})$ are orthogonal. Hence, the correlation between $f(\boldsymbol{A}_{\cdot 1:k}\boldsymbol{z}_{1:k})$ and $f - f(\boldsymbol{A}_{\cdot 1:k}\boldsymbol{z}_{1:k})$ are expected to be small.**

There are two advantages associated with our proposed definitions of delta dimension.

1. The first advantage is that the delta dimension ties directly to the PGM $\boldsymbol{A}$. As pointed out in Chapter 1 that the efficiency of the QMC is intricately related to the choice of PGM; i.e. $\boldsymbol{A}$. Hence the delta dimension d_D tells us how many columns of $\boldsymbol{A}$ (i.e. the number of components of $\boldsymbol{x}$) we need in order to achieve significant

contribution of the total variance. Hence if d_D is low relative to d, then the function f under the PGM $\boldsymbol{A}$ is said to have low effective dimension in that the first d_D components of $\boldsymbol{x}$ are sufficient to capture at least α proportion of the total variance.

2. The second advantage lies in its simplicity and its ease of calculating d_D. This is especially important as this feature allows us to integrate the delta dimension with the design of PGM.

Note that the delta dimension is defined in term of a function that depends on a set of normal random variables. For a general function $f(\boldsymbol{x})$ where $\boldsymbol{x}$ is a d-dimensional random vector that needs not be normally distributed, the delta dimension of f is then defined after applying the delta transformation. Hence the reason for naming the proposed measure as the delta dimension. To see this more explicitly, let us suppose that we can transform f to a function $\psi\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_k(\boldsymbol{x})\}$. We then let $Y = \{f_1(\boldsymbol{x}),\ f_2(\boldsymbol{x})\ ,\ ...\ ,\ f_k(\boldsymbol{x})\}$. By applying the first-order vector Taylor expansion to each of the function f_i with respect to $\boldsymbol{x}$ around an arbitrary vector $\hat{\boldsymbol{x}}$ for $i = 1, ..., k$, we have

$$\boldsymbol{Y} \approx \boldsymbol{Y}\left(\hat{\boldsymbol{x}}\right) + \nabla \boldsymbol{Y}\left(\hat{\boldsymbol{x}}\right)\left(\boldsymbol{x} - \hat{\boldsymbol{x}}\right). \tag{2.11}$$

Then, applying the Delta method (see Oehlert, 1992), the covariance matrix of $\boldsymbol{Y}$, denoted as $\boldsymbol{\Sigma}_f$, is defined as follows:

$$Var(\boldsymbol{Y}) = \boldsymbol{\Sigma}_f = \boldsymbol{J}\boldsymbol{\Sigma}\boldsymbol{J}^T \tag{2.12}$$

where the linear approximation of $\boldsymbol{J}$ is the Jacobian matrix of $\boldsymbol{Y}$ at $\boldsymbol{x}$; i.e.

$$\boldsymbol{J} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_d} \\ \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_2}{\partial x_d} \\ \cdots & \cdots & \cdots \\ \frac{\partial f_k}{\partial x_1} & \cdots & \frac{\partial f_k}{\partial x_d} \end{pmatrix}.$$

Then, according to (2.11) and the Delta method, we can construct the effective dimension

of the approximated function $\psi(\boldsymbol{A}_f \boldsymbol{z})$, where

$$f = \psi(f_1, \cdots, f_k) \approx \psi(\boldsymbol{A}_f \boldsymbol{z}),$$

$\boldsymbol{A}_f \boldsymbol{A}_f^T = \boldsymbol{\Sigma}_f$, and $\boldsymbol{z}$ is a vector of standard normal random variables.

Corresponding to the delta dimension defined in (2.5.1), we denote D_D as its random variable. The cumulative mass function of D_D can be defined as follows:

$$\Pr(D_D \leq i) := F_{D_D}(i) = \frac{Var(f(z_1, \cdots, z_i))}{Var(f(z_1, \cdots, z_d))} = \frac{Var(f(\boldsymbol{A}_{\cdot 1:i} \boldsymbol{z_{1:i}})}{Var(f(\boldsymbol{A}\boldsymbol{z}))},$$

for $i \in \{1, \cdots, d\}$.

2.5.1 Other Dimension-Based Distributional Quantities

In this subsection, we introduce two additional measures of effective dimension known as the tail dimension and the conditional tail dimension. These two measures apply to any given dimension distribution. Let D denote a discrete dimension random variable admitting values of $\{1, 2, \ldots, d\}$ and F_D denote its cumulative mass function. Note that D can be one of the D_S, D_T, or D_D discussed above. The tail dimension is then defined as follows:

Definition 2.5.2. *(Tail Dimension) The tail dimension of the dimension random variable D at the 100α % level, denoted by $Tail_\alpha(D)$, is defined as*

$$Tail_\alpha(D) = \inf\{k \in \{1, ..., d\} : F_D(k) \geq \alpha\}. \tag{2.13}$$

Note that if D is D_S, D_T, or D_S, then the $Tail_\alpha(D)$ recovers their respective effective dimensions d_S, d_S, or d_D.

Associated with $Tail_\alpha(D)$, we define the conditional tail dimension as follows:

Definition 2.5.3. *(Conditional Tail Dimension) The conditional tail dimension of D at the 100α % level, denoted by $CTD_\alpha(D)$, is defined as*

$$CTD_\alpha(D) = \frac{1}{1-\alpha}\int_\alpha^1 Tail_s(D)ds. \tag{2.14}$$

Intuitively $CTD_\alpha(D)$ measures the average dimension conditional on being greater than the effective dimension. Note that in the context of risk management with D interpreted as the loss random variable, then both (2.13) and (2.14) become the popular risk measures known as the Value at Risk and the Conditional Tail Expectation, respectively.

As the dimension random variable is a discrete random variable, the equivalent definition of (2.14) can be defined as follows:

$$CTD_\alpha(D) = \frac{(\beta^* - \alpha)Tail_\alpha(D) + (1-\beta^*)\mathbb{E}(D|D > Tail_\alpha(D))}{1-\alpha}, \tag{2.15}$$

where $\beta^* = \max\{\beta : Tail_\beta(D) = Tail_\alpha(D)\}$.

We remark that two functions may have the same variance and the same effective dimension, yet their conditional tail dimension can be very different. We argue that while the effective dimension provides some measure of efficiency of QMC, the conditional tail dimension can be another useful indicator. To see this, let us consider an arbitrary function f and assume the following decomposition applies:

$$f = f_{D\leq Tail} + f_{D>Tail},$$

where $Tail = Tail_\alpha(D)$. In the above decomposition, $f_{D\leq Tail}$ is a function that captures up to $Tail$ components of the function while $f_{D>Tail}$ depends on components that are greater than $Tail$ components. Both functions $f_{D\leq Tail}$ and $f_{D>Tail}$ do not need to be orthogonal.

To continue, it is useful to introduce the concept of QMC-friendly, as defined in Wang and Sloan (2011) and Wang and Tan (2013). Let $Var^{MC}(g)$ and $Var^{QMC}(g)$ denote the

variance of the estimator of g using the method of MC and QMC, respectively.

Definition 2.5.4. *(QMC-friendly) A function g is called "QMC-friendly" if the variance reduction ratio (VRR) of g defined as*

$$VRR(g) = \frac{Var^{MC}(g)}{Var^{QMC}(g)}$$

is sufficiently large; i.e., $VRR(g) \gg 1$. Note: in QMC method, we apply low-discrepancy sequences instead of random sequences to achieve this.

In other words, if g is QMC-friendly, then relative to the MC method, QMC is an extremely effective method for estimating g.

By denoting $VRR_{D\leq k}$ as $VRR(f_{D\leq k})$ and $VRR_{D>k}$ as $VRR(f_{D>k})$, we have the following result.

Theorem 2.5.1. *If $f_{D\leq Tail}$ is QMC-friendly, i.e., $VRR(f_{D\leq Tail})$ is sufficiently large, then the QMC variance of f, i.e. the variance of the estimate f using QMC sequences, is dominated by $f_{D>Tail}$.*

Proof:

$$\begin{aligned}
Var^{QMC}(f) &= Var^{QMC}(f_{D\leq Tail}) + Var^{QMC}(f_{D>Tail}) \\
&\quad + 2\rho^{QMC}(f_{D\leq Tail}, f_{D>Tail})\sqrt{Var^{QMC}(f_{D\leq Tail}) \times Var^{QMC}(f_{D>Tail})} \\
&= \frac{Var^{MC}(f_{D\leq Tail})}{VRR_{D\leq Tail}} + \frac{Var^{MC}(f_{D>Tail})}{VRR_{D>Tail}} \\
&\quad + 2\rho^{QMC}(f_{D\leq Tail}, f_{D>Tail})\sqrt{\frac{Var^{MC}(f_{D\leq Tail})}{VRR_{D\leq Tail}} \times \frac{Var^{MC}(f_{D>Tail})}{VRR_{D>Tail}}} \\
&\to \frac{Var^{MC}(f_{D>Tail})}{VRR_{D>Tail}} \\
&= Var^{QMC}(f_{D>Tail}),
\end{aligned}$$

where $\rho^{QMC}(f_{D\leq Tail}, f_{D>Tail})$ denotes the correlation between the QMC estimators of $f_{D\leq Tail}$ and $f_{D>Tail}$. Note that in deriving the above results, we assume that $VRR(f_{D\leq Tail})$

is sufficiently large (due to the assumption of QMC-friendly) so that $\frac{Var^{MC}(f_{D\leq Tail})}{VRR_{D\leq Tail}} \to 0$. Hence, the QMC variance of f is dominated by $f_{D\geq Tail}$. □

The above theorem suggests that the effect as attributed to the $f_{D>Tail}$ cannot be neglected. In fact, the following example suggests that its conditional tail dimension can be a useful indicator for determining efficiency of QMC.

Example 2.5.1. *(The importance of CTD) In this example we are interested in the role of effective dimension and the conditional tail dimension on the efficiency of QMC when estimating $f_i, i = 0, \ldots, 8$. We use f_0 as the benchmark. The functions $f_i, i = 0, \ldots, 8$, are shown in Table 2.1, assuming $z_i \in U(0, 1)$, for all i. These functions are chosen so that their mean, their effective dimensions (both superposition and truncation) are exactly the same. In particular, $E(f_i) = 0$ and $Var(f_i) = 1$, for $i = 0, \ldots, 8$. For each function, we compute its variance, effective dimension, $VRR^{QMC}(f)/VRR^{QMC}(f_i)$, and $CTD_\alpha(D)$.*

Table 2.1: Impact of effective dimension, conditional tail dimension on the efficiency of QMC

i	f_i	Variance	d_S or d_T	$\frac{Var^{QMC}(f_0)}{Var^{QMC}(f_i)}$	$CTD_\alpha(D)$
0	$f_0 = \sqrt{0.9}f_1 + \sqrt{0.099}f_2 + \sqrt{0.001}f_4$	1	4	$= 1$	4.24
1	$f_1 = \sqrt{12}(z_1 - \frac{1}{2})$	1	1	> 1	1
2	$f_2 = \sqrt{12^4}\prod_{i=1}^{4}(z_i - \frac{1}{2})$	1	4	< 1	4
3	$f_3 = \sqrt{12^8}\prod_{i=1}^{8}(z_i - \frac{1}{2})$	1	8	< 1	8
4	$f_4 = \sqrt{12^{16}}\prod_{i=1}^{16}(z_i - \frac{1}{2})$	1	16	< 1	16
5	$f_5 = \sqrt{0.95}f_1 + \sqrt{0.05}f_2$	1	1	> 1	4
6	$f_6 = \sqrt{0.95}f_1 + \sqrt{0.05}f_3$	1	1	< 1	8
7	$f_7 = \sqrt{0.95}f_1 + \sqrt{0.01}f_2 + \sqrt{0.04}f_4$	1	1	< 1	13.6
8	$f_8 = \sqrt{0.95}f_1 + \sqrt{0.05}f_4$	1	1	< 1	16

Note that for our benchmark function f_0, it has variance of 1, effective dimension of 4, and conditional tail dimension of 4.24. This implies that if effective dimension is a good indicator of the efficiency of QMC, then $f_i, i = 1, \ldots, 8$, with effective dimension smaller than 4 should be more efficient than the benchmark; i.e. $\frac{Var^{QMC}(f_0)}{Var^{QMC}(f_i)} > 1$. The results seem

to be consistent for functions f_i for $i = 1, \ldots, 5$. For functions f_6, f_7 and f_8, they are less efficient than f_0 even though their effective dimensions are one. The measure given by the CTD according to (2.15) appears to be a better indicator. For the same function f_6, f_7 and f_8, their CTD ranging from 8 to 16, which are substantially larger than the CTD of f_0 of 4.24. Hence there is no reason why these functions are better than the benchmark function f_0, even though their effective dimension is one.

2.6 Conclusion

We proposed a new measure of effective dimension known as the delta dimension. Comparing to the superposition dimension and truncation dimension, the delta dimension is a simple and easy to compute, and it helps to decompose a function into the summary of an analytical tractable function and a remainder whose variance is close to zero. This also provides a further research opportunity on seeking better QMC-based algorithms for evaluating high-dimensional integral. We also proposed the tail dimension and the conditional tail dimension to quantify the effective dimensions, and proved that the conditional tailed dimension is a better measurement than the effective dimension.

Key Drawbacks of this chapter

1. When using Delta dimension random variable, $f_{D \leq Tail}$ may be correlated with $f_{D>Tail}$ although their first order approximations are uncorrelated, and hence, larger $f_{D \leq Tail}$ may not necessarily give smaller $f_{D>Tail}$
2. Although tail dimension and conditional tail dimension are good measures of the efficiency of QMC methods, it is difficult to design a PGM that minimizes these ratios.

Chapter 3

Dimension Reduction in Quasi-Monte Carlo

Key contributions of this chapter

1. Propose a new path generation method (PGM) which we denote as the delta control (DC) method;
2. Demonstrate that the DC method is a general method in that it encompasses many existing PGMs as its special cases. Furthermore because of its flexibility, we provide three implementations of DC, which we label them as FC, FP, and FMIX methods.
3. Extensive numerical results are provided to demonstrate the competitive efficiency of the DC-based methods.
4. The DC method is not only powerful for high-dimensional problems, but its effectiveness remains even on problems with (multiple) discontinuities. This feature distinguishes DC method from other existing PGMs.
5. Demonstrate through the numerical results, the conditional tail dimension (CTD) can be a good measure of the efficiency of the QMC-based PGM.

3.1 Introduction

Because of the complexity of derivative securities and the sophistication of financial models, the integrals associated with finance applications typically cannot be evaluated analytically and efficient numerical methods are demanded. Consequently, Monte Carlo (MC) method, which is first introduced in quantitative finance by Boyle (1977), becomes a popular numerical method. The MC method, however, is often being criticized to be inefficient despite its convergence rate is independent of the dimension, as it only attains a convergence rate of $O(1/\sqrt{N})$ where N is the number of simulation trials.

Since 90s, several studies have surfaced advocating the use of quasi-Monte Carlo (QMC) methods to high-dimensional finance applications. The QMC offers a convergence rate of $O\left(N^{-1}(\log N)^d\right)$ in dimension d. This rate is asymptotically more efficient than that of the MC, especially in low dimension cases. The results in Joy et al. (1996) and Paskov and Traub (1995) show that the QMC yields a much higher accuracy than the MC method, even for dimension of several hundreds. As a result of both theoretical and empirical findings, there is a surge of interest among financial industries and academics in using the method of QMC.

However, it has been reported that there are some types of problems for which QMC may not be friendly in the sense that the QMC may not be effective. To highlight the issue, let us first look at a fundamental property that explains theoretical background of the QMC method. The QMC method can be regarded as a deterministic counterpart to the MC method, that is, the quadrature points are deterministically chosen in $[0,1]^d$. From the well-known Koksma-Hlawka inequality (see Niederreiter, 1992), the approximation error, denoted by $|err|$, is bounded by a product of two factors,

$$|err| \leq V_{HK}(f) D^*(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N). \tag{3.1}$$

The factor $D^*(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N)$ is called star discrepancy which quantifies the uniformity of

the points $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N$. The second factor $V_{HK}(f)$ in inequality (3.1) is called the bounded variation in the sense of Hardy and Krause.

The Koksma-Hlawka inequality provides important insights about factors that influence the performance of the QMC. The dimension and the smoothness or discontinuity of the integrand are two such key factors. It is shown that the QMC method yields a deterministic error bound $O(N^{-1}(\log N)^d)$ for functions of a finite variation of d dimensions. This convergence rate is asymptotically faster than that of the MC, but it clearly depends on the dimension d.

In the QMC, a low-discrepancy sequence that satisfies the discrepancy bound of D^* is utilized for the simulation, hence, theoretically it outperform the MC asymptotically . In the case for low dimensional integration, the QMC attains approximately a rate of $O(N^{-1}(\log N)^d)$ that outperforms the MC. In a higher dimensional case, however, the powers of $\log(N)$ in the convergence rate of the QMC are not negligible for practical sample sizes, i.e., the theoretical higher asymptotic convergence rate of the QMC is not achievable for practical applications in high dimensions. In non-finance applications, in particular, Bratley et al. (1992) indicated that a QMC method offers no practical advantage over MC, even for problems with relatively lower dimensions.

Another important factor for a success of QMC in high dimensions lies in the distinction between nominal dimension and effective dimension. Using the "analysis of variance" (ANOVA) decomposition of a function, Caflisch et al. (1997) defined two notions of the effective dimension: truncation dimension and superposition dimension. (See Chapter 2) Essentially, the truncation dimension indicates the number of important variables which predominantly capture the evaluation function, and the superposition dimension measures to what extent the low-order ANOVA terms dominate the function.

The argument implies that numerical accuracy of the QMC in practice can be improved if we can effectively reduce the dimension of the function of interest. Many dimension reduction techniques such as the Brownian bridge (BB) (Moskowitz and Caflisch 1996),

the principal component analysis (PCA) (Acworth et al. 1998), the linear transformation (LT) (Imai and Tan 2006), and fast orthogonal transformation (FOT) method (Leobacher 2012), have been proposed along this line to increase the efficiency of the QMC.

The efficiency of QMC also depends on the continuity of the function. If the function is discontinuous, its variation in the sense of Hardy Krause might be unbounded and hence the Koksma-Hlawka inequality is no longer applicable. Discontinuities may have a detrimental effect on the efficiency of QMC, as pointed out in Berblinger et al. (1997), Morokoff and Caflisch (1995) and Moskowitz and Caflisch (1996). The orthogonal transformation (OT) proposed in Wang and Tan (2013) provides a QMC-friendly path generation method (PGM), even in the presence of discontinuity.

Motivated by all of the above issues associated with QMC, this chapter proposes a new PGM that has the potential of enhancing numerical efficiency of the QMC method. The new PGM is denoted as the delta control (DC) approach. As we will shortly explain, the DC has least the following properties:

- The DC approach can be viewed as a general approach in that it encompasses many existing PGMs as its special case;
- The numerical examples demonstrate that the DC method is competitively more efficient than the existing methods;
- Finally, the DC method can handle problems involving multiple discontinuity and high-dimensionality at the same time. This is the feature that offers significant competitive advantage of our proposed DC method, relative to other existing methods.

The remaining of this chapter is organized as follows. The next section provides a brief overview of the PGMs. Section 3.3.2, which is the core contribution of this chapter, describes our proposed DC-based PGM. At high level, the proposed DC method can be described as follows. We first demonstrate that by a judiciously transforming a given function of interest to another equivalent function, the nominal dimension can be artificially

reduced. More importantly, through the first order Taylor expansion and the delta method, we show that a new PGM can be constructed. As formally shown in Section 3.3.2, the new PGM can be formulated in terms of the Jacobian matrix $\boldsymbol{J}$ and the functional covariance matrix $\boldsymbol{A}_f$ (see (3.11)). Armed with this result, the remaining task is to determine an appropriate functional transformation, and hence the Jacobian matrix $\boldsymbol{J}$ and the functional covariance matrix $\boldsymbol{A}_f$. As discussed in Section 3.3.2, the choice of the functional transformation depends on the structure of the original function of interest. The existing PGM matrix such as that based on the STD, BB, PCA can be exploited to provide further dimension reduction. In fact, depend on the choice of the functional transformation, many of the existing PGM can be derived as special cases of the DC method. Because of the flexibility of defining the functional transformation, we propose three possible implementations of the DC method, and their relative efficiencies are assessed in the numerical illustrations in Section 3.4.

3.2 A Brief Review of Path Generation Method (PGM)

Recall that in Chapter 1 we briefly review the MC and QMC approaches to estimating high-dimensional European-style derivative securities. Under the Black-Scholes model, the key to pricing high-dimensional European-style derivative securities boils down to evaluating the following expectation involving normal random variables:

$$V(g) := \mathbb{E}\left[g(\boldsymbol{x})\right] = \frac{e^{-rT}}{(2\pi)^{d/2}\sqrt{\det \boldsymbol{\Sigma}}} \int_{\mathbb{R}^d} g(\boldsymbol{x}) \exp\left(-\frac{1}{2}\,\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right) d\boldsymbol{x}, \tag{3.2}$$

where $\boldsymbol{x} \sim N_d(\boldsymbol{0}, \boldsymbol{\Sigma})$ denotes a d-dimensional normal vector with mean vector $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma}$, the function $g : \mathbb{R}^d \mapsto \mathbb{R}$ captures the payoff of the derivative security at maturity T, and r is the risk-free rate, T is time of the expiration date of the option, $\boldsymbol{\Sigma}$ represents a covariance matrix.

By exploiting the transformation $\boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{z}$ where $\boldsymbol{z} = (z_1, \ldots, z_d)^T \sim N_d(\boldsymbol{0}, \mathbf{I}_d)$, $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{\Sigma}$,

and each normal variate can be transformed by $\boldsymbol{z} = \Phi^{-1}(\boldsymbol{u}) = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d))^T$ with $\boldsymbol{u} \in (0,1)^d$, where Φ^{-1} represents the inverse of the cumulative distribution function of the standard normal, it is easy to see that (3.2) simplifies to

$$V(g) = \frac{e^{-rT}}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} G(\boldsymbol{A}\boldsymbol{z}) \exp\left(-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{z}\right) d\boldsymbol{z} = e^{-rT} \int_{(0,1)^d} G(\boldsymbol{A}\,\Phi^{-1}(\boldsymbol{u}))\, d\boldsymbol{u}. \tag{3.3}$$

Note that the above integral involves only the transformed function $G(\cdot)$ that depends explicitly on the decomposed matrix $\boldsymbol{A}$ and the d-dimensional points $\boldsymbol{u} \in (0,1)^d$. This expression also implies that its value can be estimated via sampling method. In particular for the QMC method, let $\mathcal{P} := \{\mathbf{u}_i, i = 1, \ldots, N\}$ be a low discrepancy point set over the unit cube $[0,1]^d$, then the QMC estimate of (3.3) is given by

$$Q(g, \boldsymbol{A}, \mathcal{P}) = \frac{e^{-rT}}{N} \sum_{k=1}^{N} G(\boldsymbol{A}\,\boldsymbol{z}_k) = \frac{e^{-rT}}{N} \sum_{k=1}^{N} G(\boldsymbol{A}\,\Phi^{-1}(\mathbf{u}_k)). \tag{3.4}$$

Note that the above QMC estimate is a function of the point set $\mathcal{P}$, the payoff function G and the decomposed matrix $\boldsymbol{A}$. More importantly, for a given point set $\mathcal{P}$ and a payoff function G, it is well-known that the precision of the QMC-based estimator $Q(g, \boldsymbol{A}, \mathcal{P})$ is highly dependent on the choice of $\boldsymbol{A}$. In the context of the Black-Scholes model, the decomposed matrix $\boldsymbol{A}$ determines how a trajectory of a Brownian motion is being generated. For this reason, $\boldsymbol{A}$ is known as a path-generation matrix of the Brownian motion.

A key insight to the above transformation is that the path-generation matrix $\boldsymbol{A}$ can be arbitrary as long as it satisfies $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{\Sigma}$. Hence different selection of $\boldsymbol{A}$ yields different path generation method (PGM), and hence different estimator of (3.4). As such PGM such as that based on the Brownian bridge (BB) (Moskowitz and Caflisch 1996), the principal component analysis (PCA) (Acworth et al 1998), the linear transformation (LT) (Imai and Tan 2006), orthogonal transformation on discontinuous function (OT) (Wang and Tan 2012) and fast orthogonal transformation (FOT) method (Leobacher 2012), have been proposed. These methods increase the efficiency of QMC via dimension reduction.

To conclude this section, it should be emphasized that if the point set $\mathcal{P}$ were from a random sequence, then the resulting estimator $Q(g, \boldsymbol{A}, \mathcal{P})$ becomes the estimator of the MC method. The efficiency of MC, on the other hand, does not depend on the choice of $\boldsymbol{A}$.

3.3 Delta Control (DC) Method: A New PGM

3.3.1 The Delta Method

The delta method is a method of deriving the approximate distribution of a nonlinear function of an estimator from the approximate distribution of the estimator itself. If $\boldsymbol{g}$ is a non-linear, differentiable vector-to-vector function, the best linear approximation, which is the Taylor series up through linear terms, is

$$\boldsymbol{g}(\boldsymbol{y}) - \boldsymbol{g}(\boldsymbol{x}) \approx \nabla \boldsymbol{g}(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x}),$$

where $\boldsymbol{x}$, $\boldsymbol{y}$ are arbitrary random variables, $\boldsymbol{J} := \nabla \boldsymbol{g}(\boldsymbol{x})$ is the matrix of partial derivatives with respect to $\boldsymbol{x}$, also called the Jacobian matrix. If $g_i(\boldsymbol{x})$ denotes the i-th component of vector $\boldsymbol{g}(\boldsymbol{x})$, then the (i, j)-th component of $\boldsymbol{J}$ is $\frac{\partial g_i(\boldsymbol{x})}{\partial x_j}$.

The delta method is particularly useful when $\hat{\boldsymbol{\theta}}$ is an estimator and $\boldsymbol{\theta}$ is the unknown true (vector) parameter value it estimates, and the delta method says

$$\boldsymbol{g}(\hat{\boldsymbol{\theta}}) - \boldsymbol{g}(\boldsymbol{\theta}) \approx \nabla \boldsymbol{g}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

It is not necessary that $\boldsymbol{\theta}$ and $\boldsymbol{g}(\boldsymbol{\theta})$ be vectors of the same dimension. Hence it is not necessary that $\nabla \boldsymbol{g}(\boldsymbol{\theta})$ be a square matrix. Here are some properties of Delta method:

1. The delta method is a good approximation for sufficiently small values of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ and a bad approximation for sufficiently large values of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$.

2. Overall, it is good if those "sufficiently large" values have small probability.

3. The delta method is particularly easy to use when the distribution of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ is multivariate normal, exactly or approximately.

4. The reason this is easy is that normal distributions are determined by their mean vector and variance matrix, and there is a theorem which gives the mean vector and variance matrix of a linear function of a random vector.

Theorem 3.3.1. *Theorem of Delta Method of Functions of Normal R.V. Suppose* $\hat{\boldsymbol{\theta}}$ *is normal with mean vector* $\boldsymbol{\theta}$ *and variance-covariance matrix* $\boldsymbol{\Sigma}$*, then* $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ *has mean* $\mathbf{0}$ *and variance-covariance* $\boldsymbol{\Sigma}$*, and*

$$E[\boldsymbol{g}(\hat{\boldsymbol{\theta}}) - \boldsymbol{g}(\boldsymbol{\theta})] \approx 0$$

$$Var[\boldsymbol{g}(\hat{\boldsymbol{\theta}}) - \boldsymbol{g}(\boldsymbol{\theta})] \approx \nabla \boldsymbol{g}(\boldsymbol{\theta}) \boldsymbol{\Sigma} \nabla \boldsymbol{g}(\boldsymbol{\theta})^T$$

Remark 3.3.1. *In this Chapter, we will focus on controlling functions* $f_i = g_i$ *and* $\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$*, hence, we have*

$$\begin{aligned} E(\boldsymbol{f}(\boldsymbol{x})) &\approx E(\boldsymbol{f}(\mathbf{0})), \\ Var(\boldsymbol{f}(\boldsymbol{x})) &\approx \boldsymbol{J} \boldsymbol{\Sigma} \boldsymbol{J}^T, \end{aligned}$$

where $\boldsymbol{J}$ *is the derivative of* $\boldsymbol{f}$ *with respect to* $\boldsymbol{x}$ *evaluated at* $\mathbf{0}$*.* $\boldsymbol{f}(\boldsymbol{x}) \approx E(\boldsymbol{f}(\mathbf{0})) + \boldsymbol{J}\boldsymbol{x}$*, i.e, this is a vector of first order Taylor approximation of* $\boldsymbol{f}(\boldsymbol{x})$*.*

3.3.2 Our New PGM

In this section we describe our proposed PGM, which is denoted as the Delta Control (DC) method. In the first part of this section we first explain why a functional transformation leads to a new PGM that is defined in term of the Jacobian matrix $\boldsymbol{J}$ and the functional covariance matrix $\boldsymbol{A}_f$ with potential lower nominal dimension. Based on these results, the second part of this section focuses on the choice of the functional transformation, and

hence the proposed new PGM.

Recall that fundamentally we are interested in providing an efficient QMC-based algorithm for estimating

$$\mathbb{E}\left[g(\boldsymbol{x})\right], \qquad \text{where } \boldsymbol{x} \sim N_d(\mathbf{0}, \boldsymbol{\Sigma}). \tag{3.5}$$

To continue, let us introduce a new function $\psi : \mathbb{R}^k \mapsto \mathbb{R}$, $k < d$, such that

$$g(\boldsymbol{x}) = \psi(\boldsymbol{f}) = \psi(f_1(\boldsymbol{x}), \cdots, f_k(\boldsymbol{x})), \tag{3.6}$$

where $\boldsymbol{f} = (f_1, \ldots, f_k)$ with $f_i : \mathbb{R}^d \mapsto \mathbb{R}$.

Note that the original problem is given by the function $g(\cdot)$ in terms of the normal random vector $\boldsymbol{x}$ of d (nominal) dimensions. The functional transformation (3.6) enables us to view the original problem equivalently as $\psi(\cdot)$ but depends on $\boldsymbol{f}$ of k functions $f_1, \ldots, f_k$. By construction, the nominal dimension of the problem is artificially reduced from d to k.

As an illustration, suppose we are interested in pricing an Asian option with its payoff at maturity given by $\max\left(\frac{1}{d}\sum_{i=1}^{d} S_i - K, 0\right)$, where S_i denotes the stock price at time i and K is the strike price. A possible candidate for ψ and $\boldsymbol{f}$ are

$$\psi(\boldsymbol{f}) = \max\left(\sum_{i=1}^{d} f_i - K, 0\right) \qquad \text{and} \qquad f_i = \tfrac{1}{d} S_i. \tag{3.7}$$

Another plausible set of candidates is

$$\psi(\boldsymbol{f}) = \max\left(f_1 - K, 0\right) \qquad \text{and} \qquad f_1 := \tfrac{1}{d}\sum_{i=1}^{d} S_i \tag{3.8}$$

In this case, $f_i, i = 2, \ldots, d$ can be arbitrary. We will discuss how to determine the functions ψ and $\boldsymbol{f}$ in the second part of this section.

In what follows, we argue that the introduction of ψ via (3.6) provides us with another way of defining a new PGM. To see this, let us consider the following delta approximation

based on the first order Taylor expansion around an arbitrary point $\boldsymbol{x}_0$:

$$f_i(\boldsymbol{x}) - f_i(\boldsymbol{x}_0) \approx \nabla f_i(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0), \tag{3.9}$$

where $\boldsymbol{J} := \begin{pmatrix} \nabla f_1(\boldsymbol{x}_0) \\ \vdots \\ \nabla f_k(\boldsymbol{x}_0) \end{pmatrix}$ is the Jacobian matrix of $\{f_1, \cdots, f_k\}$ with respect to $\{x_1, \cdots, x_d\}$ and is evaluated at $\boldsymbol{x}_0$. In other words, we have

$$\boldsymbol{J} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_d} \\ \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_2}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1} & \cdots & \frac{\partial f_k}{\partial x_d} \end{pmatrix}_{|\boldsymbol{x}=\boldsymbol{x}_0}.$$

Assuming that $\boldsymbol{x}_0 = \boldsymbol{0}$ and $\boldsymbol{f}(\boldsymbol{x}_0) = \boldsymbol{0}$, the function ψ in equation (3.6) can be written as

$$\psi(\boldsymbol{f}(\boldsymbol{x})) \approx \psi(\boldsymbol{J}\boldsymbol{x}). \tag{3.10}$$

This result indicates that, under the delta method, the vector $\{f_1, \cdots, f_k\}$ can be approximated by a linear function of the normal vector $\boldsymbol{x}$. More precisely we have

$$\begin{pmatrix} f_1 \\ \vdots \\ f_k \end{pmatrix} \approx \boldsymbol{J}\boldsymbol{x} \sim N_d(\boldsymbol{0}, \boldsymbol{\Sigma_f}),$$

where $\boldsymbol{\Sigma_f} = \boldsymbol{J}\boldsymbol{\Sigma}\boldsymbol{J}^{\mathrm{T}}$.

Let $\boldsymbol{A}$ and $\boldsymbol{A_f}$ denote the decomposed matrices of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma_f}$, respectively; i.e., $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{\Sigma}$ and hence an appropriate choice of $\boldsymbol{A_f}\boldsymbol{A_f}^{\mathrm{T}} = \boldsymbol{\Sigma}_f$. It is therefore easy to see that the decomposed matrices of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma_f}$ are related to the Jacobian matrix $\boldsymbol{J}$ via

$$\boldsymbol{A} = \boldsymbol{J}^{-1}\boldsymbol{A}_f. \tag{3.11}$$

The above relation demonstrates the reason for considering the transformation given in (3.6). From the vector-valued function $\boldsymbol{f}$, we obtain the corresponding $\boldsymbol{J}$ and $\boldsymbol{A}_f$. This in turn enables us to derive a new PGM $\boldsymbol{A}$ based on (3.11) for estimating (3.5). This is named as Delta control (DC) path generation method.

We now proceed to the selection of ψ and $\boldsymbol{f}$. This is important to the DC method as it explicitly determines $\boldsymbol{J}$, and hence $\boldsymbol{A}_f$ and $\boldsymbol{A}$. The idea underlies the selection of $\boldsymbol{f}$ is such that the effective dimension of the function ψ is as small as possible. Recall that in Chapter 2 an extensive discussion was provided with respect to the effective dimension of a function.

Ideally the functions $\boldsymbol{f}$ and hence $\boldsymbol{A}_f$ should be chosen in such a way that the effective dimension of the transformed function ψ is optimally minimized. In practice this can be a challenging goal. Instead of seeking a best overall choice of $\boldsymbol{f}$, here we argue that most of the practical problems can be divided into the following three categories and hence heuristic argument is provided to facilitate the selection of a reasonable choice of $\boldsymbol{f}$ and hence $\boldsymbol{A}_f$. The resulting methods, which are special cases of the DC method, are denoted as FC, FP, and FMIX. The numerical examples to be conducted in the next section attests to the effectiveness of these methods. We now describe the following three implementations of the DC methods.

FC Method The method of FC is motivated by the possibility that in some cases, the function $\boldsymbol{f}$ may have one highly dominating component. Assuming that f_1 is the dominant function, then the function ψ can be approximated as

$$\psi(f_1, \cdots, f_k) \approx \phi(f_1),$$

where ϕ is a function that depends on the dominating function f_1 in terms of variability. In this case, the Cholesky decomposition is suitable since it focuses on the first column of the matrix. In fact, under this particular setting its effective dimension can optimally be reduced to one. A practical example that fits into this category is

a weighted arithmetic average option (see Example 3.8). We refer to this method as functional Cholesky construction (FC).

FP Method At the other extreme, let us suppose that the function $\psi(f_1, \cdots, f_k)$ is of the following form

$$\psi(f_1, \cdots, f_k) \approx \phi(f_1 + f_2 + \cdots + f_k),$$

where the sum $f_1 + f_2 + \cdots + f_k$ emphasizes that each of these functions contributes equally to the function ϕ. For function with this structure, it is found that the construction based on the principal component analysis (PCA) can be very effective. In fact numerous studies have documented that applying PCA to the above ϕ can optimally reduce the superposition dimension of ψ to one. See Wang and Sloan (2011). We refer to this method as functional principal component construction (FP). **Note: this is a very special case where all the k components are equally important, however, when there is no information concluding that some function is more important than others, PCA method is still a good way to construct a good PGM.**

FMIX Method While the above discussion (heuristically) addresses the two extremes cases for which the methods of FC and FP can be effective, it should be pointed out that in practice the decomposed matrix $\boldsymbol{A}_f$ needs not to be strictly derived based on either of these two methods. In particular, most problems do not necessary fall into the two extremes as prescribed by the above two categories. Our proposed third method is a mixture of both FC and FP methods so that the resulting method is referred as the "FMIX" method. Under the FMIX method and for $k_1 < k$, the first k_1 columns of $\boldsymbol{A}_f$ are derived from the method of FC and the remaining columns are based on the FP method. For example, by inspecting from $\boldsymbol{f} = (f_1, f_2, \cdots, f_k)$ with corresponding Jacobian matrix $\boldsymbol{J}$ and functional covariance matrix $\boldsymbol{\Sigma}_f = \boldsymbol{J}\boldsymbol{\Sigma}\boldsymbol{J}^T$, the method of FC is first applied to the functions $f_1, \cdots, f_{k_1}$. This produces the first k_1 columns of $\boldsymbol{A}_f$, i.e. $(\boldsymbol{A}_f)_{\cdot 1:k_1}$. This in turn leads to the first k_1 columns of the

PGM $\boldsymbol{A}$, i.e. $\boldsymbol{A}_{\cdot 1:k_1} = \boldsymbol{J}^{-1}(\boldsymbol{A}_f)_{\cdot 1:k_1}$. The residual columns are then determined by the FP method; i.e. $(\boldsymbol{A}_f)_{\cdot k_1+1:d}(\boldsymbol{A}_f)^T_{\cdot k_1+1:d} = \boldsymbol{J}(\boldsymbol{\Sigma} - \boldsymbol{A}_{\cdot 1:k_1}\boldsymbol{A}^T_{\cdot 1:k_1})\boldsymbol{J}^T$. Then, the first k columns of $\boldsymbol{A}$ can be constructed as

$$\boldsymbol{A}_{\cdot 1:k} = [\boldsymbol{A}_{\cdot 1:k_1}, \boldsymbol{A}_{\cdot k_1+1:k}],$$

where $\boldsymbol{A}_{\cdot k_1+1:k} = \boldsymbol{J}^{-1}(\boldsymbol{A}_f)_{\cdot k_1+1:d}$.

We now make following remarks:

Remark 3.3.2.

1. *When $k < d$, our methods determine the first k columns of the PGM $\boldsymbol{A}$, i.e. $\boldsymbol{A}_{\cdot 1:k}$. The remaining $d-k$ columns of $\boldsymbol{A}$ are constructed arbitrarily as long as $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{\Sigma}$ so that*

$$E[\psi(f_1(\boldsymbol{x}), \cdots, f_k(\boldsymbol{x}))] = E[g(\boldsymbol{x})]$$

according to Equation (3.6), where $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{z}$ and $\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{I})$.

2. *It is desirable to decrease the dimension k of the function ψ. A smaller k indicates that nominal dimension of the transformed function has been decreased.*

3. *Even if the dimension k is not reduced, it is also promising when the structure of the function ψ has a form that can be exploited by either the methods of FC, or FP, or even FMIX.*

4. *Since $\boldsymbol{J}$ is set as a k by d matrix, it can be computationally intensive for large k and d. For an efficient computation of Jacobian matrix and its inverse, see Xu et al. (2016). Note that as $\boldsymbol{J}$ may not be a full rank matrix, the inverse of $\boldsymbol{J}$ is not unique. See Moore (1920). Hence, there exist many different ways of constructing the optimal PGM $\boldsymbol{A}$ to achieve the same QMC efficiency under DC approach.*

5. *Finally, it should be emphasized that although the proposed method is based on the approximation method in (3.9) for obtaining $\boldsymbol{A_f}$, it uses the original decomposed matrix*

$\boldsymbol{A}$ in utilizing the QMC. It guarantees that the QMC-based estimator is unbiased.

To conclude this section, we point out that the proposed DC approach can be interpreted as a general class of method of deriving the PGM $\boldsymbol{A}$. Hence many of the existing PGM is a special case of the proposed DC method. To see this suppose we have $f_i(x) = x_i, i = 1, \ldots, k$. This is a trivial case in that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_f$. Hence any PGM used to determine $\boldsymbol{A}_f$ is also the same method that determines the PGM directly from $\boldsymbol{A}$.

As another example, by setting $f_1(\boldsymbol{x}) := x_d$, $f_2(\boldsymbol{x}) := x_{d/2}$, ..., and applying FC, the resulting PGM reduces to BB method. On the other hand, if $\psi(\boldsymbol{f}) := f_1(\boldsymbol{x})$ with $f_1(\boldsymbol{x}) = g(\boldsymbol{x})$ and apply the FC in the second phase, it is then equivalent to the LT method in which the only first column of the orthogonal matrix is optimized.

3.4 Numerical Results

In this section, we demonstrate that the proposed methods are applicable to a wide range of practical problems, and then show that they significantly enhance numerical efficiency of QMC.

To compare the efficiency of our proposed methods relative to the crude MC methods, we provide the variance reduction ratio (VRR). For each option, we simulate its price using crude MC method and various QMC-based methods. In addition to estimating the option price, we also compute the sample variance of the estimated price. We gauge the relative efficiency of QMC-based method to MC method by computing the VRR defined by

$$\text{VRR} := \frac{Var^{MC}(f)}{Var^{QMC}(f)}. \tag{3.12}$$

Here $Var^{MC}(f)$ denotes the sample variance of the problem f estimated by the crude MC method and $Var^{QMC}(f)$ represents the sample variance of the problem f estimated by the respective QMC-based method. A VRR greater than one implies that the QMC-

based method is more efficient than the crude MC method. The larger the VRR the more favourable the underlying QMC-based method.

For the crude MC method, we use the 64-bit Mersenne Twister pseudo-random generator. For the methods involving QMC, we use the Sobol's Quasirandom Sequence Generator by Bratley and Fox (1988), and a random linear scramble combined with a random digital shift by Hong and Hickernell (2003) and Matousek (1998).

For the methods involving QMC, their sample variances are obtained by independent batches. By denoting N as the number of low discrepancy points and M as the number of independent batches, then $\{\boldsymbol{u}_j^{(k)} \in [0,1]^d; j = 1, \ldots N, k = 1, \ldots, M\}$ corresponds to the j-th d-dimensional low discrepancy point of the k-th independent batch. An estimate of the option price for the k-th independent batch is given by

$$\hat{\mu}^{(k)} := \frac{1}{N} \sum_{j=1}^{N} g\left(\boldsymbol{u}_j^{(k)}\right). \tag{3.13}$$

In practice, $\boldsymbol{u}_j^{(k)}$ needs to be transformed to the desired random variates depending on the distributional assumption of the underlying assets. The option price is then approximated by the pooled estimate over M independent estimates $\{\hat{\mu}^{(k)}, k = 1, \ldots, M\}$ as

$$\hat{\mu} := \frac{1}{M} \sum_{k=1}^{M} \hat{\mu}^{(k)}. \tag{3.14}$$

The sample variance, which measures the variability of $\hat{\mu}$, is then computed as

$$\hat{\sigma}^2 := \frac{1}{M-1} \sum_{k=1}^{M} \left(\hat{\mu}^{(k)} - \hat{\mu}\right)^2. \tag{3.15}$$

In other to ensure fairness between MC and QMC methods, the number of points for the MC method is MN so that the variance of the MC estimate can be produced in the usual manner.

Throughout this section, we assume the underlying asset price process follows the Black-Scholes model with the following parameter values:

- initial asset price $S_0 = 100$,
- time to maturity $T = 1$,
- risk-free rate $r = 0.1$, and
- volatility $\sigma = 0.2$.

In addition to the proposed DC-based methods, for comparison we also consider the following methods which are commonly used in the QMC methods (see also Chapter 1):

1. Standard approach (STD) with $\boldsymbol{A}$ corresponds to the Cholesky decomposition of $\boldsymbol{\Sigma}$;
2. Principal component analysis (PCA) method of Acworth et al. (1998);
3. Brownian bridge (BB) construction of Moskowitz and Caflisch (1996);
4. Linear Transformation (LT) construction of Imai and Tan (2006).

To compute the estimators (3.13)-(3.15) we use $M = 100$ and $N = 4096$ for each of the above methods. Finally, by setting $\alpha = 0.95$, whenever applicable we also produce $F_D(1), F_D(2), F_D(3)$, and $CTD_\alpha(D)$, where $D \in \{D_S, D_T, D_D\}$. As defined in Chapter 2, $F_D(\cdot)$ denotes the cumulative mass function (CMF) of the dimension random variable D and $CTD_\alpha(D)$ corresponds to the conditional tail dimension (2.14) at α confidence level. The random variables D_S, D_T, and D_D represent, respectively, the superposition, truncation and delta dimension random variables. The $CTD_\alpha(D)$ for $D \in \{D_S, D_T, D_D\}$ are computed according to the (2.15). Recall also that comparing their CMF allows us to address the relative magnitude of their effective dimension; i.e. the $Tail_\alpha(D)$ for $D \in \{D_S, D_T, D_D\}$ as defined in (2.13).

3.4.1 Arithmetic Asian Option

For the first example, we consider discretely monitored arithmetic Asian call options with different exercise prices and different monitoring frequencies. Their payoffs are given by

$$h_1(\boldsymbol{S}) = \max\left(\frac{1}{d}\sum_{i=1}^{d} S_i - K, 0\right),$$

where $\boldsymbol{S} = (S_1, \ldots, S_d)$ with S_i represents i-th monitored underlying asset price, and K is the exercise price of the option. The arithmetic Asian option is one of the most popular financial instruments to hedge an operating profit in nonfinancial industries. Option of this type has generated a significant interest among academics due to its lack of tractability and hence numerous numerical procedures have been proposed.

For the FP method, we set $\psi(\boldsymbol{f}) = \max\left(\sum_{i=1}^{d} f_i - K, 0\right)$ with $f_i = \frac{1}{d}S_i$. For the FC method, we set $\psi(\boldsymbol{f}) = \max(f_1 - K, 0)$ with $f_1 := \frac{1}{d}\sum_{i=1}^{d} S_i$ and $f_i, i = 2, \ldots, d$ can be arbitrary. In the examples we simply follow the order of the Brownian bridge construction.

Table 3.1: **VRRs of Arithmetic Asian Options** $(h_1(\boldsymbol{S}))$ **on various PGMs**

d	K	STD	PCA	BB	LT	FP	FC
16	90	247.50	13318.52	3286.68	14257.02	16469.31	15087.19
16	100	75.76	9808.65	1500.36	9241.58	**11694.19**	10217.05
16	110	28.32	3911.06	553.64	4471.20	5248.01	4242.79
64	90	118.68	17331.31	3239.18	12621.27	21272.76	15987.61
64	100	42.33	7912.42	1433.32	6973.41	**13907.24**	7886.25
64	110	12.67	5016.97	464.16	4393.73	6167.23	4830.13

Table 3.2: **Cumulative Mass Function and Conditional Tail Dimension of $h_1(\boldsymbol{S})$ with $K = 100$ and $\alpha = 95\%$**

d	CMF	STD	PCA	BB	LT	FP	FC
Truncation Dimension(D_T)							
16	$F_{D_T}(1)$	0.136436	0.987793	0.752961	0.999399	0.983821	1.000000
16	$F_{D_T}(2)$	0.262657	0.997401	0.933346	0.999581	0.997911	1.000000
16	$F_{D_T}(3)$	0.384212	0.999182	0.954294	0.999581	0.999135	1.000000
16	CTD_α	13.041883	1.329084	9.302353	**1.032612**	1.388176	**1.000049**
64	$F_{D_T}(1)$	0.030108	0.987554	0.728587	0.999629	0.983768	0.999715
64	$F_{D_T}(2)$	0.064833	0.998105	0.927200	0.999682	0.997665	0.999808
64	$F_{D_T}(3)$	0.100569	0.999545	0.950177	0.999749	0.999082	0.999827
64	CTD_α	49.325668	1.311369	33.437341	**1.041029**	1.410484	**1.019190**
Delta Dimension(D_D)							
16	$F_{D_D}(1)$	0.199431	0.977257	0.789979	0.982863	0.973734	0.982970
16	$F_{D_D}(2)$	0.335630	0.990446	0.939414	0.986819	0.990664	0.989711
16	$F_{D_D}(3)$	0.452092	0.994205	0.957359	0.988008	0.993963	0.991132
16	CTD_α	12.950284	**2.128261**	8.919705	2.472921	**2.190194**	2.351469
64	$F_{D_D}(1)$	0.067515	0.977229	0.773450	0.982731	0.973656	0.983093
64	$F_{D_D}(2)$	0.118362	0.990393	0.934347	0.989794	0.990520	0.989187
64	$F_{D_D}(3)$	0.161749	0.994055	0.953493	0.991752	0.993755	0.990568
64	CTD_α	49.453242	**2.593105**	30.883266	3.425480	**2.672298**	3.195196

By assuming $d \in \{16, 64\}$ and $K \in \{90, 100, 110\}$, the results are depicted in Tables 3.1 and 3.2. Table 3.1 reports VRRs of the PGMs STD, PCA, BB, as well as our proposed FP and FC. Table 3.2 presents the CMF and CTD for the truncation and delta dimension random variables. Based on these results, we make the following remarks:

- The VRRs reported in Table 3.1 are greater than 1, signifying that the QMC-based methods are more efficient than the classical MC.
- The dimension reduction methods PCA, BB, LT, FP and FC significantly outperform the QMC-based STD, demonstrating the tremendous efficiency can be gained from using methods that reduce the effective dimension of the problem.
- More importantly both implementations of FP and FC are very competitive, particularly the Delta-based FP which outperforms the rest of the methods.

- The impact of PGMs on the dimension reduction can also be seen from Table 3.2. Apart from the STD, the remaining methods' CMF are very close to 1, indicating that the effective dimension under these methods tend to be very small. In fact for the methods of LT and FC, the effective dimension reduces to the best case of one at 99% level. Note that the corresponding CMF for the superposition dimension variable is not reported since its $F_{D_S}(1)$ is already very close to 1 for all PGMs and hence their effective dimension are reduced to one in the superposition sense.

- The superposition dimension $F_{D_S}(1)$ is close to 1 and hence its CTD is also close to 1. The truncation and delta dimensions' CTDs suggest that PCA, LT, FP and FC are better than STD and BB. However, the truncation CTDs prefer LT and FC methods, while the delta CTDs prefer PCA and FP methods although their VRRs are all very comparable to each other.

3.4.2 Weighted Arithmetic Average Options

The second example is a weighted average option whose payoff function is given by

$$h_2(\boldsymbol{S}) = \max\left(\sum_{i=1}^{d} w_i S_i - K, 0\right).$$

where w_i represents the weight corresponds to the i-th monitoring asset price S_i. In our examples, we consider two sets of weights:

- Increasing weights such that $w_i = ci$ for $i = 1, ..., d$, where c is a normalized constant;

- Decreasing weights such that $w_i = ci^{-1}$ for $i = 1, ..., d$, where c is a normalized constant.

Controlling weights in these manners allows us to examine the impact these weights may have on the numerical efficiency of the PGMs.

For the FP method, we design $\psi(\boldsymbol{f}) = \max\left(\sum_{i=1}^{d} f_i - K, 0\right)$ with $f_i = w_i S_i$. For the FC, we

set $\psi(\boldsymbol{f}) = \max(f_1 - K, 0)$ with $f_1 := \sum_{i=1}^{d} w_i S_i$. As in the previous example, $f_i, i = 2, \ldots, d$ can be arbitrary and in our examples we simply follow the order of the Brownian bridge construction.

Table 3.3: **VRRs of Weighted Arithmetic Average Asian Option** ($h_2(\boldsymbol{S})$) **with increasing weights on various PGMs**

d	K	STD	PCA	BB	LT	FP	FC
16	90	310.55	10535.70	3947.59	10638.26	10667.03	11032.83
16	100	74.53	9073.61	1859.78	8771.67	8610.59	**9194.50**
16	110	23.51	6187.36	732.58	5972.13	5022.56	5602.22
64	90	71.30	17339.32	4290.79	16239.49	16307.56	17678.54
64	100	25.83	10234.41	2517.37	10368.61	10115.28	**10461.20**
64	110	12.82	4733.65	966.90	4694.12	4496.15	4907.44

Table 3.4: **Cumulative Mass Function and Conditional Tail Dimension of** $h_2(\boldsymbol{S})$ **with increasing weights,** $K = 100$ **and** $\alpha = 95\%$

d	CMF	STD	PCA	BB	LT	FP	FC
Truncation Dimension(D_T)							
16	$F_{D_T}(1)$	0.089534	0.995550	0.842003	0.999417	0.978965	1.000000
16	$F_{D_T}(2)$	0.180833	0.999771	0.947809	0.999730	0.996946	1.000000
16	$F_{D_T}(3)$	0.278347	0.999987	0.950538	0.999749	0.999042	1.000000
16	CTD_α	13.860213	**1.095333**	10.677013	**1.036053**	1.528120	**1.000000**
64	$F_{D_T}(1)$	0.016695	0.996474	0.815324	1.000000	0.980169	0.999860
64	$F_{D_T}(2)$	0.039725	0.999890	0.938918	1.000000	0.996841	0.999860
64	$F_{D_T}(3)$	0.063909	0.999990	0.942124	1.000000	0.998796	0.999874
64	CTD_α	52.676944	**1.072936**	38.316367	**1.000000**	1.495347	1.019084
Delta Dimension(D_D)							
16	$F_{D_D}(1)$	0.142967	0.983420	0.863754	0.986989	0.973797	0.987485
16	$F_{D_D}(2)$	0.250099	0.993159	0.952060	0.990024	0.992016	0.992666
16	$F_{D_D}(3)$	0.348544	0.995608	0.955243	0.990174	0.995406	0.994516
16	CTD_α	13.869771	**1.848697**	9.964406	2.467868	**2.022111**	**1.996717**
64	$F_{D_D}(1)$	0.044253	0.983590	0.845682	0.987463	0.973843	0.987529
64	$F_{D_D}(2)$	0.081948	0.993194	0.945215	0.992714	0.991970	0.992041
64	$F_{D_D}(3)$	0.114672	0.995471	0.948736	0.993512	0.995285	0.993926
64	CTD_α	52.725242	**2.226043**	37.905897	2.863392	**2.371733**	**2.765151**

Table 3.5: **VRRs of Weighted Arithmetic Average Asian Option ($h_2(\boldsymbol{S})$) with decreasing weights on various PGMs**

d	K	STD	PCA	BB	LT	FP	FC
16	90	1134.90	5480.53	2967.61	26506.31	25571.58	29849.25
16	100	138.89	1586.89	867.78	15981.52	12151.16	**16866.35**
16	110	43.24	434.65	160.59	2693.42	2104.61	2958.46
64	90	767.28	6021.82	5198.30	20489.94	18734.41	20443.36
64	100	43.04	853.26	483.45	14884.76	10404.32	**15562.73**
64	110	4.49	103.71	53.64	1612.56	702.49	1613.44

Table 3.6: **Cumulative Mass Function and Conditional Tail Dimension of $h_2(\boldsymbol{S})$ with decreasing weights, $K = 100$ and $\alpha = 95\%$**

d	CMF	STD	PCA	BB	LT	FP	FC
Truncation Dimension(D_T)							
16	$F_{D_T}(1)$	0.331484	0.811373	0.534034	0.997988	0.911462	0.999412
16	$F_{D_T}(2)$	0.518053	0.924270	0.787230	0.998371	0.992173	0.999412
16	$F_{D_T}(3)$	0.648316	0.960244	0.904404	0.999080	0.998897	0.999412
16	CTD_α	10.895650	4.994556	7.060890	1.171092	2.182018	**1.075775**
64	$F_{D_T}(1)$	0.137104	0.775593	0.492752	0.998350	0.809732	0.999047
64	$F_{D_T}(2)$	0.231132	0.895025	0.747383	0.998599	0.976601	0.999398
64	$F_{D_T}(3)$	0.306772	0.934386	0.871662	0.998712	0.996363	0.999816
64	CTD_α	41.028551	9.281665	17.915896	1.168267	2.567660	**1.051791**
Delta Dimension(D_D)							
16	$F_{D_D}(1)$	0.404419	0.845702	0.604571	0.974229	0.902069	0.974721
16	$F_{D_D}(2)$	0.578288	0.937609	0.826404	0.979261	0.978685	0.984448
16	$F_{D_D}(3)$	0.691687	0.965889	0.919691	0.986342	0.989646	0.988891
16	CTD_α	10.897150	4.991014	7.008862	3.306013	3.106338	**2.786054**
64	$F_{D_D}(1)$	0.208022	0.817471	0.569521	0.972525	0.821409	0.973138
64	$F_{D_D}(2)$	0.308536	0.913212	0.792957	0.983175	0.960875	0.982737
64	$F_{D_D}(3)$	0.382861	0.944580	0.892412	0.985528	0.983089	0.987540
64	CTD_α	41.459278	8.989668	18.029935	4.822205	4.567622	**4.021737**

By assuming $d \in \{16, 64\}$ and $K \in \{90, 100, 110\}$, the VRRs are depicted in Tables 3.3 and 3.5 for the options with increasing and decreasing weights, respectively. Similarly the CMFs for the truncation dimension variable are shown in Tables 3.4 and 3.6 for the respective cases with $K = 100$. We draw the following remarks based on these results:

- Similar to the last example, all the QMC-based PGMs are more efficient than the

classical MC methods and that STD is the least efficient.

- While the QMC-based PGMs are efficient, it is of interest to note that the weights can have a non-trivial effect on these PGMs. In particular, the efficiency of PCA and BB declines as the weight changes from increasing to decreasing. In contrast, PGMs such as STD, LT, FP and FC exhibit an increased in performance for the same change in the weights (at least for at-the-money and in-the-money options).

- The FC method appears to be the most efficient in most cases, regardless of the weighting functions.

- The dimension distribution is also affected by the weighting function. While the method of STD is the least effective at dimension reduction, its effective dimension is lower with decreasing weights. On the other hand, the dimension reduction is less effective under the decreasing weighting function for both PCA and BB. This may in part account for the decline in their efficiency as we change from increasing weights to decreasing weights. In contrast, in term of the dimension distribution, LT, FP and FC appear to be less affected by the weights. In most cases, their effective dimensions reduce to one.

- The superposition dimension has CMF(1) close to 1 and hence their CTDs are also close to 1. The truncation and delta CTDs both tell us that PCA,LT,FP and FC outperform STD and BB when the weights are increasing and LT,FP and FC outperform others when the weights are decreasing.

- When the weights are increasing, both truncation CTDs prefer PCA, LT and FC, however, delta CTDs prefer PCA,FP and FC although there is no significant evidence which PGM among PCA,LT,FP and FC is the best in the Table 3.3. It is still interesting to see that truncation dimension prefers LT comparing to PCA when they have similar VRRs.

- When the weights are decreasing, both truncation and delta CTDs prefer FC, which

is consistent with the VRR Table 3.5. It is also interesting to see that delta dimension prefers PCA comparing to LT when they have similar VRRs.

- **Note: comparing to the LT method, the CMF(3) and CTD of FC method is smaller under both truncation sense or delta sense, and hence, FC has smaller tail dimension than LT.**

3.4.3 Options depending on the average of distances from the current price

We introduce call options that depend on the average of distances from the current underlying price. Their payoffs are given by

$$h_3(\boldsymbol{S}) = \max\left(\frac{1}{d}\sum_{i=1}^{d}|S_i - S_0| - K, 0\right).$$

This option is of interest due to its non-differentiability and hence method such as the LT cannot be applied.

For the FP method, we design $\psi(\boldsymbol{f}) = \max\left(\sum_{i=1}^{d}|f_i| - K, 0\right)$ with $f_i = \frac{1}{d}(S_i - S_0)$. For the FC, we set $f_1 = \frac{1}{d}(S_d - S_0)$, $f_2 = \frac{1}{d}(S_{d/2} - S_0)$, up to f_d, which follows the order of Brownian motion. In this case, we equivalently use the functional BB method.

Table 3.7: **VRRs of $(h_3(\boldsymbol{S}))$ on various PGMs**

d	K	STD	**PCA**	BB	FP	FC
16	0	31.79	1534.34	849.18	1728.70	1108.88
16	10	43.77	**2320.29**	535.06	**2774.19**	1821.79
16	20	16.48	1457.71	200.19	1465.40	1400.23
64	0	8.26	1987.36	1239.02	2118.27	1474.87
64	10	18.60	**2179.21**	484.14	**2234.62**	997.16
64	20	4.65	1069.29	139.98	1250.37	1149.24

Table 3.8: **Cumulative Mass Function and Conditional Tail Dimension of $h_3(\boldsymbol{S})$ with $K = 10$ and $\alpha = 95\%$**

d	CMF	STD	PCA	BB	FP	FC
Truncation Dimension(D_T)						
16	$F_{D_T}(1)$	0.072142	0.985949	0.695662	0.981837	0.991169
16	$F_{D_T}(2)$	0.158394	0.996444	0.911486	0.996475	0.992986
16	$F_{D_T}(3)$	0.260861	0.998475	0.932838	0.998416	0.993185
16	CTD_α	13.621666	**1.438138**	10.904632	**1.521332**	1.694619
64	$F_{D_T}(1)$	0.010905	0.986334	0.661553	0.981544	0.990931
64	$F_{D_T}(2)$	0.026074	0.996882	0.901062	0.996746	0.993035
64	$F_{D_T}(3)$	0.043904	0.998922	0.924635	0.998868	0.993192
64	CTD_α	51.522081	**1.381330**	37.985794	**1.487865**	1.778738
Delta Dimension(D_D)						
16	$F_{D_D}(1)$	0.042682	0.960698	0.692985	0.954831	0.970577
16	$F_{D_D}(2)$	0.147527	0.978197	0.905910	0.977762	0.978928
16	$F_{D_D}(3)$	0.268306	0.984560	0.931679	0.984206	0.980564
16	CTD_α	13.475805	**3.521117**	10.785763	**3.688295**	3.832335
64	$F_{D_D}(1)$	0.000648	0.956501	0.660790	0.950375	0.966746
64	$F_{D_D}(2)$	0.007178	0.974807	0.892191	0.974182	0.974742
64	$F_{D_D}(3)$	0.020799	0.981433	0.920343	0.980867	0.976437
64	CTD_α	51.597895	**5.716308**	37.942458	**5.927360**	6.750866

By assuming $d \in \{16, 64\}$ and $K \in \{0, 10, 20\}$, the results are depicted in Tables 3.7 and 3.8 for the VRRs and the CMF, respectively. Despite the non-differentiability of the options, PCA, FP and FC are very competitive efficient, though FP is marginally more efficient. It is interesting to see that both truncation and delta CTDs suggest to use PCA and FP methods for $K = 100$ which is consistent with the VRR Table 3.7. Note that the superposition CTD is either close to 1 or becomes computationally burdensome to compute. See Algorithm 2.4.2. However according to Remark 2.4.1, we could use the CMF of D_T as the lower bound of the CMF of D_S.

3.4.4 Options depending on the weighted average of distances from the current price

The next examples are weighted version of the previous options. The payoff is given by

$$h_4(\boldsymbol{S}) = \max\left(\sum_{i=1}^{d} w_i|S_i - S_0| - K, 0\right).$$

As in the second example, increasing and decreasing weights are similarly considered.

For FP method, we design $\psi(\boldsymbol{f}) = \max(\sum_{i=1}^{d} w_i|f_i| - K, 0)$ with $f_i = w_i(S_i - S_0)$ For FC method, we design $f_1 = 2_d(S_d - S_0)$,$f_2 = x_{d/2}(S_{d/2} - S_0)$,and so on, which follows the order of Brownian bridge.

Table 3.9: **VRRs of $(h_4(\boldsymbol{S}))$ with increasing weights on various PGMs**

d	K	STD	PCA	BB	FP	FC
16	0	22.34	2146.58	989.36	2946.04	2620.01
16	10	43.33	**4466.03**	1027.15	**4824.06**	**5122.60**
16	20	20.14	2976.58	359.22	2738.42	2945.61
64	0	8.47	2531.32	1275.78	3500.13	2181.50
64	10	8.88	**2695.50**	583.67	**2728.13**	**2053.75**
64	20	4.73	2215.31	302.61	1919.12	2021.63

Table 3.10: **Cumulative Mass Function and Conditional Tail Dimension of $h_4(\boldsymbol{S})$ with increasing weights, $K = 10$ and $\alpha = 95\%$**

d	CMF	STD	PCA	BB	FP	FC
Truncation Dimension(D_T)						
16	$F_{D_T}(1)$	0.049632	0.988290	0.799296	0.975586	0.996568
16	$F_{D_T}(2)$	0.106920	0.999331	0.926690	0.996458	0.997309
16	$F_{D_T}(3)$	0.180660	0.999758	0.929526	0.998886	0.998211
16	CTD_α	14.262948	**1.267263**	11.293142	**1.622831**	**1.222979**
64	$F_{D_T}(1)$	0.005636	0.989635	0.763304	0.977070	0.996372
64	$F_{D_T}(2)$	0.016876	0.999408	0.912535	0.996310	0.997104
64	$F_{D_T}(3)$	0.028996	0.999785	0.915866	0.998498	0.997855
64	CTD_α	54.291382	**1.230051**	40.230670	**1.580066**	**1.306913**
Delta Dimension(D_D)						
16	$F_{D_D}(1)$	0.037201	0.982125	0.823560	0.969869	0.987278
16	$F_{D_D}(2)$	0.116203	0.991921	0.938689	0.991209	0.992915
16	$F_{D_D}(3)$	0.209725	0.994667	0.942456	0.993976	0.993966
16	CTD_α	14.125204	**1.985574**	11.096306	**2.466742**	**2.029439**
64	$F_{D_D}(1)$	0.001367	0.981782	0.797710	0.969237	0.986812
64	$F_{D_D}(2)$	0.008901	0.991637	0.928472	0.990729	0.992031
64	$F_{D_D}(3)$	0.021262	0.994324	0.932633	0.993512	0.993260
64	CTD_α	53.883849	**2.482498**	39.275251	**3.486524**	**2.910098**

Table 3.11: **VRR of $(h_4(\boldsymbol{S}))$ with decreasing weights on various PGMs**

d	K	STD	PCA	BB	FP	FC
16	0	58.99	213.94	393.00	697.85	274.72
16	10	30.60	275.49	119.46	**897.93**	389.67
16	20	8.00	54.00	29.80	150.03	139.68
64	0	22.99	235.23	507.42	562.36	217.65
64	10	5.50	104.25	68.95	**361.51**	205.30
64	20	1.44	14.74	6.77	21.86	42.13

Table 3.12: **Cumulative Mass Function and Conditional Tail Dimension of $h_4(\boldsymbol{S})$ with decreasing weights, $K = 10$ and $\alpha = 95\%$**

d	CMF	STD	PCA	BB	FP	FC
Truncation Dimension(D_T)						
16	$F_{D_T}(1)$	0.189611	0.770873	0.438667	0.818438	0.972019
16	$F_{D_T}(2)$	0.354426	0.906672	0.737658	0.982591	0.977499
16	$F_{D_T}(3)$	0.497427	0.947969	0.873661	0.996506	0.983936
16	CTD_α	12.069912	5.834269	9.019824	**2.446149**	3.307101
64	$F_{D_T}(1)$	0.042133	0.715551	0.350273	0.610479	0.963672
64	$F_{D_T}(2)$	0.087492	0.874103	0.669438	0.945026	0.970698
64	$F_{D_T}(3)$	0.129178	0.922981	0.826581	0.989155	0.979647
64	CTD_α	45.982774	10.624976	32.784212	**3.304701**	4.948579
Delta Dimension(D_D)						
16	$F_{D_D}(1)$	0.093246	0.685335	0.312617	0.787437	0.929017
16	$F_{D_D}(2)$	0.273322	0.852717	0.653176	0.953908	0.946151
16	$F_{D_D}(3)$	0.435073	0.907693	0.822133	0.977395	0.956563
16	CTD_α	12.352352	8.326780	9.442277	**4.283664**	6.758462
64	$F_{D_D}(1)$	0.000454	0.548416	0.172148	0.527816	0.879478
64	$F_{D_D}(2)$	0.008358	0.742357	0.501227	0.881771	0.900565
64	$F_{D_D}(3)$	0.027148	0.811595	0.695579	0.944130	0.915012
64	CTD_α	48.711965	28.375361	35.843713	**7.921431**	19.042265

By assuming $d \in \{16, 64\}$ and $K \in \{0, 10, 20\}$, the VRRs are depicted in Tables 3.9 and 3.11 for the options with increasing and decreasing weights, respectively. Similarly the CMFs for the dimension variable are shown in Tables 3.10 and 3.12 for the respective cases with $K = 10$. We draw the following remarks based on these results:

- A key conclusion can be drawn from these results is that the non-trivial impact of the weights on the efficiency of PGMs is clearly highlighted in options of this type. More specifically, PCA, FP and FC are extremely effective for pricing options with increasing weights. For example, for the range of parameter values we considered, the VRRs for PCA range from as low of 2146.58 to as high as 4466.03. Similar VRRs for the FP are 1919.12.4 to 3500.13.
- The PCA is particularly pruned to the weight change from increasing to decreasing.
- In particular, the efficiency of PCA drops at least by 10 times to as high as 150 times.

- Similar comparison for the FP method ranging from 4 times to 88 times. Hence impact of the weights on FP is relatively less. Nevertheless, the method of FP still outperforms the other PGMs, regardless of the weights.
- When the weights are increasing, both CTDs prefer PCA,FP and FC methods; When they are decreasing, both prefer FP as the best PGM which are all consistent with Tables 3.9 and 3.11. Hence, CTDs are truly good measures of tail dimensionality when the performance of different PGMs have significant difference.

3.4.5 Single Barrier combining with high dimensionality

The fifth example considers single barrier problem. The payoff function of the Asian option with knock-out feature at the maturity is defined as

$$h_5(\boldsymbol{S}) = \max(S_R - K, 0) I_{\{S_{d/4} \leq H\}}(S), \tag{3.16}$$

where

$$S_R = \sum_{i=1}^{d} \frac{1}{d} |S_i - S_0| \tag{3.17}$$

where K is the strike price, H is the barrier.

This option is of interest because it shows that not only discontinuity but also non-differentiability will affect the efficiency of PGM. Hence, a good PGM should be able to reflect both of these issues concurrently.

In this example, we include the orthogonal transformation (OT) method in Wang and Tan (2012) with the initial decomposition matrix A_0 being the Cholesky decomposition. Wang and Tan (2012) first designed the optimal transformation method controlling the discontinuity under single barrier options. However, they does not discuss the optimal choice of A_0 for an arbitrary function, and also does not consider the function of interest in the payoff function except the barrier.

For FMIX method, we assign $f_1 = S_{d/4}$, $f_i = \frac{1}{d}(S_{i-1} - S_0)$,for $i = 2, \cdots, d/4 - 1$, and $f_i = \frac{1}{d}(S_i - S_0)$, for $i = d/4 + 1, \cdots, d$. FMIX will handle the barrier by the FC and the rest of dimensions by the FP method. In other words, we first apply FC with Cholesky decompostion of $\boldsymbol{\Sigma_f} = \boldsymbol{J\Sigma J^T}$, and get the first column $\boldsymbol{A_{f\cdot 1}}$ and the corresponding $\boldsymbol{A}_{\cdot 1} = \boldsymbol{J}^{-1}\boldsymbol{A_{f\cdot 1}}$. Next FP method is applied on the residual, i.e. $\boldsymbol{J}(\boldsymbol{\Sigma} - \boldsymbol{A}_{\cdot 1}\boldsymbol{A}_{\cdot 1}^T)\boldsymbol{J}^T$, and obtain the corresponding decomposition matrix $(\boldsymbol{A}_f)_{residual}$. This in turn allows us to determine $\boldsymbol{A}_{residual} = \boldsymbol{J}^{-1}(\boldsymbol{A}_f)_{residual}$ and finally $\boldsymbol{A} = [\boldsymbol{A}_{\cdot 1}, \boldsymbol{A}_{residual}]$.

Note: the comparison of CMFs of dimension random variables are not important here as the discontinuity has no direct relationship with the variance contribution. i.e. the larger CMFs may not give a better result for a discontinuous function. For further illustration of the effect of discontinuity, see Wang and Tan (2013).

Table 3.13: **VRRs of Modified Asian Options with Knock-out Feature at Maturity ($h_5(\boldsymbol{S})$) on various PGMs**

d	K	H	STD	PCA	BB	OT	FP	FC	FMIX
16	0	100	6.97	6.88	16.04	23.52	10.94	11.51	840.52
16	0	120	4.69	6.86	13.77	15.28	7.14	12.62	677.64
16	0	140	18.99	84.92	111.20	19.16	99.63	75.86	487.74
16	10	100	5.87	9.79	13.40	8.53	11.16	11.31	184.87
16	10	120	5.65	9.35	16.84	22.05	9.70	16.55	530.87
16	10	140	34.19	94.43	119.31	33.95	107.01	73.62	405.13
64	0	100	4.29	10.63	22.42	5.04	11.58	14.96	712.21
64	0	120	2.29	7.55	12.56	7.35	7.49	9.08	893.93
64	0	140	9.88	75.34	96.11	12.28	81.40	81.61	601.89
64	10	100	2.69	11.42	14.43	6.08	9.64	14.27	278.23
64	10	120	2.82	9.70	14.45	6.04	9.79	12.33	476.24
64	10	140	9.34	90.05	96.50	14.37	102.22	84.92	528.59

The results are shown in Table 3.13 for parameter values $d \in \{16, 64\}, K \in \{0, 10\}$, and $H \in \{120, 140\}$. The effectiveness of FMIX is clearly demonstrated in these results. This example also confirms the importance of handling both high-dimensionality and discontinuity concurrently. The performance of the remaining PGMs are comparable similar. PCA and BB, for example, attempt to increase the efficiency of QMC via dimension reduction.

Hence these methods manage to reduce the effective dimension but not able to handle the discontinuity. As shown in Wang and Tan (2013), the discontinuity of the option is critical to the efficiency of QMC. Hence under some specific cases their OT is able to address option with discontinuity. The drawback of their approach is that they did not address the high-dimensionality of the option. Hence the performance of these methods are comparable similar.

3.5 Conclusion

We proposed the delta control (DC) path generation method which is a general method to reduce the nominal dimension of any arbitrary function f. Under our new designed DC approach, we could easily find optimal PGM by combining sophisticated methods on the functional covariance matrix. We also showed that the DC method encompasses many existing PGMs as its special case and importantly it can handle high dimensionality and multiple non-differentiability at the same time.

Drawbacks of this chapter

1. The method of DC is more complicated than other dimension reduction methods such as BB and PCA.
2. We have proposed three implementations of DC methods and in practice it is not clear which of these methods is most efficient.
3. For an arbitrary problem, there is no obvious way of defining $f_i, i = 1, \ldots, k$.
4. The method of DC requires the calculation of the Jacobian matrix $\boldsymbol{J}$. This imposes additional initial overhead computational time.

Chapter 4

Severity in Quasi-Monte Carlo

Key contributions of this chapter

1. Define differentiable mimic functions to evaluate variables $\{f_1, \ldots, f_k\}$ which have unknown distributions.
2. Propose a new measure that allows us to identify the relative importance of the various sub-functions. For the g function in (4.1), this is equivalent to determining in what order a PGM should be applied to. We accomplish this objective by proposing a new measure known as *severity* of a function. The severity measure allows us to determine a picking order or a hierarchical order of the relative importance of the functions.
3. Propose a new PGM that explicitly exploits the hierarchical order of functions. It turns out that once the preferred order is determined, the flexibility of the LT method implies that an optimal generation matrix can be determined accordingly. We label the resulting PGM that integrates the LT method with the hierarchical order as the hierarchical liner transformation (HLT) method.
4. Some justifications are provided to demonstrate the importance of severity for multiple non-differentiable or discontinuous functions.

4.1 Introduction

When using a PGM to simulate trajectory of stock prices at time periods $1, 2, \ldots, d$, there is an implicit ordering of how stock prices are being simulated. For example, the classical STD method, with its path generation matrix derived from the Cholesky decomposition, simulates the stock prices given S_0 in the order $\{\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_d\}$. On the other hand, the PGM based on BB generates stock prices in the following order $\{\hat{S}_d, \hat{S}_{d/2}, \hat{S}_{3d/4}, \ldots, \hat{S}_1\}$, assuming d is a power of 2. Similarly the LT method constructs the path generation matrix by optimizing its columns starting from first column, then second column, and iteratively until the d-th column. A natural question to consider is that if there exist a better "ordering" for these PGMs? More specifically, is it always optimal to first simulate the terminal stock price for the BB method? Similarly, for the LT method is it always optimal to optimize the columns starting from the first column? The "optimal" ordering of the BB construction has been addressed in Lin and Wang (2008) who demonstrate that under their prescribed optimality, it is never optimal to first simulate the terminal stock price. In fact they formally establish that the first optimal stock price to be simulated is $\hat{S}_{3d/4}$.

In this chapter we similarly investigate the "preferred" ordering of applying a PGM. We are still interested in estimating $\mathbb{E}[g(\boldsymbol{x})]$ where $\boldsymbol{x} \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$. However, we are more interested in function g that has a more complicated structure. An example of interest is of the following form:

$$g(\boldsymbol{x}) = f_1(\boldsymbol{x}) \times f_2(\boldsymbol{x}) \times f_3(\boldsymbol{x}) = \mathbb{I}_{\{g_1(\boldsymbol{x}) \leq H_1\}} \times \max(g_2(\boldsymbol{x}), H_2) \times g_3(\boldsymbol{x})), \tag{4.1}$$

where H_1, H_2 are constants, g_1, g_2, and g_3 are continuous functions and $f_1(\boldsymbol{x}) = \mathbb{I}_{\{g_1(\boldsymbol{x}) \leq H_1\}}$, $f_2(\boldsymbol{x}) = \max(g_2(\boldsymbol{x}), H_2)$ and $f_3(\boldsymbol{x}) = g_3(\boldsymbol{x})$.

There are at least three reasons that motivated us to consider (4.1):

1. Derivative securities with their payoff functions similar to that in (4.1) are quite

common. Some notable examples are the knock-in and knock-out options.

2. The discontinuity induced by the indicator function increases the difficulty for the QMC-based PGM considerably. The numerical results in Wang and Tan (2013) show that the nature of the discontinuity has a tremendous effect on the efficiency of QMC. While the orthogonal transformation (OT) proposed by Wang and Tan (2013) is able to recover the high precision of QMC even with the discontinuity, there are two issues with their method.

   (a) The OT method applies only to discontinuity with very specified structure and with only one discontinuity.

   (b) The OT method focuses only on discontinuity re-alignment; i.e. if focuses only on f_1 while ignoring f_2 and f_3. As argued in Imai and Tan (2014), the high-dimensionality of f_2 and f_3 can severely affect the overall effectiveness of the QMC-based PGM. Hence it is important to have a PGM that has the capability of handling both discontinuity and high-dimensionality concurrently.

3. If g is continuous, we postulate that the efficiency gain from using the "optimal ordering" of a PGM (apart from STD) is negligible. The numerical results conducted in Lin and Wang (2008) appear to support this assertion. While the "optimal" BB of Lin and Wang (2008) outperforms the standard BB, the efficiency gains as shown in their numerical examples, however, are quite minimum.

In summary, the following describes the two objectives of this chapter are as follows:

1. Propose a new measure that allows us to identify the relative importance of the various sub-functions. For the g function in (4.1), this is equivalent to determining in what order a PGM should be applied to. In other words, should a PGM be first applied to f_1, f_2, or f_3? We accomplish this objective by proposing a new measure known as *severity* of a function. The severity measure allows us to determine a picking order or a hierarchical order of the relative importance of the functions.

Formal definition of severity will be provided in the next section.

2. Propose a new PGM that explicitly exploits the hierarchical order of functions. It turns out that once the preferred order is determined, the flexibility of the LT method implies that an optimal generation matrix can be determined accordingly. We label the resulting PGM that integrates the LT method with the hierarchical order as the hierarchical liner transformation (HLT) method. HLT is presented in Section 4.3 and the numerical illustration is given in Section 4.4

4.2 Severity

We begin the section by re-expressing the function $g(\boldsymbol{x})$ as follows:

$$g(\boldsymbol{x}) = \psi(f_1(\boldsymbol{x}), \ldots, f_k(\boldsymbol{x})), \tag{4.2}$$

where $k \leq n$, and $\psi : \mathbb{R}^k \mapsto \mathbb{R}$. We implicitly assume that each component $f_i, i = 1, \ldots, k$ should be separately handled due to, for instance, discontinuity or non-differentiability. A sample of $\psi_f(\cdot)$ is of the form $\psi(f_1, f_2, f_3) = f_1 f_2 f_3$, as in (4.1). Note that the function ψ is differentiable with respect to each component f_i. We also assume the existence of expectation and variance of each component.

In what follows, we assume that a given random vector $\boldsymbol{f} = (f_1, \ldots, f_k)$ can further be classified into three groups $(F_1, \boldsymbol{F}_2, \boldsymbol{F}_3) := \boldsymbol{F}$, where $F_1 \in \{f_1, \ldots, f_k\}$ is a random variable, $\boldsymbol{F}_2 \in \{(f_1, \ldots, f_k) \backslash F_1\}$ is a random vector of k_2 dimension, and $\boldsymbol{F}_3$ is the residual random vector of dimension $k - 1 - k_2$. For a given vector $\boldsymbol{X}$, we denote $dim(X)$ as its dimension so that $dim(\boldsymbol{F}_2) = k_2$ and $dim(\boldsymbol{F}) = k$.

For a given permutation $\boldsymbol{F} = (F_1, \boldsymbol{F}_2, \boldsymbol{F}_3)$ from $\boldsymbol{f} = (f_1, \ldots, f_k)$, we define ξ_f with respect

to $(F_1, \boldsymbol{F}_2, \boldsymbol{F}_3)$ as follows:

$$\xi_f\left(F_1, \boldsymbol{F}_2, \boldsymbol{F}_3\right) = \mathrm{Var}_{(F_1, \boldsymbol{F}_2)}\left(\mathbb{E}_{\boldsymbol{F}_3|(F_1, \boldsymbol{F}_2)}\left[\psi\left(\boldsymbol{F}\right)\right]\right), \tag{4.3}$$

where $\mathrm{Var}_{(F_1, \boldsymbol{F}_2)}$ represents the variance with respect to the distribution for the random vector $(F_1, \boldsymbol{F}_2)$, while $\mathbb{E}_{\boldsymbol{F}_3|(F_1, \boldsymbol{F}_2)}$ represents the conditional expectation with respect to the conditional distribution for $\boldsymbol{F}_3\,|(F_1, \boldsymbol{F}_2)$.

Note that since $\psi(\boldsymbol{f}) = \psi(F_1, \boldsymbol{F}_2, \boldsymbol{F}_3)$ and from the low of total variance, we have

$$\begin{aligned}
\mathrm{Var}(\psi(F_1, \boldsymbol{F}_2, \boldsymbol{F}_3)) &= \mathrm{Var}_{(F_1, \boldsymbol{F}_2)}\left(\mathbb{E}_{\boldsymbol{F}_3|(F_1, \boldsymbol{F}_2)}\left[\psi\left(\boldsymbol{F}\right)\right]\right) + \mathbb{E}_{(F_1, \boldsymbol{F}_2)}\left(\mathrm{Var}_{\boldsymbol{F}_3|(F_1, \boldsymbol{F}_2)}\left[\psi\left(\boldsymbol{F}\right)\right]\right) \\
&= \xi_f\left(F_1, \boldsymbol{F}_2, \boldsymbol{F}_3\right) + \mathbb{E}_{(F_1, \boldsymbol{F}_2)}\left(\mathrm{Var}_{\boldsymbol{F}_3|(F_1, \boldsymbol{F}_2)}\left[\psi\left(\boldsymbol{F}\right)\right]\right).
\end{aligned}$$

Due to the effectiveness of the dimension reduction, we expect that conditional variance $Var_{\boldsymbol{F}_3|(F_1, \boldsymbol{F}_2)}\left[\psi\left(\boldsymbol{F}\right)\right]$ to be as small as possible. Hence the quantity $\xi_f\left(F_1, \boldsymbol{F}_2, \boldsymbol{F}_3\right)$ will be the dominating term in the above variance decomposition. For this reason, we focus on $\xi_f\left(F_1, \boldsymbol{F}_2, \boldsymbol{F}_3\right)$ and the notion of severity will be defined in terms of this quantity.

We now introduce our definition of severity in terms of the function ξ_f.

Definition 4.2.1. *The level-1 of severity for each component $f_i, i = 1, \ldots, k$ in the function ψ, denoted by $SV^1(f_i)$, is defined as follows:*

$$SV^1\left(f_i\right) = \xi_f(F_1, \boldsymbol{F}_2, \boldsymbol{F}_3), \tag{4.4}$$

where $F_1 = f_i, \boldsymbol{F}_2 = \emptyset, \boldsymbol{F}_3 = \{f_1, \ldots, f_{i-1}, f_{i+1}, \ldots, f_k\}$.

Intuitively, the level-1 severity extracts the effect of the function f_i on the variance of the function ψ. Hence, the larger the value of its severity, the larger its impact for estimating the expectation of ψ. Accordingly, the first hierarchical severity order of the function ψ is optimally chosen as the one that yields the greatest level-1 severity. Let $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)$ represents the optimal hierarchical severity order of the function ψ so that π_i denotes the

i-th highest hierarchical severity order. Thus, the optimal first hierarchical severity order π_1 is obtained as follows:

$$\pi_1 := \arg\max_{i=1,\ldots,k} SV^1(f_i). \tag{4.5}$$

We now proceed to the second hierarchical severity order of the function ψ. This requires us to define the level-2 severity, conditioning on the optimal level-1 severity.

Definition 4.2.2. *The level-2 of severity for component $f_i, i \in \{1,\ldots,k\} \setminus \{\pi_1\}$ in the function ψ given the first component f_{π_1} , denoted by SV^2, is defined as follows:*

$$SV^2\left(f_i \,|f_{\pi_1}\right) = \xi_f(F_1, \boldsymbol{F}_2, \boldsymbol{F}_3), \tag{4.6}$$

with $F_1 = f_i, \boldsymbol{F}_2 = f_{\pi_1}$, and $\boldsymbol{F}_3$ are the residual $(k-2)$ functions.

From the level-2 severity, the second optimal hierarchical severity order π_2 is then determined by

$$\pi_2 := \arg\max_{i\in\{1,\ldots,k\}\setminus\{\pi_1\}} SV^2\left(f_i \,|f_{\pi_1}\right). \tag{4.7}$$

Analogously, m-th hierarchical severity order can be similarly determined optimally as the maximum of all level-m severity. In other words, we have

Definition 4.2.3. *Given the $m-1$ components $\pi_1,\ldots,\pi_{m-1}$, the level-m of severity for component $f_i, i \in \{1,\ldots,k\} \setminus \{\pi_1,\ldots,\pi_{m-1}\}$ in the function ψ_f , denoted by SV^m, is defined as follows*

$$SV^m\left(f_i \,|f_{\pi_r}, r = 1,\ldots,m-1\right) = \xi_f(F_1, \boldsymbol{F}_2, \boldsymbol{F}_3), \tag{4.8}$$

with $F_1 = f_i, \boldsymbol{F}_2 = f_{\pi_1,\ldots,\pi_{m-1}}$, and $\boldsymbol{F}_3$ is the residual $k-m$ functions .

The m-th hierarchical severity order is then determined as

$$\pi_m := \arg\max_{i\in\{1,\ldots,k\}\setminus\{\pi_1,\ldots,\pi_{m-1}\}} SV^m\left(f_i \,|f_{\pi_r}, r = 1,\ldots,m-1\right). \tag{4.9}$$

4.2.1 Approximating ξ

In general, evaluating the function ξ is very difficult and in most cases the distributions of the functions $\{f_1, f_2, \ldots, f_k\}$ are unknown. Furthermore, in many cases the function f_i is discontinuous. To facilitate the calculation of ξ, we introduce a concept named as the *mimic function.* Formally the mimic function is defined as follows:

Definition 4.2.4. *A function h is called a mimic function of f if the following conditions hold.*

1. *h is continuous and differentiable with respect to $\boldsymbol{x}$,*
2. *$f|h$ is constant.*

Going back to the functions f_1, f_2, and f_3 defined in (4.1), then possible mimic functions for f_1, f_2, and f_3 are $h_1(\boldsymbol{x}) = g_1(\boldsymbol{x}), h_2(\boldsymbol{x}) = g_2(\boldsymbol{x})$, and $h_3(\boldsymbol{x}) = g_3(\boldsymbol{x})$. Note that under the assumption that $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$, then the mimic function is a continuous differentiable function in terms of the normal random variables. **Note: The severity only measures the importance of general functions $\{f_1, \cdots, f_k\}$. $\{g_1, \cdots, g_k\}$ are the mimic functions of $\{f_1, \cdots, f_k\}$ where $\{g_1, \cdots, g_k\}$ are differentiable functions and can be used to construct PGM.**

In what follows, we assume that $\psi(\boldsymbol{x})$ in ((4.2)) can be approximated using the mimic functions. In other words, we have

$$\psi(f_1, \cdots, f_k) \approx \psi_h(h_1(\boldsymbol{x}), \ldots, h_k(\boldsymbol{x})), \tag{4.10}$$

where h_i is a mimic function of f_i, for $i = 1, \ldots, k$.

To proceed, it is useful to recall the following properties associated with the normal random vectors. First, any normal vector $\boldsymbol{x}$ can be expressed as $\boldsymbol{x} = \boldsymbol{A}^{ch}\boldsymbol{z}$ where $\boldsymbol{A}^{ch}$ denotes the Cholesky decomposition of the covariance matrix Σ_x, i.e., $\Sigma_x = \boldsymbol{A}^{ch}\boldsymbol{A}^{ch^{\mathrm{T}}}$ and $\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{I})$. Second, since $\boldsymbol{A}^{ch}$ is a lower triangular matrix, the first component $h_1(\boldsymbol{x})$ can be expressed

only by z_1. In general, j-th component $h_j(\boldsymbol{x})$ can be expressed by z_i for $i = 1, \ldots, j$.

Based on these observations, we will rewrite ((4.10)) as a function of $\boldsymbol{z}$ so that it can be analyzed explicitly. Let $\boldsymbol{y} = (y_1, \boldsymbol{y}_2, \boldsymbol{y}_3)$ denote a permutation of the random vector $\boldsymbol{z} = (z_1, \ldots, z_k)$, where we denote y_1 as the current random variable, i.e. the random variable we would control, $\boldsymbol{y}_2$ as selected random variables, and $\boldsymbol{y}_3$ as the residual random variables.

We define ξ_y with respect to $(y_1, \boldsymbol{y}_2, \boldsymbol{y}_3)$ as follows:

$$\xi_y\left(y_1, \boldsymbol{y}_2, \boldsymbol{y}_3\right) = \mathrm{Var}_{(y_1, \boldsymbol{y}_2)}\left(\mathbb{E}_{\boldsymbol{y}_3|(y_1, \boldsymbol{y}_2)}\left[\psi\left(y_1, \boldsymbol{y}_2, \boldsymbol{y}_3\right)\right]\right). \tag{4.11}$$

The normality of the random vector $\boldsymbol{y}$ facilitates the evaluation of the above function $\xi_y\left(y_1, \boldsymbol{y}_2, \boldsymbol{y}_3\right)$.

$$\mu_{\boldsymbol{y}_3}\left(y_1, \boldsymbol{y}_2\right) := \mathbb{E}_{\boldsymbol{y}_3|(y_1, \boldsymbol{y}_2)}\left[\psi\left(\boldsymbol{y}\right)\right] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \psi_f\left(f_1\left(\boldsymbol{A}^{ch}\tilde{\boldsymbol{y}}\right), \ldots, f_k\left(\boldsymbol{A}^{ch}\tilde{\boldsymbol{y}}\right)\right) \phi_{\boldsymbol{y}_3}\left(\boldsymbol{y}_3\right) d\boldsymbol{y}_3, \tag{4.12}$$

where $\tilde{\boldsymbol{y}}^{\mathrm{T}} = (y_1, \boldsymbol{y}_2, \boldsymbol{y}_3)$, and $\phi_{\boldsymbol{y}_3}\left(\boldsymbol{y}_3\right)$ is given by $dim(\boldsymbol{y}_3)$ products of density functions of the standard normal, that is,

$$\phi_{\boldsymbol{y}_3}\left(\boldsymbol{y}_3\right) = \phi\left(y_3^1\right) \cdots \phi\left(y_3^{dim(\boldsymbol{y}_3)}\right),$$

where y_3^j is the j^{th} element of the vector $\boldsymbol{y}_3$, and $d\boldsymbol{y}_3 = dy_3^1 \cdots dy_3^{dim(\boldsymbol{y}_3)}$.

Using eqs. ((4.11)) and ((4.12)), we can write

$$\xi_y(\boldsymbol{y}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\mu_{\boldsymbol{y}_3}\left(y_1, \boldsymbol{y}_2\right) - \mathbb{E}_{(y_1, \boldsymbol{y}_2)}\left[\mu_{\boldsymbol{y}_3}\left(y_1, \boldsymbol{y}_2\right)\right]\right)^2 \phi_{(y_1, \boldsymbol{y}_2)}\left(y_1, \boldsymbol{y}_2\right) d\boldsymbol{y}_{(y_1, \boldsymbol{y}_2)}, \tag{4.13}$$

and

$$\mathbb{E}_{(y_1, \boldsymbol{y}_2)}\left[\mu_{\boldsymbol{y}_3}\left(y_1, \boldsymbol{y}_2\right)\right] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mu_{\boldsymbol{y}_3}\left(y_1, \boldsymbol{y}_2\right) \phi_{(y_1, \boldsymbol{y}_2)}\left(y_1, \boldsymbol{y}_2\right) d\boldsymbol{y}_{(y_1, \boldsymbol{y}_2)}, \tag{4.14}$$

where

$$\phi_{y_1,\boldsymbol{y}_2}(y_1,\boldsymbol{y}_2)=\phi(y_1)\phi\left(y_2^1\right)\cdots\phi\left(y_2^{dim(\boldsymbol{y}_2)}\right),$$

and

$$d\boldsymbol{y}_{(y_1,\boldsymbol{y}_2)}=dy_1dy_2^1\cdots dy_2^{dim(\boldsymbol{y}_2)}.$$

In practice the multiple integrals in (4.13) and (4.14) can be evaluated using standard MC methods.

Note that in order to exploit the idea of severity as advocated in this chapter, we need an additional pre-computational effort for identifying the function's hierarchical severity order. For practical application, it is anticipated that the additional computational cost could be reduced for the following reasons:

1. First is that the severity values do not need to be calculated with high precision. We are only concerned with the ordering, rather than their individual values. This implies that the severity value can be approximated quickly using a small simulation sample such as $N_s = 1024$, as in our numerical examples to be presented at the end of this chapter.

2. If k is large, then it may be quite costly to obtain the entire hierarchical severity orders $\pi_1, \pi_2, \ldots, \pi_k$. In view of the effectiveness of the dimension reduction methods, in practice it is sufficient to deduce the severity up to second or third level. The efficiency gain from having higher level of severity is of diminishing return. Note that if we were to compute up to severity l, the floating point operations of computation is $O(N_s\frac{k!}{(k-l)!})$.

3. If the additional computational overhead is still a concern, then the following approximation may be used. The idea behind the proposed approximation is as follows. Instead of calculating the exact expectation in ((4.12)), we replace it by a single value evaluated at some representative point. More precisely, an approximated function

can be defined as

$$\tilde{\mu}_{\boldsymbol{y}_3}(y_1, \boldsymbol{y}_2) = \psi_f\left(f_1\left(\boldsymbol{A}^{ch}\hat{\boldsymbol{y}}\right), \ldots, f_k\left(\boldsymbol{A}^{ch}\hat{\boldsymbol{y}}\right)\right) \tag{4.15}$$

where $\hat{\boldsymbol{y}}$ is given by $\hat{\boldsymbol{y}}^{\mathrm{T}} = (y_1, \boldsymbol{y}_2, \boldsymbol{y}_3')$ with $\boldsymbol{y}_3' \equiv \boldsymbol{0}$ under condition that $\boldsymbol{y}_3$ is under the standard normal law with mean zero. Since $dim(\boldsymbol{y}_3)$ is much larger than $dim(\boldsymbol{y}_2)$ in determining the low-level of severity, this approximation significantly decreases the computational burden, though it leads to some bias. Accordingly, the function ξ_y in eq.((4.11)) is replaced by $\tilde{\xi}_y$ which is given by

$$\tilde{\xi}_y(\boldsymbol{y}) = \mathrm{Var}_{(y_1, \boldsymbol{y}_2)}\left(\tilde{\mu}_{\boldsymbol{y}_3}(y_1, \boldsymbol{y}_2)\right). \tag{4.16}$$

In our numerical results, we will still use ((4.13)), instead of the approximation ((4.15)).

4.3 Hierarchical Linear Transformation (HLT)

Recall that the LT method of Imai and Tan (2006) achieves the dimension reduction by seeking a generation matrix that optimally minimizes the truncation dimension of the function. The key assumptions of this method are the differentiability of the function and the normality of the underlying state variables. For non-Gaussian applications, see the Generalized Linear Transformation (GLT) of Imai and Tan (2009).

While the original implementation of the LT method implies that the columns of the generation matrix $\boldsymbol{A}$ are determined iteratively starting from column 1 until column d, in theory the columns of the generation matrix can be optimized in any arbitrary order (after a suitable permutation). This feature suggests that the LT method can be implemented in such a way that reflects the hierarchical severity order of a function determined from the last section. To distinguish from the original implementation of the LT method that produces $\boldsymbol{A}$, we use the notation $\boldsymbol{A}_\pi$ to denote the generation matrix derived as a result

of the hierarchical severity order $\boldsymbol{\pi}$. We refer to the resulting PGM as the Hierarchical Linear Transformation (HLT).

Once the hierarchical order of severity is found, the HLT method applies the LT method to each component of the function. We now describe in greater details the implementation of HLT. Let us assume that $\boldsymbol{x}$ is a normal vector, $\boldsymbol{x} \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$ and let $\boldsymbol{C}^{chol}$ be a decomposed matrix of the covariance matrix Σ. Then we can write $\boldsymbol{x} = \boldsymbol{C}^{chol}\boldsymbol{\varepsilon}$ where $\boldsymbol{\varepsilon}$ stands for a standard normal vector, i.e., $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_d)^{\mathrm{T}} \sim N_d(\mathbf{0}, \boldsymbol{I})$.

With a slight abuse of notation, we denote by $f_{\pi(i)}(\boldsymbol{\varepsilon})$ the component with the i-th largest severity. The LT method essentially proposes the optimal orthogonal matrix $\boldsymbol{A}^*$ and replace $\boldsymbol{\varepsilon}$ by $\boldsymbol{A}^*\boldsymbol{\varepsilon}$ for the LT path-generation, i.e, $f_{\pi(i)}(\boldsymbol{A}^*\boldsymbol{\varepsilon})$.

In the HLT method, we first select the component $f_{\pi(1)}$ that has the largest value of severity. By applying first-order Taylor expansion to the mimic function $g_{\pi(1)}$ at arbitrary point $\boldsymbol{\varepsilon} = \hat{\boldsymbol{\varepsilon}} + \Delta\boldsymbol{\varepsilon}$:

$$g_{\pi(1)}(\boldsymbol{\varepsilon}) \approx g_{\pi(1)}(\hat{\boldsymbol{\varepsilon}}) + \sum_{l=1}^{d} \left.\frac{\partial g_{\pi(1)}}{\partial \varepsilon_l}\right|_{\varepsilon=\hat{\varepsilon}} \Delta\varepsilon_l \tag{4.17}$$

The candidate vector $\boldsymbol{b}_1$ of the 1st column of the orthogonal matrix, denoted by $\boldsymbol{A}^*_{\cdot 1}$, can be given by

$$\boldsymbol{b}_1^{\mathrm{T}} = \left(\frac{\partial g_{\pi(1)}}{\partial \varepsilon_1}, \ldots, \frac{\partial g_{\pi(1)}}{\partial \varepsilon_d}\right). \tag{4.18}$$

In the LT method, we usually set $\hat{\boldsymbol{\varepsilon}} = \mathbf{0}$. To be more precise, the first column of the orthogonal matrix is given by

$$\boldsymbol{A}^*_{\cdot 1} = \pm\frac{\boldsymbol{b}}{|\boldsymbol{b}|},$$

where $|\boldsymbol{b}|$ denotes the norm of the vector $\boldsymbol{b}$. In the case when the vector $\boldsymbol{x}$ is under non-Gaussian distribution, it is replaced by

$$\boldsymbol{b}_1^{\mathrm{T}} = \left(\frac{\partial g_{\pi(1)}}{\partial x_1}\frac{\phi(\varepsilon_1)}{h(x_1)}, \ldots, \frac{\partial g_{\pi(1)}}{\partial x_d}\frac{\phi(\varepsilon_d)}{h(x_d)}\right), \tag{4.19}$$

where h represents a density function for x_i and ϕ is a standard normal density. For the

technical details of the LT paper, we refer the readers to Imai and Tan (2006, 2014).

In the HLT method, the j-th column can be determined based on the function $f_{\pi(j)}$ in exactly the same manner, i.e, the candidate vector for j-th column of $\boldsymbol{A}^*$ is calculated by

$$\boldsymbol{b}_j^{\mathrm{T}} = \left(\frac{\partial g_{\pi(j)}}{\partial \varepsilon_1}, \ldots, \frac{\partial g_{\pi(j)}}{\partial \varepsilon_d} \right), \tag{4.20}$$

for $j = 1, \ldots, k$.

Note that if we let $g_i = x_i$, then severity measure will determine the optimal permutation of Brownian bridge. This is because Brownian bridge is the re-ordering of Chelosky decomposition. In a more general case, e.g. Levy process, the severity measure could be used to determine the optimal order of bridges. For example, the normal inverse Gaussian process could be simulated by Inverse Gaussian bridges.

It is straightforward to construct the optimal orthogonal matrix $\boldsymbol{A}^*$. Let $\boldsymbol{B}$ be a d by d matrix that is given by

$$\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k, \cdots, \boldsymbol{b}_d) .$$

Note that j-th columns of $\boldsymbol{B}$ with $k+1 \leq j \leq d$ are arbitrary as long as it is non-singular. The optimal orthogonal matrix $\boldsymbol{A}^*$ can be obtained by applying the QR-decomposition to the matrix $\boldsymbol{B}$, that is,

$$\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{R},$$

where $\boldsymbol{Q}$ and $\boldsymbol{R}$ are an orthogonal matrix and an upper triangular matrix, respectively. We can set $\boldsymbol{A}^* = \boldsymbol{Q}$ for the HLT method.

When the function $f_{\pi(j)}$ is differentiable, it is easy to implement. It is, however, often the case when $f_{\pi(j)}$ is not differentiable or even discontinuous. Typical instances include indicator functions or max(min) operations. In this case, we proposes to use a differentiable function that mimics the original function for specifying the orthogonal matrix. For instance, in the case of our example given in ((4.1)) we replace f_1, f_2 and f_3 by g_1, g_2 and g_3 to determine the candidate vector $\boldsymbol{b}$ in ((4.20)).

In summary, the HLT contains two types of linear transformations. The first type corresponds to a permutation $\boldsymbol{A}_\pi$ that depends on the severity in the decreasing order.

The second type of transformation is accomplished after the permutation in which j-th column of the orthogonal matrix is determined based on the function $f_{\pi(j)}$, separately, i.e. the first column of $\boldsymbol{A}_\pi$ corresponds to the $f_{\pi(1)}$, the second column of $\boldsymbol{A}_\pi$ corresponds to the $f_{\pi(2)}$, and etc.

4.4 Numerical Illustrations

In this section we provide some numerical evidence supporting the importance of severity as well as its effect on the efficiency of QMC-based PGM. Subsection 4.4.1 provides additional insight on the severity ranking while Subsection 4.4.2 assesses the efficiency of the proposed HLT method relative to some other existing PGMs.

4.4.1 The Importance of Severity

Let us begin with the following example to demonstrate the importance of severity order for an efficient implementation of QMC:

$$\psi(\boldsymbol{x}) = \mathbb{I}_{\{\sum_{j=1}^{d} w_j^1 x_j \leq H_1\}} \mathbb{I}_{\{\sum_{j=1}^{d} w_j^2 x_j \leq H_2\}} \mathbb{I}_{\{\sum_{j=1}^{d} w_j^3 x_j \leq H_3\}}, \tag{4.21}$$

where $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ and for $i = 1, ..., d$, $w_i^1 = 1/d$, $w_i^2 = ci$, and $w_i^3 = ci^{-1}$, representing the cases with equal weight, increasing weight, and decreasing weight, respectively. In the latter two weights c is a normalized constant with their weights sum to one. This is an interesting function as it is notoriously challenging for QMC due to the presence of three discontinuities. By letting $f_1 = \mathbb{I}_{\{\sum_{j=1}^{d} w_j^1 x_j \leq H_1\}}$, $f_2 = \mathbb{I}_{\{\sum_{j=1}^{d} w_j^2 x_j \leq H_2\}}$, and $f_3 = \mathbb{I}_{\{\sum_{j=1}^{d} w_j^3 x_j \leq H_3\}}$, we are interested in finding out if the order of applying the PGM to each of these functions

f_1, f_2, and f_3 matters?

For a given set of ordering, we use the HLT to first determine the optimal generation matrix $\boldsymbol{A}_\pi$. The resulting generation matrix is then used to simulate $\boldsymbol{f}$ in order to estimate the value of (4.21). Note that f_1, f_2, and f_3 are discontinuous functions so that they cannot be applied to HLT directly. Instead, the generation matrix is optimally determined based on their respective mimic functions. More specifically, the HLT is implemented based on h_1, h_2, and h_3 where $h_1 = \sum_{j=1}^{d} w_j^1 x_j$, $h_2 = \sum_{j=1}^{d} w_j^2 x_j$, and $h_3 = \sum_{j=1}^{d} w_j^3 x_j$.

With only three functions f_1, f_2, and f_3, there are six possible permutations of ordering. This implies that we can exhaust all six possible permutations and compare the efficiency of the resulting estimator. To gauge the efficiency among these six possible permutations, we resort to the Variance Reduction Ratio (VRR), which is defined as $VRR(f) = \frac{Var^{MC}(g)}{Var^{QMC}(g)}$, where $Var^{MC}(g)$ means the variance of g under MC estimate and $Var^{QMC}(g)$ means the variance of g under QMC-based HLT estimate. The resulting VRRs for $d \in \{16, 64\}$ and $H_1, H_2, H_3 \in \{-1, 0, 1\}$ are reported in the last six columns of Table 4.1.

One immediately conclusion can be drawn from these results is that the relative efficiency of HLT is highly dependent on the ordering. Just to illustrate, let us consider the case with $d = 64, H_1 = 0, H_2 = 0$, and $H_3 = 1$. If the HLT has been applied in the order of $\{f_3, f_2, f_1\}$, then the gain is almost 18 times more efficient than the corresponding MC method. Had we chosen the order to be $\{f_1, f_2, f_3\}$ or $\{f_1, f_3, f_2\}$, we can achieve a remarkable VRR of 8553. These results clearly indicate the ordering can have a profound effect on the efficiency of PGM.

Along with the VRRs, the table also displays two additional information. One is the expected value of ψ estimated based on the MC method. The other is the hierarchical order of severity estimated based on the MC method as described in the previous section. What is of significant interest is to note that the ranking provided by the severity coincides with the VRRs' results. For the case we just considered, using ordering of either $\{f_1, f_2, f_3\}$ or $\{f_1, f_3, f_2\}$ leads to a remarkable VRR of 8553. The hierarchical order of severity

also suggests that the first ranking of $\{f_1, f_2, f_3\}$ should be selected. For all other set of parameter values, the severity ranking is also consistent with the best permutation set of ordering.

Table 4.1: The VRR for the 6 possible permutations based on (4.21)

d	H_1	H_2	H_3	$\mathbb{E}(\psi)$ (MC)	Severity Ranking	[1,2,3]	[1,3,2]	[2,1,3]	[2,3,1]	[3,1,2]	[3,2,1]
16	-1	0	1	0.0481	[1,2,3]	807.56	807.56	50.47	27.32	33.90	25.07
16	-1	1	0	0.0484	[1,2,3]	836.48	836.48	44.97	31.32	35.71	25.84
16	0	-1	1	0.0849	[2,1,3]	79.43	26.77	1349.76	1349.76	19.99	43.39
16	0	1	-1	0.0054	[3,1,2]	6.27	12.17	4.27	9.97	92.05	92.05
16	1	-1	0	0.0854	[2,1,3]	62.08	24.23	791.87	657.01	24.14	45.89
16	1	0	-1	0.0056	[3,1,2]	5.04	11.20	5.44	9.35	92.45	92.45
64	-1	0	1	0.0433	[1,2,3]	8552.61	8552.61	42.80	24.44	36.22	17.91
64	-1	1	0	0.0411	[1,2,3]	4410.59	4410.59	43.46	24.05	37.26	19.93
64	0	-1	1	0.0914	[2,1,3]	64.17	35.32	1333.91	1333.91	18.73	37.78
64	0	1	-1	0.0004	[3,1,2]	2.19	2.56	2.33	2.50	6.00	6.00
64	1	-1	0	0.0848	[2,3,1]	62.09	30.95	409.16	544.71	16.12	44.69
64	1	0	-1	0.0003	[3,1,2]	2.04	2.99	2.06	1.90	5.14	5.14

4.4.2 Relative Efficiency of HLT

In this subsection, we consider two additional example to demonstrate the relative efficiency of the HLT method, hence highlighting the importance of severity ranking. The first example is of the following form:

$$\psi(\boldsymbol{x}) = \max\{\sum_{j=1}^{d} w_j^1 S_j, H_1\}\mathbb{I}_{\{\sum_{j=1}^{d} w_j^2 S_j \leq H_2\}}\mathbb{I}_{\{\sum_{j=1}^{d} w_j^3 S_j \leq H_3\}} \tag{4.22}$$

where the weights are defined similarly as in the previous example. This is an option involving a double barriers. If both of these barriers are not breached over the life of the option, then the payoff of the option reduces to an Asian-style.

By considering $d \in \{16, 64\}$ and $H_1, H_2, H_3 \in \{90, 100, 110\}$, Table 4.2 provides a comparison of the VRRs for the PGMs STD, BB, PCA, and HLT. Note that for the method of

HLT, its implementation is based on the optimal severity rankings that are depicted in the 5-th column of the table.

Among the four PGMs, both STD and BB are inferior while the efficiency of PCA is highly sensitive to the parameter values. These results may not be surprising since these PGMs are not designed to handle problems with multiple discontinuities. In all cases, the superiority of the HLT is obvious, confirming the importance of the severity measure.

Table 4.2: VRRs of (4.22) based on STD, BB, PCA, and HLT

d	H_1	H_2	H_3	Severity Ranking	STD	BB	PCA	HLT
16	90	100	110	[2,1,3]	3.72	11.29	143.12	186.67
16	90	110	100	[3,2,1]	3.70	8.78	11.38	150.42
16	100	90	110	[2,1,3]	3.22	12.20	131.70	187.90
16	100	110	90	[3,1,2]	2.79	5.38	4.79	97.97
16	110	90	100	[2,3,1]	2.50	10.07	55.03	138.17
16	110	100	90	[3,1,2]	4.34	6.20	6.01	72.07
64	90	100	110	[2,1,3]	3.55	13.51	151.83	158.84
64	90	110	100	[3,2,1]	2.78	7.02	9.69	112.87
64	100	90	110	[2,1,3]	2.49	11.06	120.18	137.49
64	100	110	90	[3,1,2]	1.84	4.30	4.66	42.80
64	110	90	100	[2,3,1]	2.14	9.20	59.11	98.43
64	110	100	90	[3,1,2]	2.14	3.21	3.54	41.45

Next we consider another function with the following structure:

$$\psi(\S) = \frac{\max\{\sum_{j=1}^{d} w_j^1 S_j, H_1\}}{\max\{\sum_{j=1}^{d} w_j^2 S_j, H_1\} + \max\{\sum_{j=1}^{d} w_j^3 S_j, H_1\}}. \tag{4.23}$$

We similarly use the same set of parameter values as in the previous example. The results are displayed in Table 4.3. While (4.23) is a rather unusual function, nevertheless it can be used as a toy example to assess the relative efficiency of the various PGMs. Interestingly, the VRRs of the HLT can be as high as over 60,000 times more efficient that the MC method. This compares favourably to all other PGMs considered in Table 4.3.

Table 4.3: VRRs of (4.23) based on STD, BB, PCA, and HLT

d	H_1	H_2	H_3	Severity	STD	BB	PCA	HLT
16	90	100	110	[1,2,3]	21.44	461.94	7,260.33	57,359.28
16	90	110	100	[1,2,3]	13.93	599.84	7,263.15	53,579.61
16	100	90	110	[1,3,2]	2.92	75.80	1,350.39	11,773.56
16	100	110	90	[1,2,3]	4.78	68.64	1,974.98	7,339.18
16	110	90	100	[1,3,2]	14.55	308.99	4,497.46	61,239.65
16	110	100	90	[1,3,2]	16.78	269.55	6,497.61	39,673.45
64	90	100	110	[1,3,2]	21.89	530.61	8,814.78	47,169.85
64	90	110	100	[1,2,3]	14.55	644.04	5,543.73	49,640.41
64	100	90	110	[1,2,3]	4.54	104.24	1,483.45	9,185.10
64	100	110	90	[2,1,3]	4.73	137.05	4,064.18	9,615.43
64	110	90	100	[1,3,2]	7.28	276.42	3,328.65	30,760.94
64	110	100	90	[1,3,2]	9.23	276.32	8,382.32	35,098.15

4.4.3 Comparison to He and Wang (2014)

Next we consider another function with the following structure:

$$\psi(\S) = I_{S_1 \leq H} I_{S_2 \leq H} I_{S_3 \leq H}. \tag{4.24}$$

We similarly use the same set of parameter values as in the previous example. The results are displayed in Table 4.4. The order of HW is $\{S_3, S_2, S_1\}$ accordingly, and the order of HLT is $\{S_3, S_1, S_2\}$. This is because conditional on S_3, the residual variance of S_1 is larger than that of S_2.

Table 4.4: VRRs

d	H	STD ([1,2,3])	BB	PCA	HW ([3,2,1])	HLT ([3,1,2])
16	95	23.23	2.89	4.24	19.00	26.35
16	100	22.05	5.41	5.86	27.68	40.80
16	105	27.30	5.34	6.62	38.53	48.60
64	95	10.84	1.37	2.72	5.53	10.15
64	100	26.17	3.59	5.21	30.98	54.96
64	105	22.27	2.69	3.63	45.46	36.64

When H is small, the discontinuity has a significant impact on S_1; When H is large, the

discontinuity has a significant impact on S_3; When the discontinuity has no significant impact on a particular function, severity is a good measure of variance contribution.

4.5 Conclusion

We proposed a severity measure to identify the relative importance of the various sub-functions for evaluating an arbitrary function. By incorporating this measure, the mimic functions and hierarchical linear transformation PGM are proposed to explicitly exploits the hierarchical order of functions. Importantly, this method could efficiently handle the functions with multiple non-differentiability or discontinuities.

Drawbacks of this chapter

1. Severity is not the only way of quantifying the importance of the various sub-functions.
2. When there is axis parallel in the discontinuous functions, referring to Wang and Tan (2013), He and Wang (2014), our proposed severity measure may not necessary the best measure.

Chapter 5

Dimension Reduction on Large Portfolio of Insurance Pricing

Key contributions of this chapter

1. Extend the idea of low discrepancy sequences to the problem of selecting the representative synthetic variable annuity (VA) policies.
2. Propose a new algorithm, named as green mesh method, that is not only able to approximate the large VAs portfolio with a high degree of precision, but also is simple, easy to implement, and offers real-time application.
3. The green mesh method has the additional advantage of being portable, i.e. we only need to incur the start-up cost once and the results can be re-cycled or re-used to effectively price other arbitrary large VAs portfolios.

5.1 Introduction

In this chapter, we extend our dimension reduction on pricing large portfolio of insurance contracts. In the last few decades variable annuities (VAs) have become one of the innovative investment-linked insurance products for the retirees. VA is a type of annu-

ity that offers investors the opportunity to generate higher rates of returns by investing in equity and bond subaccounts. Its innovation stems from a variety of embedded guaranteed riders wrapped around a traditional investment-linked insurance product. These guaranteed minimum benefits are collectively denoted as GMxB, where "x" determines the nature of benefits. For example, the guaranteed minimum death benefit (GMDB) guarantees the beneficiary of a VA holder to receive the greater of the sub-account value of the total purchase payments upon the death of the VA holder. The guaranteed minimal accumulation benefit (GMAB), guaranteed minimal income benefit (GMIB) and guaranteed minimal withdrawal benefit (GMWB) are examples of living benefit protection options. More specifically the GMAB and GMIB provide accumulation and income protection for a fixed number of year contingent on survival of the VA holder, respectively. The GMWD guarantees a specified amount for withdrawals during the life of the contract as long as both the amount that is withdrawn within the policy year and the total amount that is withdrawn over the term of the policy stay within certain limits. See the monograph by Hardy (2003) for a detailed discussion on these products.

The appealing features of these guarantees spark considerable growth of the VA markets around the world. According to the Insured Retirement Institute, the VA total sales in the U.S. were $130 billion and $138 billion for 2015 and 2014, respectively. Consequently many insurance companies are managing large VAs portfolio involving hundreds of thousands of policies. This in turn exposes insurance companies to significant financial risks. Hence heightens the need for an effective risk management program (such as the calculation of VA's sensitivity or Greeks to underlying market risk factors) for VAs.

The sophistication and challenges of the embedded guarantees also stimulate a phenomenon interest among academics in proposing novel approaches to pricing and hedging VAs. Because of the complexity of these products, closed-form pricing formulas exist only in some rather restrictive modelling assumptions and simplified guarantee features (see e.g. Boyle and Tian, 2008; Lin et al., 2009; Ng and Li, 2011; Tiong, 2013). In most other cases, the pricing and hedging of VAs resort to numerical methods such as numerical solution

to partial differential equations (PDE, Azimzadeh and Forsyth, 2015; Peng, et al., 2012; Shen, et al., 2016, Forsyth and Vetzal, 2014), tree approach (Hyndman and Wenger, 2014; Dai, et al., 2015; Yang and Dai, 2013), Monte Carlo simulation approach (Bauer, et al., 2008, 2012; Hardy, 2000; Huang and Kwok, 2016; Jiang and Chang, 2010, Boyle, et al., 2001), Fourier Space Time-Stepping algorithm (Ignatieva et al., 2016). See also Bacinello et al. (2016), Zineb and Gobet (2106), Fan et al. (2015), Luo and Shevchenko (2015), Steinortha and Mitchell (2015), for some recent advances on pricing and hedging of VAs.

The aforementioned VA pricing and hedging methodologies tend to be very specialized, customized to a single VA with a specific type of GMxB, and are computationally too demanding to scale to large portfolio of VAs. This is a critical concern as hedging the large portfolio of VA policies dynamically calls for calculation of the VA's sensitivity parameters (i.e. Greeks) in real-time. The need for the capability of pricing and hedging large VAs efficiently has prompted a new direction of research inquiry. The objective is no longer on seeking a method that has the capability of pricing every single policy with a very high precision and hence can be extremely time-consuming, but rather is to seek some compromised solutions that have the capability of pricing hundreds of thousands of policies in real time while retaining an acceptable level of accuracy. This is precisely the motivation by Gan (2013), Gan and Lin (2015, 2016), Gan and Valdez (2016), Hejazi and Jackson (2016), Xu, et al. (2016), Hejazi, et al.(2015), among others. Collectively we refer these methods as the large VAs portfolio valuation methods.

In all of the valuation methods for the large VAs portfolio, the underpinning idea is as follow. Instead of pricing each and every single policy in the large portfolio of VAs, a "representative" set of policies is first selected and their corresponding quantities of interest, such as prices, Greeks, etc, are valuated. These quantities of interest are then used to approximate the required values for each of the policy in the entire large portfolio of VAs. If such approximation yields a reasonable accuracy and that the number of representative policies is very small relative to all the policies in the VAs portfolio, then a significant reduction in the computational time is possible since we now only need to evaluate the

representative set of VA policies.

The above argument hinges on a number of other additional assumptions. First, we need to have an effective way of determining the representative VA policies. By representative, we loosely mean that the selected VA policies provide a good representation of the entire large VAs portfolio. Data clustering is typically used to achieve this objective. Second, we need to have an effective way of exploiting the quantities of interest from the representative VA policies to provide a good approximation to the entire large VAs portfolio. Machine learning based methods such as the Kriging method or other spatial interpolation methods have been proposed to accomplish this task.

The experimental results provided the above mentioned papers have been encouraging. While these methods have already achieved a significant reduction in computational time, there are some potential outstanding issues. We identify the following six issues and we summon that an efficient large VAs portfolio valuation algorithm should adequately address all of these issues:

1. the complexity of the proposed algorithm,
2. the cost of finding representative VA policies,
3. the cost of initial training set, if any,
4. the cost of estimating the entire large VAs from the representative VA policies,
5. the computer memory constraint,
6. the portability to other large VAs portfolio valuation.

Inevitably these issues become more pronounced with the size of the VAs portfolios and the representative VA policies. More concretely, if Monte Carlo method were to price a VAs portfolio consisting of 200,000 policies, the time needed is 1042 seconds, as reported in Table 3 of Gan (2013). If one were to implement the method proposed by Gan (2013), the computational time reduces remarkably by about 70 times with 100 representative

VA policies. However, if one were to increase the representative VA policies to 2000, the reduction in computational time drops from 70 to 14 times. While we are still able to achieve a 14-fold reduction in computational time, the deterioration of the proposed method is obvious.

Many of the existing large VAs portfolio valuation algorithms incur a start-up cost, such as the cost of finding the representative VA policies, before it can be used to approximate all the policies in the VAs portfolio. For some algorithms the initial set-up cost can be computational intensive and hence the property of portability becomes more important. By portability, we mean that we only need to incur the start-up cost once and the results can be re-cycled or re-used. In our context, the portability property implies that we only need to invest once on finding the representative set of VA policies so that the same set of representative VA policies can be re-cycled to effectively price other arbitrary large VAs portfolios. Unfortunately most of the existing large VAs portfolio valuation algorithms do not possess this property. For this reason we refer our proposed large VA valuation method as the green mesh method.

The objective of this paper is to provide another compromised solution attempting to alleviate all of the issues mentioned above. More specifically, our proposed solution has a number of appealing features including its simplicity, ease of implementation, less computer memory, etc. More importantly, the overall computational time is comparatively less and hence our proposed method is a potential real-time solution to the problem of interest. Finally, unlike most other competitive algorithms, our proposed method is portable.

The remaining chapter is organized as follows. Section 5.2 provides a brief overview of some of the large VA valuation algorithms. Our focus is on the method of Gan (2013) and Gan and Lin (2015) and hence a more detailed description is provided. Section 5.3 introduces and describes our proposed green mesh method. Section 5.4 compares and contracts our proposed method to that of Gan (2013) and Gan and Lin (2015).

5.2 Review of some existing large VAs portfolio valuation methods

As pointed out in the preceding section that a few approaches have been proposed to tackle the valuation of large VAs portfolio of n policies. Broadly speaking, the existing methods typically involve the following four steps:

Step 1: Determine a set of k representative synthetic VA policies.

Step 2: Mapping the set of k representative synthetic VA policies onto the set of k representative VA policies that are in the large VAs portfolio.

Step 3: Estimate the quantities of interest of the k representative VA policies.

Step 4: Estimate the quantities of interest of the entire large VAs portfolio from the k representative VA policies.

The above algorithm is best illustrated by considering the k-prototype data clustering method proposed by Gan (2013) and Gan and Lin (2015). The description of this algorithm, including notation, are largely abstracted from these two references. Let $\boldsymbol{x}_i, i = 1, 2, \ldots, n$, represent the i-th VA policy in the large VAs portfolio and $\mathcal{X}$ denote the set containing all the n VA policies; i.e. $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$. Without loss of generality, we assume that each VA policy $\boldsymbol{x} \in \mathcal{X}$ can further be expressed as $\boldsymbol{x} = (x_1, \ldots, x_d)$, where x_j corresponds to the j-th attribute of the VA policy $\boldsymbol{x}$. In our context, examples of attributes are gender, age, account value, guarantee types, etc., so that any arbitrary VA policy is completely specified by its d attributes. We further assume that the attributes can be categorized as either quantitative or categorical. For convenience, the attributes are arranged in such a way that the first d_1 attributes, i.e. $x_j, j = 1, \ldots, d_1$, are quantitative while the remaining $d - d_1$ attributes, i.e. $x_j, j = d_1 + 1, \ldots, d$, are categorical. A possible measure of the

closeness between two policies $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathcal{X}$ is given by (Huang, 1998; Huang et al., 2005)

$$D(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{j=1}^{d_1} w_j (x_j - y_j)^2 + \sum_{j=d_1+1}^{d} w_j \delta(x_j, y_j)}, \tag{5.1}$$

where $w_j > 0$ is a weight assigned to the j-th attribute, and $\delta(x_j, y_j)$ is the simple matching distance defined for the categorical variable as

$$\delta(x_j, y_j) = \begin{cases} 0, & if \quad x_j = y_j, \\ 1, & if \quad x_j \neq y_j. \end{cases}$$

To proceed with Step 1 of determining the set of k representative synthetic VA policies, the k-prototype data clustering method proposed by Gan (2013) and Gan and Lin (2015) boils down to first optimally partitioning the portfolio of n VA policies into k clusters. The k representative synthetic VA policies are then defined as the centers or prototypes of the k clusters. By defining C_j as the j-th cluster and $\boldsymbol{\mu}_j$ as its prototype, the k-prototype data clustering method for determining the optimal clusters (and hence the optimal cluster centers) involves minimizing the following function:

$$P = \sum_{j=1}^{k} \sum_{\boldsymbol{x} \in C_j} D^2(\boldsymbol{x}, \boldsymbol{\mu}_j). \tag{5.2}$$

The above minimization needs to be carried out iteratively in order to optimally determining the membership of C_j and its $\boldsymbol{\mu}_j$. This in turn involves the following four sub-steps:

Step 1a: Initialize cluster center.

Step 1b: Update cluster memberships.

Step 1c: Update cluster centers.

Step 1d: Repeat Step 1b and Step 1c until some desirable stop conditions.

We refer readers to Gan (2013) and Gan and Lin (2015) for the mathematical details of

the above k-prototype algorithm.

After Step 1 of the k-prototype algorithm, we optimally obtain the k representative synthetic VA policies which correspond to the clusters centers $\boldsymbol{\mu}_j, j = 1, \ldots, k$. Step 2 is then concerned with mapping the representative synthetic VA policies to the representative VA policies that are actually within the large VAs portfolio. This objective can be achieved via the nearest neighbour method. By denoting $\boldsymbol{z}_j, j = 1, \ldots, k$, as the j-th representative VA policy, then $\boldsymbol{z}_j$ is optimally selected as one that is closest to the j-th representative synthetic VA policy $\boldsymbol{\mu}_j$. Using (5.1) as the measure of closeness, $\boldsymbol{z}_j$ is the solution to the following nearest neighbour minimization problem:

$$\boldsymbol{z}_j = \underset{\boldsymbol{x} \in \mathcal{X}}{\operatorname{argmin}}\, D(\boldsymbol{x}, \boldsymbol{\mu}_j).$$

Note that $\boldsymbol{\mu}_j$ may not be one of the policies in $\mathcal{X}$ but $\boldsymbol{z}_j$ is. The resulting k representative VA policies are further assumed to satisfy

$$D(\boldsymbol{z}_r, \boldsymbol{z}_s) > 0$$

for all $1 \leq r < s \leq k$ to ensure that these policies are mutually distinct.

Once the k representative VA policies are determined, the aim of Step 3 is to estimate the corresponding quantities of interest, such as prices and sensitivities. This is typically achieved using the Monte Carlo (MC) method, the PDE approach, or other efficient numerical methods that were mentioned in the introduction. The resulting estimate of the quantity of interest corresponding to the representative VA policy $\boldsymbol{z}_j$ is denoted by f_j.

The final step of the algorithm focuses on estimating the entire large VAs portfolio from the k representative VA policies. The method advocated by Gan (2013) and Gan and Lin (2015) is to rely on the Kriging method (Isaaks and Srivastava, 1990). For each VA policy

$\boldsymbol{x}_i \in \mathcal{X}, i = 1, \ldots, n$, the corresponding quantity of interest, $\hat{f}_i$, is then estimated from

$$\hat{f}_i = \sum_{j=1}^{k} w_{ij} y_j \tag{5.3}$$

where $w_{i1}, w_{i2}, \ldots, w_{ik}$ are the Kriging weights. These weights, in turn, are obtained by solving a system of $k + 1$ linear equations.

Note that the algorithm proposed by Gan (2013) and Gan and Lin (2015) draws tools from data clustering and machine learning. For this reason we refer their method as the clustering-Kriging method. The experiments conducted by Gan (2013) and Gan and Lin (2015) indicate that their proposed method can be used to approximate large VAs portfolio with an acceptable accuracy and in a reasonable computational time. However, there are some potential complications associated with the underlying algorithm and hence in practice the above algorithm is often implemented sub-optimally. The key complications are attributed to the k-prototype clustering method and the Kriging method.

Let us now discuss these issues in details. The k-prototype clustering method of determining the representative synthetic VA policies can be extremely time consuming, especially for large n and k. This is also noted in Gan (2013, page 797) that *"if n and k are large (e.g., $n > 10000$ and $k > 20$), the k-prototypes algorithm will be very slow as it needs to perform many distance calculations."* The problem of solving for the optimal clusters using the clustering method is a known difficult problem. In computational complexity theory, this problem is classified as a NP-hard problem. (see Aloise et al. 2009; Dasgupta 2007). This feature severely limits the practicality of using k-prototype clustering method to identify the representative VA policies as in most cases n is in the order of hundreds of thousands and k is in the order of hundreds or thousands. Because of the computational complexity, in practice the k-prototype clustering method is typically implemented sub-optimally via a so-called "divide and conquer" approach. Instead of determining the optimal clusters directly from the large VAs portfolio, the "divide and conquer" strategy involves first partitioning the VAs portfolio into numerous sub-portfolios and then identifying "sub-clusters"

from the sub-portfolio. By construction, the sub-portfolio is many times smaller than the original portfolio and hence the optimal sub-clusters within each sub-portfolio can be obtained in a more manageable time. The optimal clusters of the large VAs portfolio is then the collection of all sub-clusters from the sub-portfolios. While the sub-clusters are optimal to each sub-portfolio, it is conceivable that their aggregation may not be optimal to the entire large VAs portfolio, and hence cast doubt on the "representativeness" of the synthetic VA polices obtained from the centers of the clusters.

The second key issue is attributed to using the Kriging method of approximating the large VAs portfolio. For a large set of representative VA policies, the Kriging method breaks down due to the demand in computer memory and computational time. This issue is also acknowledged by Gan (2013) that *"a large number of representative policies would make solving the linear equation system in Eq. [(5.3)] impractical because solving a large linear equation system requires lots of computer memory and time.*

We conclude this section by briefly mentioning other works related to the large valuation of VAs portfolio. For example, Hejazi, el al. (2015) extended the method of Gan (2013) by investigating two other spatial interpolation techniques, known as the Inverse Distance Weighting and the Radial Basis Function, in addition to the Kriging. Gan and Lin (2016) propose a two-level metamodeling approach for efficient Greek calculation of VAs and Hejazi and Jackson (2016) advocate using the neutral networks. The large VAs portfolio valuation method proposed by Xu, et. al (2016) exploit a moment matching scenario selection method based on the Johnson curve and then combine with the classical machine learning methods, such as neural network, regression tree and random forest, to predict the quantity of interest of the large VAs portfolio. By using the generalized beta of the second kind, in conjunction with the regression method and the Latin hypercube sampling method, Gan and Valdez (2016) provide another method of approximating large VAs portfolios.

5.3 Proposed Green Mesh Method

As we have discussed in the preceding section that the existing large VAs portfolio valuation methods can be quite complicated and computationally intensive, especially when we wish to increase the accuracy of the underlying method by using a larger set of representative VA policies. Recognizing the limitations of these methods, we now propose a new algorithm that is not only able to approximate the large VAs portfolio with a high degree of precision, but also a surprisingly simple, easy to implement, portable, and real-time application. Hence the proposed algorithm has the potential of resolving all the limitations of the existing methods. We refer our proposed approach as the green mesh method. Hence the determination of the representative VA policies is equivalent to the construction of green mesh. Its construction and its application to valuing large VAs portfolio can be summarized in the following four steps:

Step 1: Determine the boundary values of all attributes.

Step 2: Determine a representative set of synthetic VA policies, i.e. green mesh.

Step 3: Estimate mesh's quantities of interest.

Step 4: Estimate the quantities of interest of the entire large VAs portfolio from the green mesh.

Each of the above steps is further elaborated in the following subsections.

5.3.1 Step 1: Boundary values determination

Recall that for a VA policy $\boldsymbol{x} \in \mathcal{X}$, its d attributes are represented by $\{x_1, x_2, \ldots, x_d\}$, where the first d_1 attributes are quantitative and the remaining $d - d_1$ attributes are categorical. The first step of determining the attributes' boundaries is an easy task for the quantitative attributes. This entails finding the extremum values for each j-th quantitative attribute, with $j = 1, \ldots, d_1$, to obtain x_j^{MIN} and x_j^{MAX} that represent the j-th attribute's minimum

and maximum values, respectively. The categorical variables, on the other hand, contain a finite number of categories such as male and female for the gender attribute. Hence the categorical data do not have any intrinsic order and it is meaningless to discuss their extremals. In our application, all our categorical attributes are dichotomous; i.e. binary attributes with only two possible values. For these categorical variables; i.e., for $j = d_1 + 1, \ldots, d$, it is convenient to use x_j^{MIN} to denote one of the binary values (with probability of occurrence p_j) and x_j^{MAX} to represent the other binary value (with probability of occurrence $1 - p_j$).

In summary, after Step 1 we obtain the boundary conditions for all the attributes of our large VAs portfolio. The possible values for which all of these attributes must lie can be represented succinctly as

$$\left[x_1^{MIN}, x_1^{MAX}\right] \times \left[x_2^{MIN}, x_2^{MAX}\right] \times \cdots \times \left[x_d^{MIN}, x_d^{MAX}\right]. \tag{5.4}$$

In other words, the boundary conditions is a d-dimensional hyperrectangle (or box) although care need to be taken in interpreting the attributes that are categorical; i.e. attributes $d_1 + 1, \ldots, d$. In these cases, x_j^{MIN} and x_j^{MAX} do not represent the extremals of the j-th categorical attribute, but rather their possible values since these attributes are assumed to be dichotomy.

5.3.2 Step 2: Green mesh determination

We now proceed to Step 2. The objective of this step is to determine the representative synthetic VA policies; i.e. construct the green mesh. Let $X_j, j = 1, \ldots, d$ denote the random variable of the j-th attribute and $F_j(x)$ be its cumulative distribution function (CDF). For $j = 1, \ldots, d_1$, $F_j(x)$ is the CDF of the random variable X_j, with domain on $\left[x_j^{MIN}, x_j^{MAX}\right]$, that can be continuous or discrete depending on the nature of the attribute. For $j = d_1 + 1, \ldots, d$, X_j is a binary random variable that admits x_j^{MIN} with

probability p_j and x_j^{MAX} with probability $1 - p_j$. Despite there is no intrinsic ordering for the categorical attribute, the corresponding CDF is assumed to be defined as $F_j(x_j^{MIN}) = p_j$ and $F_j(x_j^{MAX}) = 1$.

Determining the representative synthetic VA policies boils down to sampling representative attributes from $F_j(x), j = 1, \ldots, d$, subject to the domains given by the hyperrectangle (5.4). Suppose we are interested in generating m representative synthetic VA policies and that $\boldsymbol{r}_i = (r_{i1}, r_{i2}, \ldots, r_{id}), i = 1, 2, \ldots, k$, denotes the attributes of the i-th representative synthetic VA policy. In practice there exists various ways of sampling r_{ij} from the given CDF $F_j(x)$. In this paper we use the simplest method known as the inversion method. For a given $u_{ij} \in [0, 1]$, the method of inversion implies that the corresponding r_{ij} can be determined as

$$r_{ij} = F_j^{-1}(u_{ij}), \tag{5.5}$$

where $i = 1, \ldots, k, j = 1, \ldots, d$, and $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$. Care must be taken for inverting $F_j(x)$ since the function can be continuous or right continuous. For example, suppose the random variable X_j is a quantitative attribute that is uniformly distributed on $\left[x_j^{MIN}, x_j^{MAX}\right]$. Then its CDF has the form $F_j(x) = \frac{x - x_j^{MIN}}{x_j^{MAX} - x_j^{MIN}}$, where $x_j^{MIN} \leq x \leq x_j^{MAX}$ and $x_j^{MIN} \neq x_j^{MAX}$. This implies that for a given arbitrary $u_{ij} \in [0, 1]$, the inversion method yields

$$r_{ij} = u_{ij} x_j^{MAX} + (1 - u_{ij}) x_j^{MIN}.$$

If X_j is a categorical attribute, then for $u_{ij} \in [0, 1]$ the inversion method of (5.5) becomes

$$r_{ij} = \begin{cases} x_j^{MIN} & \text{if } u_{ij} \leq p_j \\ x_j^{MAX} & \text{otherwise} \end{cases}$$

In summary, if $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{id}) \in [0, 1]^d, i = 1, 2, \ldots, k$, then the inversion method (5.5) provides a simple way of transforming each d-dimensional hypercube point $\boldsymbol{u}_i$ onto the synthetic VA policy with attributes given by $\boldsymbol{r}_i$. The linkage between the input hypercube

points $\boldsymbol{u}_i$ and the output synthetic VA policy $\boldsymbol{r}_i$, for $i = 1, 2, \ldots, k$, provides a useful clue on determining the quality of the representativeness of the synthetic VA policies. In particular, the greater the uniformity of $\boldsymbol{u}_i, i = 1, 2, \ldots, k$ over the d-dimensional hypercube $[0, 1]^d$, the better the representativeness of the synthetic VA policies.

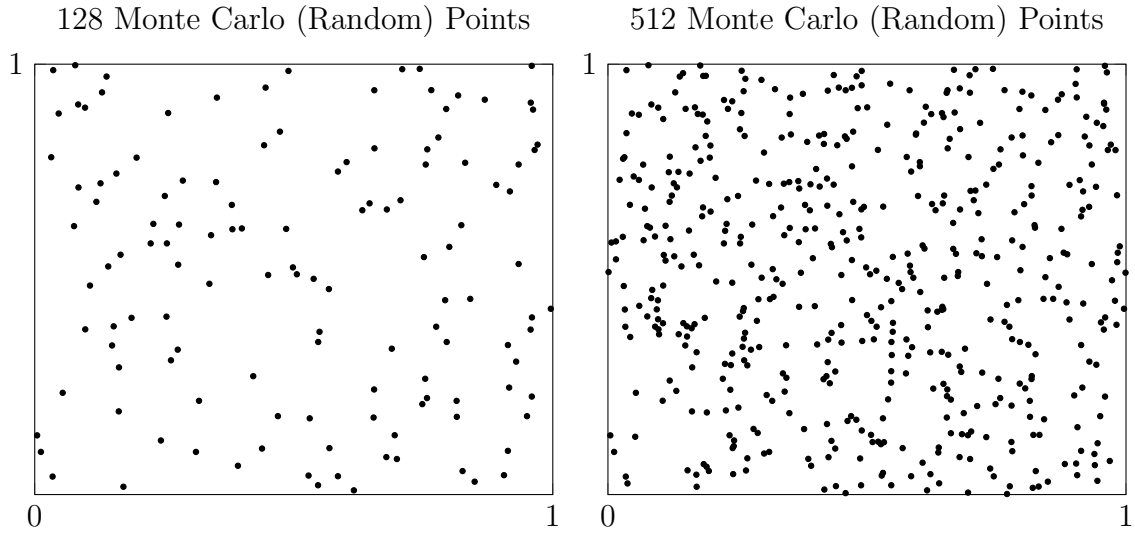

Figure 5.1: A sample of 128 and 512 Monte Carlo (random) points

There are a few ways of sampling uniformly distributed points from $[0, 1]^d$. Among them, the simplest approach is based on the Monte Carlo (MC) method that selects points randomly from $[0, 1]^d$. While this is the most common method, the points distribute uniformly on $[0, 1]^d$ at a rate of $O(1/k^{1/2})$, which is deemed to be slow for practical application. Furthermore because of the randomness, the generated finite set of points tends to have gaps and clusters and hence these points are far from being uniformly distributed on $[0, 1]^d$. This issue is highlighted in the two panels in Figure 5.1 which depict two-dimensional projections of 128 and 512 randomly selected points. Gaps and clusters are clearly visible from these plots even when we increase the number of points from 128 to 512.

A possible solution of alleviating the issues associated with the MC is to rely on the stratification sampling, which is one of the common variance reduction techniques. Suppose we are interested in sampling m points that are uniformly distributed on the one-dimensional

unit interval $[0,1]$. Rather than selecting m points randomly from $[0,1]$ as in the classical MC method, the simplest stratification method is to first divide the unit interval equally into m strata, and then randomly select a point from each of these strata to produce k stratified samples. The resulting stratified samples have exactly one point in each stratum (i.e. sub-interval), hence effectively avoiding gaps and clusters and hence the points tend to have greater uniformity on $[0,1]$.

While the stratification method has proven to be a very effective variance reduction method for one-dimensional application, stratifying high dimensions can be problematic due to the exponential growth of the stratified samples. To see this, let us suppose each dimension is divided into k strata. A direct generalization of the stratification would lead to stratifying k^d strata from $[0,1]^d$. This is computationally prohibitive even for moderate values of m and d. To prevent the exponential growth of the stratified samples, an alternate method known as the Latin hypercube sampling (LHS) has been proposed (see McKay et al., 1979). LHS, which is another popular variance reduction technique, can be an efficient and a practical high dimensional stratification method.

LHS is a multi-dimensional generalization of the basic one-dimensional stratification in the following sense. Suppose we are interested in k LHS samples from $[0,1]^d$. As in the basic stratification, each dimension is first stratified independently to produce m stratified samples of one dimension. To produce the d-dimensional stratified samples, rather than using the Cartesian product approach of producing k^d d-dimensional stratified samples, LHS samples are generated by merely concatenating all one-dimensional "marginal" k stratified samples. Algorithmically, the LHS samples can be generated as follows:

$$\hat{u}_{ij} = \frac{\pi_{ij} - 1 + u_{ij}}{k}, \qquad i = 1, \ldots, k, j = 1, \ldots, d, \tag{5.6}$$

where u_{ij} is the original randomly selected sample from $[0,1]$ and $(\pi_{1j}, \pi_{2j}, \ldots, \pi_{kj})$ are independent random permutations of $\{1, 2, \ldots, k\}$. Then $\hat{\boldsymbol{u}}_i = (\hat{u}_{i1}, \hat{u}_{i2}, \ldots, \hat{u}_{id}), i = 1, 2, \ldots, k$ is the required LHS samples from $[0,1]^d$ and the resulting LHS samples, in turn, are applied

to (5.5) to produce the necessary representative synthetic VA attributes.

The method of LHS offers at least the following three advantages. Firstly, the number of points required to produce a stratified samples is kept to a manageable size, even for high dimension. Secondly, by construction the one-dimensional projection of the stratified samples reduces to the basic one-dimensional marginal stratification of m samples. Thirdly, the intersection of the stratum in high dimension has exactly one point. Panel (a) in Figure 5.2 depicts a 2-dimensional LHS sample of 8 points. To appreciate the stratification of the Latin hypercube sampling, we have produced Panels (b), (c) and (d) for better visual effect. First note that the points in these panels are exactly the same as that in Panel (a). Second, by subdividing the unit horizontal axis into 8 rectangles of equal size as in Panel (b), each of the rectangles contains exactly one point. The same phenomenon is observed if we partition the vertical axis as in Panel (c), instead of the horizontal axis. Finally, if both axes are divided equally as shown in Panel (d), then each intersection of the vertical and horizontal rectangle again has exactly a point in the sense that once a stratum has been stratified, then there are no other stratified points along either direction (vertical or horizontal) of the intersection of the stratum. This property ensures that each marginal stratum is optimally evaluated only once.

From the above discussion and demonstration, the method of LHS seems to provides a reasonable good space-filling method for determining the representative VA policies. In fact some variants of LHS have also been used by Hejazi, el al. (2015) and Gan and Valdez (2016) in connection to valuation of large VAs portfolios. To end the discussion on LHS, we point out two potential issues of using LHS to generate the representative synthetic VA policies. Firstly, whenever we wish to increase the stratified samples of LHS, we need to completely re-generate the entire set of representative synthetic VA policies. This is because the LHS design depends on k. If k changes, then we need to re-generate a new set of LHS design. Secondly and more importantly, the LHS design is essentially a combinatorial optimization. For a given k and d, there are $(k!)^{d-1}$ possible LHS designs. This implies that if we were to optimize 20 samples in 2 dimensions, then there are 10^{36}

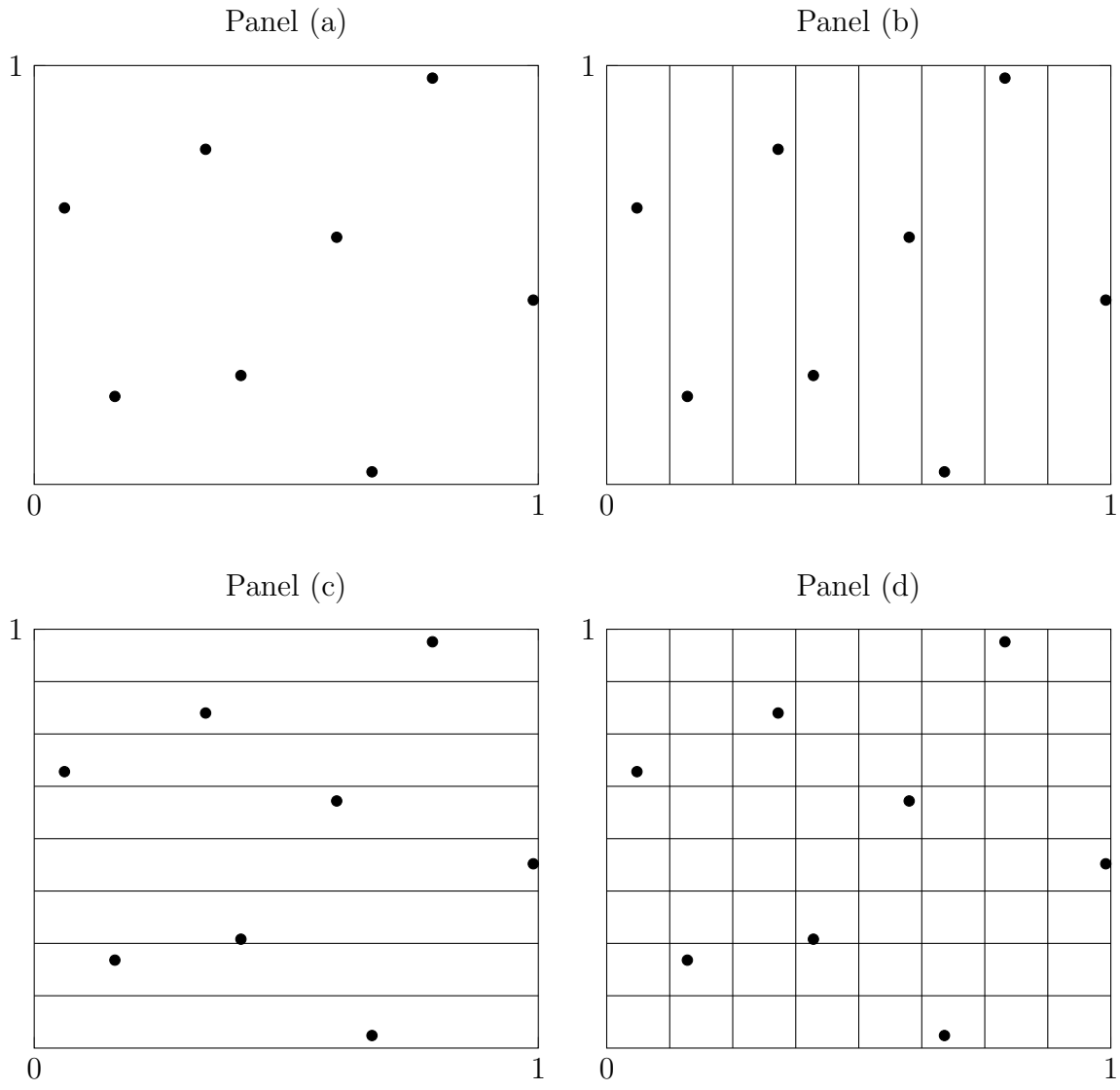

Figure 5.2: An example of 8 points LHS

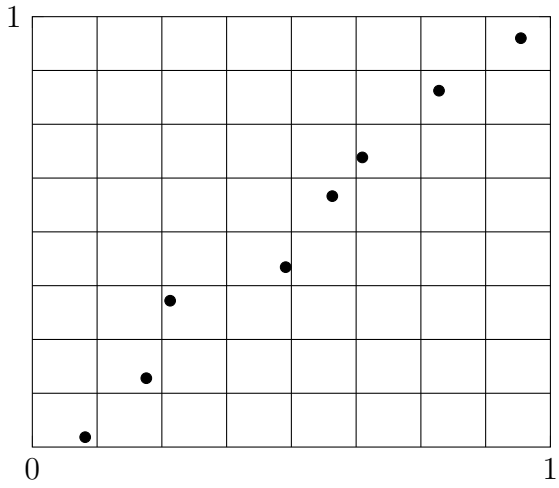

Figure 5.3: A perfectly valid 8 points LHS but with a bad distribution

possible designs to choose from. If the number of dimensions increases to 3, then we will have more than 10^{55} designs. Hence it is computational impossible to optimally select the best possible LHS design. More critically, not all of these designs have the desirable "space-filling" quality such as that shown in Figure 5.3. Although unlikely, there is nothing preventing a LHS design with the extreme case as illustrated in Figure 5.3.

Finally we present another promising sampling method based on the randomized quasi-Monte Carlo (RQMC) method. Before describing RQMC, it is useful to distinguish the quasi-Monte Carlo (QMC) method from the classical MC method. In a nutshell, the key difference between these two methods lies on the properties of the points they use. MC uses points that are randomly generated while QMC relies on specially constructed points known as low discrepancy points/sequences. The low discrepancy points have the characteristics that they are deterministic and have greater uniformity (i.e. low discrepancy) than the random points. Some popular explicit constructions of low discrepancy points/sequences satisfying these properties are attributed to Halton (1960), Sobol (1967) and Faure (1982). Theoretical justification for using points with enhanced uniformity follows from the Koksma-Hlawka inequality which basically asserts that the error of using sampling-based method of approximating a d-dimensional integral (or problem) depends

crucially on the uniformity of the points used to evaluate the integral. Hence for a given integral, points with better uniformity lead to lower upper error bound. It follows from the Koksma-Hlawka inequality that QMC attains a convergence rate $O(k^{-1+\epsilon}), \epsilon > 0$, which asymptotically converges at a much faster rate than the MC rate of $O(k^{-1/2})$. The monograph by Niederreiter (1992) provides an excellent exposition on QMC. The theoretically higher convergence rate of QMC has created a surge of interests among financial engineers and academicians in the pursuit of more effective ways of valuing complex derivative securities and other sophisticated financial products. See for example Glasserman (2004) and Joy et al. (1996).

Recent development in QMC has supported the use of a randomized version of QMC, i.e. RQMC, instead of the traditional QMC. The idea of RQMC is to introduce randomness to the deterministically generated low discrepancy points in such a way that the resulting set of points is randomized while still retaining the desirable properties of low discrepancy. Some advantages of RQMC are (i) improve the quality of the low discrepancy points; (ii) permit the construction of confidence interval to gauge the precision of RQMC estimator; (iii) under some additional smoothness conditions, the RQMC's root mean square error of integration using a class of randomized low discrepancy points is $O(k^{-1.5+\epsilon})$, which is smaller than the unrandomized QMC error of $O(k^{-1+\epsilon})$. See for example Tan and Boyle (2000), Owen (1997) and Lemieux (2009). For these reasons our proposed construction of green mesh will be based on RQMC with randomized Sobol points, other randomized low discrepancy points can similarly be used. The two panels in Figure 5.4 plot two-dimensional randomized Sobol points of 128 and 512 points. Compared to the MC points, the randomized Sobol points appear to be more evenly dispersed over the entire unit-square.

5.3.3 Step 3: Estimating green mesh's quantities of interest

Once the representative set of synthetic VA policies has been determined from Step 2, we need to compute their quantities of interest, such as their market value and their dollar

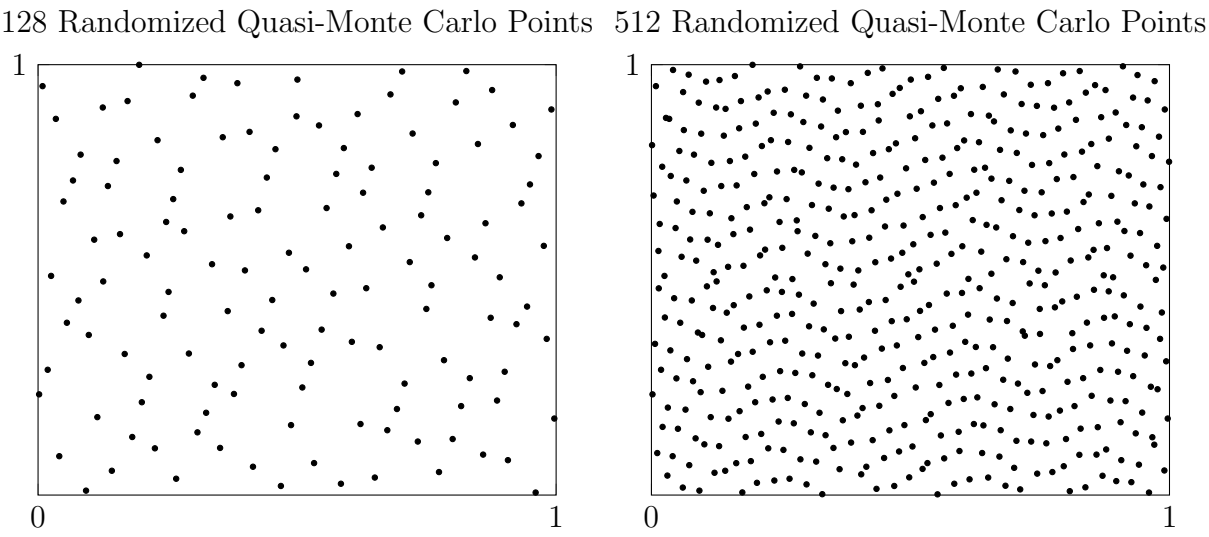

Figure 5.4: A sample of 128 and 512 randomized Quasi-Monte Carlo points

delta. As in Gan (2013) and Gan and Lin (2015), the simplest method is to rely on the MC simulation method, which we also adopt for our proposed method. In this step, we will also be estimating the gradient of each representative synthetic VA policy, which is needed in Step 4.

5.3.4 Step 4: Large VAs portfolio approximation from the green mesh

Once the green mesh and its corresponding quantities of interest have been determined from Steps 2 and 3, respectively, the remaining task is to provide an efficient way of approximating the large VAs portfolio from the green mesh. Recall that Gan (2013) and Gan and Lin (2015) advocate using the Kriging method, which unfortunately is computationally burdensome, in term of both time and memory requirement. To alleviate these issues, we propose a very simple approximation technique by combining both the nearest neighbour method and the Taylor's method. In particular, let $\boldsymbol{x} \in \mathcal{X}$ be one of the VA policies and $f(\boldsymbol{x})$ be its quantity of interest we are interested in approximating. Our proposed

approximation method involves the following two substeps:

Step 4a: Use the nearest neighbour method to find the green mesh (i.e. the representative synthetic VA policy that is closest to $\boldsymbol{x}$. This can be determined by resorting to distance measure such as that given by (5.1). Let $\boldsymbol{y}$ be the synthetic VA policy that is closest to $\boldsymbol{x}$.

Step 4b: $f(\boldsymbol{x})$ can be estimated from the nearest neighbour $\boldsymbol{y}$ via the following Taylor approximation:

$$f(\boldsymbol{x}) \approx f(\boldsymbol{y}) + \nabla f(\boldsymbol{y})(\boldsymbol{x} - \boldsymbol{y}), \tag{5.7}$$

where $\nabla f(\boldsymbol{y})$ is the gradient of f with respect to $\boldsymbol{y}$.

Step 4b is simply the first order Taylor approximation and hence it requires the pre-computation of gradients for each synthetic VA policy $\boldsymbol{y}$. These values were pre-computed from Step 3. We assume that the given VA policy $\boldsymbol{x}$ is sufficiently close to the synthetic VA policy $\boldsymbol{y}$ so that the locally linear approximation yields sufficient accuracy. In situation where the linear approximation is questionable, higher order Taylor approximation can be applied to enhance the accuracy.

The above four steps complete the description of our proposed large VAs portfolio valuation method. There are some important features of our proposed method. Other than its simplicity and its ease of implementation, the most appealing feature is that if the boundaries of all the attributes are set correctly, then we only need to determine the representative set of synthetic VA policies once. This implies that if we subsequently change the composition of the VAs portfolio such as changing its size, then the same set of representative synthetic VA policies can be used repeatedly. This is the portability of the algorithm that we emphasized earlier. Also, our proposed algorithm avoids the mapping of the synthetic VA policy to the actual VA policy within the portfolio. This avoid using the nearest neighbour method to furnish the mapping and hence reduce some computational effort.

5.4 Large VAs portfolio valuation

The objective of this section is to provide some numerical evidences on the relative efficiency of our proposed mesh methods. The set up of our numerical experiment largely follow that of Gan (2013) and Gan and Lin (2015). In particular, we first create two large VAs portfolios consisting of 100,000 and 200,000 policies, respectively. The VA policies are randomly generated based on the attributes given in Table 5.1. The attributes with a discrete set of values are assumed to be selected with equal probability. Examples of attributes satisfying this property are "Age", "GMWB withdrawal rate", "Maturity". "Guarantee type" and "Gender". The latter two are the categorical attributes, with "Guarantee type" specifies the type of guarantees of the VA policy; i.e. GMDB only or GMDB + GMWB, and "Gender" admits either "male" or "female". The remaining attribute "Premium" is a quantitative variable with its value distributes uniformly on the range $[10000, 500000]$.

Attributes	Values
Guarantee type	GMDB only, GMDB + GMWB
Gender	Male, Female
Age	20, 21, . . . , 60
Premium	[10000, 500000]
GMWB withdrawal rate	0.04,0.05,. . . ,0.08
Maturity	10, 11, . . . , 25

Table 5.1: Variable annuity contract specification

The two VAs portfolios constructed above can then be used to test against various large VAs portfolio valuation methods. As in Gan (2013) and Gan and Lin (2015), we are interested in the VAs portfolio's market value and its sensitivity parameter dollar delta. We assume that all the VA policies in our benchmark portfolios are written on the same underlying fund so that to estimate their market values and dollar deltas, we need to simulate trajectories of the fund over the maturity of the VA policies. Under the lognormality assumption, the yearly fund can be simulated according to

$$S_t = S_{t-1} e^{r - \frac{1}{2}\sigma^2 + \sigma Z_t}, \ t = 1, 2, \cdots \tag{5.8}$$

where r is the annual interest rate, σ is the volatility of the fund, S_t is the value of the fund in year t, and Z_t is the standardized normal variate. On the valuation date $t = 0$, the fund value is normalized to 1; i.e. $S_0 = 1$. We also set $r = 3\%$ and $\sigma = 20\%$. Note that to simulate a trajectory of the fund , we need to generate $\{Z_t, t = 1, 2, \ldots, \}$. As discussed in details in Gan (2013), the simulated trajectory can then be used to evaluate each and every VA policy within the portfolio, depending on the attributes.

The simulated VAs portfolio's market values and its dollar deltas are displayed in Table 5.2. The reported values under the heading "Benchmark" are estimated using the MC method with 10,000 trajectories and applying to all the policies in the portfolio. For our numerical experiment, we will assume that these estimates are of sufficient accuracy so that they are treated as the "true values" and hence become the benchmark for all future comparisons. The values reported under "MC-1024" are based on the same method as the "Bencmark" except that these values are estimated based on a much smaller number of trajectories, i.e. 1024 trajectories. The values for the "RQMC-1024" are also the corresponding estimates from valuing all policies in the portfolio except that the trajectories are generated based on the RQMC method with 1024 paths and randomized Sobol points. Here we should emphasize that when we simulate the trajectories of the fund, (5.8) is used iteratively for MC. For RQMC, the trajectories are generated using the randomized Sobol points coupling with the principal component construction. Detailed discussion of this method can be found in Wang and Tan (2013). To gauge the efficiency of the MC and RQMC from using almost 10-fold smaller number of trajectories, errors relative to the "Benchmark" are tabulated. The effectiveness of RQMC is clearly demonstrated, as supported by the small relative errors of less than 1%, regardless of the size of the VAs portfolio. In contrast, the MC method with the same number of simulated paths leads to relative errors of more than 7%.

Method	Market Value	Relative Error	Dollar Delta	Relative Error
$n = 100,000$				
Benchmark	736,763,597	-	-2,086,521,832	-
MC-1024	793,817,541	7.74%	-2,237,826,994	7.25%
RQMC-1024	742,849,028	0.83%	-2,094,359,793	0.38%
$n = 200,000$				
Benchmark	1,476,872,416	-	-4,182,264,066	-
MC-1024	1,591,202,285	7.74%	-4,484,725,861	7.23%
RQMC-1024	1,489,074,001	0.82%	-4,197,858,556	0.37%

Table 5.2: Market values and dollar deltas of the two large VAs portfolio of $n = 100,000$ and $200,000$ policies. The "benchmark" is the MC method with 10,000 trajectories, MC-1024 is the MC method with 1024 trajectories, and RQMC-1024 is the RQMC method with 1024 trajectories. The relative errors are the errors of the respective method relative to the "benchmark".

We now proceed to comparing the various large VAs portfolio valuation methods. By using the two large VAs portfolios constructed above, we consider two green meshes based on the methods of LHS and RQMC. Recall that our VAs portfolios consist of two categorical attributes: "Guarantee type" and "Gender". In term of the sensitivity of the attribute to the value of VAs portfolio, the categorical attribute tends to be more pronounced than the quantitative attribute. For example, the value of a VA policy crucially depends on the feature of the guarantee. This implies that two policies with different "Guarantee type" can have very different market values (and dollar delta) even though their remaining attributes are exactly the same. Similarly two VA policies with identical attributes except that one annuitant is male and the other is female, the value of the policy can be quite different due to the difference in gender's mortality. For this reason, if a VA policy is approximated by a representative synthetic VA policy for which their categorical attributes are not aligned, then the proposed Taylor approximation may not be as accurate. To avoid this situation, a possible adjustment is to implement the mesh method in the following way. Rather than selecting a single set of synthetic VA policies that is representative of all attributes (i.e. both quantitative and categorical), we can first focus on finding the representative synthetic VA policies for each possible combination of categorical attributes. For instance,

there are four possible combinations of "Guarantee type" and "Gender" in our example. Then for each stratified combination, a set of synthetic VA policies that is representative of only the quantitative attributes is determined and evaluated. Via the nearest neighbour approach and the Taylor method, these representative synthetic VA policies, in turn, are used to approximate those VA policies that have the same categorical attributes. We refer this implementation as the *conditional-green mesh* approach, to differentiate it from the standard green mesh method.

There are some advantages associated with the conditional-green mesh implementation. Firstly, the dimensionality of the conditional-green mesh method is reduced by the number of categorical attributes. This decreases the dimension of the representative synthetic VA policies so that fewer points are needed to achieve the representativeness. Secondly, the nearest neighbour method can be implemented more efficiently since we now only need to focus on the "closeness" among all the quantitative attributes. Furthermore, the number of representative synthetic VA policies conditioned on the combination of categorical attributes tends to be smaller than the total number of representative synthetic VA policies under the standard mesh method. Hence the computational time required for the nearest neighbour search is also reduced. Thirdly and more importantly, by first conditioning on the categorical attributes and then using the resulting representative synthetic VA policies to approximate the VA policies has the potential of increasing the accuracy of the Taylor's approximation since any mismatch among the categorical attributes has been eliminated.

Tables 5.3 and 5.4 give the necessary comparison for $n = 100,000$ and $200,000$ VA policies, respectively. For each large VAs portfolio, we compare and contrast five valuation methods: Gan (2013) and Gan and Lin (2015) clustering-Kriging method, mesh method and conditional mesh method based on LHS and RQMC sampling, respectively. For the each method we use four different set of representative points $k \in \{100, 500, 1000, 2000\}$ to estimate each VA policy's market value and its dollar delta. For the conditional mesh method, there are four equal probable combinations of the categorical variables so that for fair comparison, $k/4$ mesh points are used of these sub-meshes.

Due to the memory constraint and computational time, a sub-optimal clustering-Kriging method is implemented. More specifically, for the k-prototype clustering algorithm of determining the representative VA policies, the large VAs portfolio is first divided into a sub-portfolios comprises of 10,000 VA policies so that there are 10 and 20 of these sub-portfolios for $n = 100,000$ and $200,000$, respectively. Then the k-prototype clustering algorithm is applied to each sub-portfolio 100 and 50 representative VA policies for $n = 100,000$ and $200,000$, respectively, so that in aggregate there are $k = 1000$ for each of the large VAs portfolios. For $k = 2000$, the above procedure applies except that we generate 200 and 100 representative VA policies from the corresponding sub-portfolios for $n = 100,000$ and $200,000$, respectively. The Kriging method also faces the same challenge. In this case, For Kriging method, we solve the linear equation system according to equation (6) of Gan (2013) for each data point except the representative VA. i.e. For 100,000 data points and $k = 1,000$, we solve 100,000-1,000=99,000 linear equation systems; For 200,000 data points and $k = 1,000$, we solve 200,000-1,000=199,000 linear equation systems.

For each of the above method, we quantify its efficiency by comparing the following error measures:

$$\text{MAPE} = \frac{\sum_{i=1}^{n} |A_i - B_i|}{\left|\sum_{i=1}^{n} B_i\right|} \tag{5.9}$$

$$\text{MRE} = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{A_i - B_i}{B_i}\right|. \tag{5.10}$$

where A_i and B_i are, respectively, the estimated value and benchmark value for the i-th VA policy, for $i = 1, 2, \ldots, n$. MAPE denotes the mean absolute percentage error and MRE the mean relative error. In all of these cases, smaller value implies greater efficiency of the underlying method. Furthermore, these measures quantify errors with respect to pricing each VA policy in some averaging ways and do not allow cancellation of errors.

An conclusion can be drawn from the results reported in these two tables is that the

clustering-Kriging method of Gan (2013) and Gan and Lin (2015) is clearly not very efficient, with unacceptably large errors in many cases and irrespective of which error measure. The mesh methods, on the other hand, are competitively more effective than the clustering-Kriging method. Comparing among the green mesh methods, the conditional green mesh method in general leads to smaller errors, hence supporting the advantage of constructing the green mesh conditioning on the categorical variables. Finally the mesh constructed from the RQMC method yields the best performance. More specifically, the method of clustering-Kriging with 100 clusters yields MAPE of 0.4072 for estimating 100,000 VA policies' market value. If the sampling method were based on LHS, then it is more effective than Gan (2013) and leads to smaller MAPE by 2.4 and 2.9 times for standard green mesh and conditional green mesh, respectively. However, the QMC-based sampling method produces even smaller MAPE of 2.6 and 3.6 times for standard green mesh and conditional green mesh, respectively. By increasing k from 500 to 2,000, the MAPE of Gan (2013)'s method reduces from 0.4072 to 0.0652 for estimating the same set of VA policies. On the other hand, the efficiency of LHS method increased by 2.5 and 3.1 times for standard green mesh and conditional green mesh, respectively. The QMC-based sampling method gives even better performance and leads to 3.2 and 4.3 times smaller MAPE for the same set of comparison.

Method	Market Value	MAPE	MRE	Dollar Delta	MAPE	MRE
$k = 100$						
Clustering-Kriging	733,408,057	0.4072	2.9113	-2,111,059,490	0.3966	3.7120
Standard-Mesh-LHS	779,382,382	0.1708	1.1190	-2,187,748,111	0.1634	0.7207
Standard-Mesh-RQMC	774,787,351	0.1560	0.5905	-2,177,917,524	0.1426	0.4937
Conditional-Mesh-LHS	756,514,592	0.1390	1.2262	-2,142,052,361	0.1307	0.8952
Conditional-Mesh-RQMC	758,395,576	0.1119	0.7700	-2,127,495,679	0.1088	0.5904
$k = 500$						
Clustering-Kriging	743,952,147	0.1504	1.1946	-2,106,736,645	0.1395	1.4383
Standard-Mesh-LHS	750,143,449	0.0718	0.2909	-2,118,171,620	0.0689	0.2317
Standard-Mesh-RQMC	746,493,713	0.0554	0.2287	-2,108,624,022	0.0541	0.1880
Conditional-Mesh-LHS	751,246,815	0.0583	0.2489	-2,120,137,979	0.0557	0.2012
Conditional-Mesh-RQMC	746,632,631	0.0437	0.2134	-2,107,111,384	0.0450	0.1716
$k = 1,000$						
Clustering-Kriging	746,981,504	0.1004	0.7411	-2,100,747,984	0.0881	0.8729
Standard-Mesh-LHS	743,740,049	0.0439	0.1683	-2,099,803,880	0.0414	0.1396
Standard-Mesh-RQMC	743,569,931	0.0387	0.1542	-2,101,802,728	0.0380	0.1226
Conditional-Mesh-LHS	744,940,085	0.0345	0.1546	-2,103,392,558	0.0344	0.1266
Conditional-Mesh-RQMC	744,379,692	0.0291	0.1372	-2,101,114,587	0.0300	0.1154
$k = 2,000$						
Clustering-Kriging	744,091,499	0.0652	0.4334	-2,101,521,754	0.0580	0.5043
Standard-Mesh-LHS	744,368,005	0.0264	0.1057	-2,099,895,472	0.0259	0.0897
Standard-Mesh-RQMC	743,474,594	0.0206	0.0742	-2,097,515,603	0.0214	0.0624
Conditional-Mesh-LHS	744,432,495	0.0207	0.0902	-2,099,564,732	0.0214	0.0773
Conditional-Mesh-RQMC	743,848,669	0.0152	0.0669	-2,097,895,831	0.0155	0.0569

Table 5.3: Comparison of large VAs portfolio valuation methods on a portfolio of $100,000$ policies

Method	Market Value	MAPE	MRE	Dollar Delta	MAPE	MRE
$k = 100$						
Clustering-Kriging	1,486,323,825	0.4112	2.8492	-4,238,127,436	0.4195	3.4307
Standard-Mesh-LHS	1,562,487,430	0.1699	1.1450	-4,386,501,086	0.1629	0.7362
Standard-Mesh-RQMC	1,551,783,419	0.1540	0.5850	-4,362,008,069	0.1414	0.4907
Conditional-Mesh-LHS	1,517,430,764	0.1384	1.2058	-4,297,122,824	0.1303	0.8773
Conditional-Mesh-RQMC	1,519,975,189	0.1120	0.7562	-4,264,064,685	0.1088	0.5781
$k = 500$						
Clustering-Kriging	1,524,702,769	0.1600	1.2734	-4,300,058,619	0.1435	1.5098
Standard-Mesh-LHS	1,503,408,982	0.0709	0.2837	-4,245,571,030	0.0684	0.2260
Standard-Mesh-RQMC	1,496,248,048	0.0559	0.2288	-4,227,061,000	0.0546	0.1879
Conditional-Mesh-LHS	1,506,144,258	0.0580	0.2435	-4,250,124,644	0.0554	0.1973
Conditional-Mesh-RQMC	1,496,583,907	0.0434	0.2067	-4,223,483,061	0.0449	0.1673
$k = 1,000$						
Clustering-Kriging	1,475,050,093	0.1064	0.7102	-4,162,797,915	0.0969	0.8247
Standard-Mesh-LHS	1,489,935,827	0.0438	0.1681	-4,207,022,610	0.0413	0.1386
Standard-Mesh-RQMC	1,490,282,637	0.0390	0.1535	-4,213,262,468	0.0383	0.1223
Conditional-Mesh-LHS	1,493,011,025	0.0339	0.1508	-4,215,750,737	0.0340	0.1233
Conditional-Mesh-RQMC	1,492,559,392	0.0292	0.1362	-4,212,586,115	0.0301	0.1149
$k = 2,000$						
Clustering-Kriging	1,482,550,735	0.0643	0.3962	-4,185,408,010	0.0556	0.4496
Standard-Mesh-LHS	1,491,875,469	0.0261	0.1057	-4,208,431,745	0.0257	0.0896
Standard-Mesh-RQMC	1,490,238,009	0.0209	0.0740	-4,204,332,533	0.0215	0.0617
Conditional-Mesh-LHS	1,492,100,073	0.0207	0.0889	-4,207,726,068	0.0214	0.0763
Conditional-Mesh-RQMC	1,491,084,172	0.0152	0.0665	-4,205,294,903	0.0155	0.0567

Table 5.4: Comparison of large VAs portfolio valuation methods on a portfolio of $200,000$ policies

Comparing both green meshes, the conditional green mesh has the advantage of dimension reduction and that each possible categorical variables combination has been stratified with the right proportion of representative synthetic VA policies. As supported by our numerical results, the conditional green mesh is more efficient than the standard green mesh, irrespective of the sampling method.

As noted in the above illustration that the sampling method also plays an important role in determining the efficiency of green meshes. We now provide an additional insight on why the sampling method based on the RQMC is more efficient than the LHS. This can be attributed to the greater uniformity of the RQMC points, as demonstrated in Table 5.5 which displays the standard green meshes of 8 representative synthetic VA policies

constructed from LHS and RQMC sampling methods. Recall in our example we have 2 categorical attributes "Guarantee types" and " Gender", each with two equal probable values (denoted by either 0 or 1) and hence four equally likely combinations represented by $\{(0,0),(0,1),(1,0),(1,1)\}$. With 8 representative synthetic VA policies, each of these permutations should ideally have 2 policies. This patten is observed for RQMC but not LHS. The LHS sampling, for example, has three representative synthetic VA policies that are $(0,1)$ and $(1,0)$ combinations but only one representative synthetic VA policy in combinations $(0,0)$ and $(1,1)$. Therefore the RQMC sampling a more uniform sampling than the LHS sampling, even under the standard green mesh construction.

	Green Mesh (Representative synthetic VA policies)							
	1	2	3	4	5	6	7	8
LHS								
Guarantee Type	1	1	0	0	0	0	1	1
Gender	1	0	1	0	1	1	0	0
Age	49	37	29	52	59	40	23	31
Premium	59,694	114,891	188,169	342,207	213,144	405,271	290,174	460,528
Withdrawal rate	0.05	0.05	0.06	0.06	0.08	0.08	0.07	0.04
Maturity	11	12	17	19	15	22	21	25
RQMC								
Guarantee Type	1	0	1	0	1	0	1	0
Gender	0	1	1	0	1	0	0	1
Age	24	58	37	44	55	26	46	34
Premium	313,162	29,307	155,153	438,036	365,276	97,869	217,046	485,394
Withdrawal rate	0.04	0.07	0.06	0.07	0.08	0.05	0.06	0.05
Maturity	11	23	20	16	18	14	13	25

Table 5.5: Comparison of standard green mesh construction under LHS and RQMC samplings with 8 points

We now compare the computational efforts among the various methods. For our proposed green meshes, the computational time can be divided into two parts. The first part relates to the initial cost of determining a representative set of synthetic VA policies as well as determining their market values, dollar deltas and the gradients for the Taylor approximation. This computational effort, which is denoted as the start-up time, is invariant to the size of the portfolio. The second part of the computational effort corresponds to

the time it takes to approximate all VA policies in the portfolio. This is denoted as the evaluation time and is proportional to the number of policies in the portfolio. Using the same two large VAs portfolios of 100,000 and 200,000 policies, Table 5.6 reports the breakdown of the computational time for the various implementations of green meshes involving $k \in \{100, 500, 1000, 2000\}$.

Let us first comment on the start-up time of the green meshes. Firstly, the start-up time for implementing among various green meshes is very comparable. Secondly, the start-up time is proportional to k, the size of mesh points. Thirdly, and most importantly is that the start-up time is invariant to the size of the VAs portfolio. This is an important feature in that once we have produced a representative set of synthetic VA policies, the same mesh can be re-used or re-cycled on any other large VAs portfolios. We now consider the evaluation time. Similar to the start-up time, the evaluation time for the various mesh implementations are also competitively comparable and also proportional to k. However, the evaluation time in this case depends on the size of the VAs portfolio; doubling the size of the VAs portfolio doubling the evaluation time.

To appreciate the computational efficiency of our proposed green meshes, Table 5.6 also reports the computational time if the entire VAs portfolio were to be evaluated using the benchmark method (MC with 10,000 trajectories) and MC and RQMC methods (with 1,024 trajectories). If the MC method with 10,000 trajectories (i.e. the benchmark method) were to evaluate the entire portfolio of 200,000 VA policies, it would take about 8.4 hours of computational time. This is 10 times computationally more intensive than the corresponding MC and RQMC methods (with 1,024 trajectories) since the latter methods use trajectories of almost 10-fold smaller. In contrast, if the RQMC-based conditional mesh with $k = 500$ and $1,000$ were used to approximate the same large VAs portfolio, the time it takes to approximate all the VA policies in the portfolio, given the meshes, is less than 6 and 9 seconds, respectively. Even if we take into account the initial set up cost of constructing the green meshes and determining their prices and gradients, the total time it requires is still 634 and 326 times less than the benchmark, for $k = 500$ and 1000, respectively.

The saving in computational time relative to the direct valuation of large VAs portfolio is remarkable.

A more interesting comparison is to contrast the computational efficiency from our proposed green meshes to clustering-Kriging method. Recall that the implementation of the cluster-Kriging method basically involves four steps: (i) k-prototype clustering, (ii) nearest neighbouring of mapping representative synthetic VA policies to the actual VA policies, (iii) MC evaluation of representative VA policies, (iv) Kriging method of approximating all the VA policies within the portfolio. The respective time for these steps applying to the same two large VAs portfolios with $k = 1,000$ are reported in the upper panel of Table 5.7. The breakdown of these steps also provides useful insight to the time complexity of the clustering-Kriging method. The k-prototype clustering is the second most time-consuming step, leading to 174 and 351 seconds for VAs portfolios of 100,000 and 200,000 policies. The Kriging step, on the other hand, has the most demand on time. The Kriging approach of approximating all the 200,000 policies of the VAs portfolio would take about 7.14 hours and such computational effort is proportional to the size of the VAs portfolio. These results confirm that the clustering-Kriging method is not only memory intensive, but also computationally burdensome.

Along with the clustering-Kriging method, the lower panel of Table 5.7 reproduces from Table 5.6 the corresponding computational time but for the green meshes with $k = 1,000$. In this case, the start-up time is in the neighbourhood of 85 seconds and the evaluation time ranges from 4 to 11 seconds, depending on the sampling methods and the size of the VAs portfolio. Compared to the computational time of the clustering-Kriging method, the proposed green mesh is remarkably efficient, with a time saving of about 280 times. Another important difference that needs to be emphasized is that for the method of clustering-Kriging the representative VA policies depend on the VAs portfolio. Hence when we change the size of the VAs portfolio, the representative VA policies determined from the clustering approach need to be up-dated. On the other hand, this is not the case for the green mesh methods. As long as all the domains of the attributes remain unchange, the same set of

representative synthetic VA policies can be re-cycled repeatedly to approximate VA policies of any portfolio.

	$n = 100,000$		$n = 200,000$	
Benchmark	15,184		30,337	
MC-1024	1,564		3,166	
RQMC-1024	1,555		3,115	
Green Mesh	$k = 100$	$k = 500$	$k = 1000$	$k = 2000$
Start-Up Time	any n			
• Standard-Mesh-LHS	8.88	42.31	85.17	169.23
• Standard-Mesh-RQMC	9.25	42.34	85.94	172.77
• Conditional-Mesh-LHS	9.75	42.61	83.63	167.75
• Conditional-Mesh-RQMC	8.95	42.25	84.47	171.45
Evaluation Time	$n = 100,000; 200,000$			
• Standard-Mesh-LHS	1.91; 3.50	3.31; 6.56	5.38; 10.81	9.52; 19.80
• Standard-Mesh-RQMC	1.72; 3.41	2.98; 6.09	4.61; 9.55	7.88; 15.77
• Conditional-Mesh-LHS	1.89; 3.58	2.89; 5.91	4.39; 9.05	7.39; 15.38
• Conditional-Mesh-RQMC	1.83; 3.33	2.89; 5.88	4.42; 9.05	7.39; 15.27

Table 5.6: Computational time for the benchmark (MC with 10,000 trajectories), MC and RQMC with 1024 trajectories, and the green meshes with $k \in \{100, 500, 1000, 2000\}$ on large VAs portfolio of $n = 100,000$ and $200,000$ policies. For the green meshes we decompose the computational time into start-up time and evaluation time. All reported times are in seconds and run on laptop with Single Thread, 2.5 GHz CPU and Matlab program. The first and second entries of the evaluation time of the green meshes are for $100,000$ and $200,000$ VA policies, respectively.

	$n = 100,000$	$n = 200,000$
Clustering-Kriging ($k = 1,000$)		
(i) Clustering	174	351
(ii) Nearest neighbour mapping	5.92	5.97
(iii) MC evaluation	16	16
(iv) Kriging	12,886	25,709
Green Mesh ($k = 1,000$)		
Start-Up Time:		
• Standard-Mesh-LHS	85.17	85.17
• Standard-Mesh-RQMC	85.94	85.94
• Conditional-Mesh-LHS	83.63	83.63
• Conditional-Mesh-RQMC	84.47	84.47
Evaluation Time:		
• Standard-Mesh-LHS	5.38	10.81
• Standard-Mesh-RQMC	4.61	9.55
• Conditional-Mesh-LHS	4.39	9.05
• Conditional-Mesh-RQMC	4.42	9.05

Table 5.7: Comparison of computational time between Clustering-Kriging and the green meshes for $k = 1000$ on large VAs portfolio of $n = 100,000$ and $200,000$ policies. All reported times are in seconds and run on laptop with Single Thread, 2.5 GHz CPU and Matlab program.

5.5 Conclusion

We extended the idea of low discrepancy sequence in the QMC on selecting the representative synthetic variable annuity policies. The proposed green mesh method is not only able to approximate the large VAs portfolio with a high degree of precision but also a simple and real-time application. We discussed that our green mesh method needs only a little start-up cost and it's results can be re-cycled or re-used to effectively price other arbitrary large VAs portfolios. The downsize of the proposed method it does not apply to nested simulation.

Chapter 6

Dimension Reduction in Portfolio Selection

Key contributions of this chapter

1. Propose a new and novel portfolio strategy which we denote as the effective portfolio (EP) strategy
2. The EP strategy has the potential of providing better risk and reward trade-off on a given target portfolio.
3. If the target portfolio has N stocks, then the EP portfolio has the following characteristics:
   (a) it contains EPD stocks, where $EPD \ll N$;
   (b) all EPD stocks are selected from the original list of N stocks;
   (c) the EP portfolio has higher positive alpha and $\beta < 1$
4. An extensive empirical studies are conducted to demonstrate the efficiency and robustness of the EP strategy.

6.1 Introduction

In a typical portfolio selection,[1] the problem boils down to selecting appropriate assets to be included in the portfolio as well as their relative investment proportion within the portfolio. More specifically, we assume there are N risky stocks to be considered for investment and $R_i, i = 1, \ldots, N$, represents the i-th stock rate of return random variable over a single investment horizon. The i-th stock return's expected value and variance is given by $\mathbb{E}[R_i] = \mu_i$ and $Var(R_i) = \sigma_i^2 = \sigma_{ii}$, respectively, and the correlation between returns of stock i and stock j is denoted by $Corr(R_i, R_j) = \rho_{ij}, i, j = 1, \ldots, N$.

Let $w_i, i = 1, \ldots, N$, represent the portfolio weight in stock i. Positive w_i implies long position in stock i and negative w_i implies a short position. The resulting rate of return random variable R_P for a portfolio P constructed involving $\boldsymbol{R} = (R_1, \ldots, R_N)^T$ and with weight $\boldsymbol{w} = (w_1, \ldots, w_N)^T$ is given by

$$R_P = \sum_{i=1}^{N} w_i R_i = \boldsymbol{w}^T \boldsymbol{R}, \qquad \text{where } \sum_{i=1}^{N} w_i = 1. \tag{6.1}$$

The portfolio's expected return, μ_P, and its variance, σ_P^2, in turn, can be computed by

$$\mathbb{E}(R_p) = \mu_P = \sum_{i=1}^{N} w_i \boldsymbol{\mu}_i = \boldsymbol{w}^T \boldsymbol{\mu} \tag{6.2}$$

$$Var(R_P) = \sigma_P^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{i,j} = \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}, \tag{6.3}$$

with mean vector $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_N)^T$ and variance-covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1,\ldots,N}$, where $\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$, for $i \neq j$.

The above portfolio selection formulation assumes that the assets in the investment portfolio are all risky. It can be desirable to construct a portfolio that has some exposure to

[1] In this chapter N is defined as the number of stocks or the number of factors. In our context, N can be interpreted as the dimension of the problem of interest, though in other chapters we have consistently used d to denote the dimension.

the risk-free asset. By denoting r_f as the risk-free asset's deterministic return and w_0 as its portfolio weight, then the portfolio rate of return random variable (6.1) generalizes to

$$R_P = \sum_{i=0}^{N} w_i R_i = w_0 r_f + \boldsymbol{w}^T \boldsymbol{R} = \boldsymbol{w}^T \boldsymbol{R} + (1 - \boldsymbol{w}^T \mathbf{1}_N) r_f, \tag{6.4}$$

where $w_0 + \boldsymbol{w}^T \mathbf{1}_N = 1$ and $\mathbf{1}_N$ is an N-by-1 vector of ones. Similarly, the portfolio mean (6.2) and variance (6.3) become

$$\mathbb{E}(R_p) = \mu_P = w_0 r_f + \boldsymbol{w}^T \boldsymbol{\mu} = \boldsymbol{w}^T (\boldsymbol{\mu} - r_f \mathbf{1}_N) + r_f \tag{6.5}$$

$$Var(R_P) = \sigma_P^2 = \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}. \tag{6.6}$$

In the presence of risk-free asset, it is useful to express the return in excess of the risk-free return, so that (6.4) and (6.5), respectively, become

$$\tilde{R}_P = R_P - r_f = \boldsymbol{w}^T \boldsymbol{R} + \boldsymbol{w}^T \mathbf{1}_N r_f \tag{6.7}$$

$$\tilde{\mu}_P = \mu_P - r_f = \boldsymbol{w}^T (\boldsymbol{\mu} - r_f \mathbf{1}_N). \tag{6.8}$$

Throughout the chapter, we will consistently use the notation with "~" to denote its excess return.

For any given investment strategy $\boldsymbol{w}$, its risk and rewards are captured by (6.3) and (6.2) (or analogously by (6.6) and (6.5) in the presence of risk-free asset). An optimal portfolio strategy $\boldsymbol{w}$ is derived by formulating the portfolio selection problem as some optimization problem. The classical Markowitz (1952) model is a mean-variance efficient portfolio in the sense that the resulting portfolio yields the largest reward (as measured by the portfolio expected return) for a given level of portfolio risk exposure (as measured by the portfolio variance); or equivalently the smallest risk for a given level of reward.

The pioneering work of Markowitz (1952) forms the foundation of the well-known "Modern Portfolio Theory". It opens up a new line of research inquiry that is of interest to both

academics and practitioners. In particular, it is also well known that mean-variance efficient portfolios that are based on sample estimates of the first two moments perform poorly in out-of-sample. See Jobson and Korkie (1981), Frost and Savarino (1986), Michaud (1989), Black and Litterman (1990) and DeMiguel et. al. 2009. Subsequently a number of models attempting to address this issue has been proposed. Some of these extensions include imposing the shortsale constraint (see Frost and Savarino, 1988; Jagannathan and Ma, 2003), reducing the impact of estimation risk (see Kan and Zhou, 2007; Tu and Zhou, 2011), Bayesian shrinkage on covariance matrix (see Tibshirani, 1996; Ledoit and Wolf 2003, 2004, 2012), constraint on transaction cost (see Boyle and Lin, 1997; Olivares-Nadal and DeMiguel, 2015).

Another restriction under modern portfolio theory is on the pricing error. This includes parameter uncertainty, see Brown(1976), Klein and Bawa and (1976), Bawa et. al. (1979), Jorion (1986), Pástor (2000), Pástor and Stambaugh (2000), MacKinlay and Pástor (2000), Tu and Zhou (2004), Kan and Zhou (2007), and Tu and Zhou (2011), and model uncertainty, see Goldfarb and Iyengar (2003), Halldórsson and Tütüncü (2000), and Tütüncü and Koenig (2004). For the portfolio selection considering both parameter and model uncertainty, see Garlappi et al. (2007).

By resorting to an extensive empirical studies, DeMigueal at al. (2009) conclude that *"[o]f the 14 models [they] evaluate across seven empirical datasets, none is consistently better than the* $1/N$ *rule in terms of Sharpe ratio, certainty-equivalent return, or turnover".* A key reason for the "inefficiency" of the 14 sophisticated models is due to the estimation errors. This is also pointed out in Kritzman et al. (2010) that by relying on longer-term samples, the optimized portfolios may outperform $1/N$ in out-of-sample tests. However, the sophisticated optimization strategies rely heavily on the in-sample estimates, i.e. the "ex ante" perspective.

Another approach is to construct the portfolio based on economic values. In order to explain economic phenomena on asset prices and analyse relationships between variables,

Fama and French (1993) introduced the well known three-factor model. Following their work, Carhar (1997) extend the Fama-French three-factor model to include a momentum factor; Fama and French (2015) extends their own work to a five-factor pricing model; Stambaugh and Yuan(2015) incorporate the mispricing factors aggregating information across 11 prominent anomalies and produced a better-performing three-factor models.

We now provide a brief review of factor model. In general, a linear factor model assumes that the rate of return of i-th asset is given by

$$R_i = a_i + b_{i1}f_1 + b_{i2}f_2 + \cdots + b_{ik}f_k + \epsilon_i, \tag{6.9}$$

where f_j, $j = 1, \cdots, k$, are $k \geq 1$ random variables called factors, a_i is the expectation of R_i, b_{ij} is a constant and ϵ_i is called the "error" term with $E(\epsilon_i) = 0$ and $E(\epsilon_i f_j) = E(\epsilon_i)E(f_j) = 0$, for $i = 1, \cdots, N$. The factors themselves are allowed to be correlated and are meant to simplify and reduce the amount of randomness required in an analysis of our assets. When $k = 1$, this model is called a single factor model.

However the investable assets have the issue of misspecification. One form of misspecification is referred as alpha. Using the intuition behind the Treynor Index (Treynor (1965)) and Jensen's alpha (Jensen (1969)), Dybvig and Ross(1985) argues that "If the security market line tells us how much of a reward is justified for a given amount of risk, it makes intuitive sense that deviations from the security market line can be used to measure superior or inferior performance." Hence, they constructed a so called "alpha portfolio". See also Kan and Wang (2016).

Following Kan and Wang (2016), suppose $\tilde{\boldsymbol{R}} = [\tilde{\boldsymbol{R}}_{\boldsymbol{a}}, \tilde{\boldsymbol{R}}_{\boldsymbol{b}}]$, where $\tilde{\boldsymbol{R}}_{\boldsymbol{a}}$ and $\tilde{\boldsymbol{R}}_{\boldsymbol{b}}$ are the excess returns of the benchmark portfolios and the investing assets. We assume that $\tilde{\boldsymbol{R}}$ follows a multivariate normal distribution with mean

$$\tilde{\boldsymbol{\mu}} = \begin{bmatrix} \tilde{\boldsymbol{\mu}}_a \\ \tilde{\boldsymbol{\mu}}_b \end{bmatrix}$$

and covariance matrix of $\tilde{\boldsymbol{R}}$ as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix},$$

where a is the number of assets in $\tilde{\boldsymbol{R}}_{\boldsymbol{1}}$ and b is the number of assets in $\tilde{\boldsymbol{R}}_{\boldsymbol{2}}$. The investor is assumed to choose a portfolio P in order to maximize the mean-variance utility function which is $U = \mu_P - \frac{\gamma}{2}\sigma_P^2$, where γ is the investor's risk aversion coefficient, and μ_p and σ_p^2 are the mean and variance of portfolio P.

When the benchmarks are ex ante efficient, the investing assets will have zero alphas, i.e.

$$\boldsymbol{\alpha} = \tilde{\boldsymbol{\mu}}_b - \boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}\tilde{\boldsymbol{\mu}}_a = \mathbf{0}_N. \tag{6.10}$$

When the benchmarks are inefficient, the test assets will have nonzero alphas, i.e.

$$\boldsymbol{\alpha} = \tilde{\boldsymbol{\mu}}_b - \boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}\tilde{\boldsymbol{\mu}}_a \neq \mathbf{0}_N. \tag{6.11}$$

In this chapter, we attempt to address the following question. Given a portfolio strategy with investment in N stocks, is it possible to design a revised strategy with the following characteristics

(a) the revised portfolio contains only a stocks, where $a \ll N$;

(b) the a stocks are selected from the original list of N stocks;

(c) outperform the original portfolio strategy, possibly with a higher Jensen's alpha?

Property (a) ensures the revised portfolio is exposed to much less estimation errors relative to the original portfolio since it involves less stocks.

The answer to this question is yes. We accomplish this objective via the notion of effective portfolio dimension.

We make the following contributions in this paper. First, we show that only small number of stocks could dominate the whole market in terms of variance contribution. In other words, in order to enhance the portfolio performance, we could reduce the number of trading stocks without increasing the variance. Second, we show that we could construct our own factor models so that our portfolio could achieve any give beta value via dimension reduction. Third, we demonstrate that the dispersion of high performance stocks could be reduced when selecting smaller number of stocks. In other words, we could select preferred stocks or high-performance stocks which still reduce the variance of the portfolio. To be more exciting, we could construct our effective portfolio to achieve $\beta < 1$ and potential portfolio $\alpha > 0$. Fourth, we show that any sophisticated strategy could be combined with our effective portfolio strategy to enhance its performance relying on both statistical concepts and economic factors.

Section 6.2 introduces our effective portfolio strategy and the relationship to the factor model. In particular, Subsection 6.2.2 discusses the relationship between the selected and non-selected stocks, Subsection 6.2.3 addresses the question of determining effective portfolio dimension, and Subsection 6.2.4 proposes an algorithm on the ordering of stock preference. Then Section 6.3 provides ample of empirical results to support our proposed portfolio strategies.

6.2 Effective Dimension Portfolio Strategy

We begin the section by assuming that $\boldsymbol{R} = (R_1, R_2, \ldots, R_N)^T$ is multivariate normally distributed with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$ so that $\boldsymbol{R} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In other words we have

$$\boldsymbol{R} = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{Z}, \tag{6.12}$$

where $\boldsymbol{A}$ is an arbitrary decomposition matrix as long as it satisfies $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^T$ and $\boldsymbol{Z}$ is a vector of N standardized independent normally distributed random variables.

In Subsection 6.2.1 we consider a situation for which the N stocks are partitioned into two separate groups and provide a relationship between the stocks in one group and the stocks in another group. Subsection 6.2.3 discusses, effectively, the number of stocks should be included in the revised portfolio.

6.2.1 Stock Partitions

In this subsection we assume that the N stocks have been partitioned into two groups, with the first group consists of the a stocks and the second group consists of $b = N - a$ stocks. In term of the matrix notation, the representation (6.12) can be partitioned as follows:

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_a \\ \boldsymbol{R}_b \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix} + \begin{bmatrix} \boldsymbol{A}_{aa} & \boldsymbol{A}_{ab} \\ \boldsymbol{A}_{ba} & \boldsymbol{A}_{bb} \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}_a \\ \boldsymbol{Z}_b \end{bmatrix}, \tag{6.13}$$

where the subscript of each vector or matrix indicates their dimensions. As $\boldsymbol{A}$ can be arbitrarily selected as long as variance-covariance condition is satisfied, in what follows we impose an additional constraint on the decomposed $\boldsymbol{A}$ such that $\boldsymbol{A}_{ab}$ is a zero matrix.

For the partition displayed in (6.13), we assume that the stocks in the first group are simply the first a stocks from $\boldsymbol{R}$; i.e. $R_1, \ldots, R_a$ and the stocks in the second group are the remaining set of b stocks; i.e. $R_{a+1}, R_{a+2}, \ldots, R_N$. In order to cater to a general stock partitioning for which stocks from each group can be from any of the N stocks (instead of the first a stocks as in the first group described above), it is useful to introduce a permutation vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)^T$ of $\{1, \ldots, N\}$. The permutation vector $\boldsymbol{\pi}$ facilitates the partitioning of stocks so that $R_{\pi_1}, R_{\pi_2}, \ldots, R_{\pi_a}$ are assigned to the first group while $R_{\pi_{a+1}}, R_{\pi_{a+2}}, \ldots, R_{\pi_N}$ are allocated to the second group.

In order to distinguish the "permuted" notation from the "non-permuted" notation, the former notation is attached with a superscript $\boldsymbol{\pi}$. For example, the permuted N stocks is denoted by $\boldsymbol{R}^{\pi} = (R_{\pi_1}, R_{\pi_2}, \ldots, R_{\pi_N})^T$. The partition (6.13) is generalized to the following:

$$\begin{aligned} \boldsymbol{R}^{\boldsymbol{\pi}} = \begin{bmatrix} \boldsymbol{R}_a^{\boldsymbol{\pi}} \\ \boldsymbol{R}_b^{\boldsymbol{\pi}} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{\mu}_a^{\boldsymbol{\pi}} \\ \boldsymbol{\mu}_b^{\boldsymbol{\pi}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{A}_{aa}^{\boldsymbol{\pi}} & 0 \\ \boldsymbol{A}_{ba}^{\boldsymbol{\pi}} & \boldsymbol{A}_{bb}^{\boldsymbol{\pi}} \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}_a^{\boldsymbol{\pi}} \\ \boldsymbol{Z}_b^{\boldsymbol{\pi}} \end{bmatrix}. \\ &= \begin{bmatrix} \boldsymbol{\mu}_a^{\pi} + \boldsymbol{A}_{aa}^{\pi} \boldsymbol{Z}_a^{\pi} \\ \boldsymbol{\mu}_b^{\pi} + \boldsymbol{A}_{ba}^{\pi} \boldsymbol{Z}_a^{\pi} + \boldsymbol{A}_{bb}^{\pi} \boldsymbol{Z}_b^{\pi} \end{bmatrix}. \end{aligned} \tag{6.14}$$

Note that $\boldsymbol{R}_a^{\boldsymbol{\pi}} = \boldsymbol{\mu}_a^{\pi} + \boldsymbol{A}_{aa}^{\pi} \boldsymbol{Z}_a^{\pi}$ and hence $\boldsymbol{Z}_a^{\pi} = (\boldsymbol{A}_{aa}^{\pi})^{-1}(\boldsymbol{R}_a^{\pi} - \boldsymbol{\mu}_a^{\pi})$. This implies $\boldsymbol{R}_b^{\boldsymbol{\pi}}$ in (6.14) can be expressed as

$$\begin{aligned} \boldsymbol{R}_b^{\boldsymbol{\pi}} &= \boldsymbol{\mu}_b^{\pi} + \boldsymbol{A}_{ba}^{\pi} \boldsymbol{Z}_a^{\pi} + \boldsymbol{A}_{bb}^{\pi} \boldsymbol{Z}_b^{\pi} \\ &= \boldsymbol{\mu}_b^{\pi} + \boldsymbol{A}_{ba}^{\pi} (\boldsymbol{A}_{aa}^{\pi})^{-1} (\boldsymbol{R}_a^{\pi} - \boldsymbol{\mu}_a^{\pi}) + \boldsymbol{A}_{bb}^{\pi} \boldsymbol{Z}_b^{\pi} \\ &= \boldsymbol{A}_{ba}^{\pi} (\boldsymbol{A}_{aa}^{\pi})^{-1} \boldsymbol{R}_a^{\pi} + \boldsymbol{\theta}^{\pi} \end{aligned} \tag{6.15}$$

where $\boldsymbol{\theta}^{\pi} = \boldsymbol{\mu}_b^{\pi} - \boldsymbol{A}_{ba}^{\pi} (\boldsymbol{A}_{aa}^{\pi})^{-1} \boldsymbol{\mu}_a^{\pi} + \boldsymbol{A}_{bb}^{\pi} \boldsymbol{Z}_b^{\pi}$.

The relationship (6.15) demonstrates that returns in $\boldsymbol{R}_b^{\boldsymbol{\pi}}$ are the sum of the returns in $\boldsymbol{R}_a^{\boldsymbol{\pi}}$ (adjusted by $\boldsymbol{A}_{ba}^{\pi} (\boldsymbol{A}_{aa}^{\pi})^{-1}$) and the residual component $\boldsymbol{\theta}^{\pi}$. The residual component is independent of $\boldsymbol{R}_a^{\pi}$ with its mean and variance given by $\boldsymbol{\mu}_b^{\pi} - \boldsymbol{A}_{ba}^{\pi} (\boldsymbol{A}_{aa}^{\pi})^{-1} \boldsymbol{\mu}_a^{\pi}$ and $\boldsymbol{A}_{bb}^{\pi} (\boldsymbol{A}_{bb}^{\pi})^T$, respectively. More importantly, this result can also be interpreted as providing a way of replacing stocks in $\boldsymbol{R}_b^{\boldsymbol{\pi}}$ from the stocks in $\boldsymbol{R}_a^{\boldsymbol{\pi}}$. This result plays an important role in our portfolio strategy, as discussed in the next subsection.

6.2.2 Revised Portfolio: Selected Stocks vs Non-selected Stocks

We now assume that $\boldsymbol{w}$ is the original portfolio strategy constructed from N stocks. Based on the given portfolio strategy, our objective is to devise a new portfolio strategy that

uses only a subset of the original N stocks. Ideally our revised portfolio strategy should outperform the original portfolio strategy $\boldsymbol{w}$. In this aspect, the original portfolio strategy $\boldsymbol{w}$ can be interpreted as the benchmark or target portfolio and we use TP to denote the resulting target portfolio. The objective is therefore to outperform the target portfolio TP constructed by the portfolio weight $\boldsymbol{w}$.

Our revised portfolio strategy consists of a stocks and the risk-free asset. We assume that the stocks can only be selected from the original N stocks of the target portfolio TP. Which of the a stocks to be selected is dictated by the permutation vector $\boldsymbol{\pi}$. For this reason, the permutation vector $\boldsymbol{\pi}$ can be interpreted as the order of preference for investing among the N stocks; with the π_1-th stock being the most preferred and the π_N-th stock being the least. Depending on the risk attitude of the investor, the stock preference order can be based on the expected return, variance, or some other statistics or attributes of the stocks. In Section 6.2.4 we will propose some plausible ways of determining $\boldsymbol{\pi}$. For now we will simply refer the permutation vector $\boldsymbol{\pi}$ as the stock picking vector or rule.

Using the same notation as convention on the stock partition and the permutation, we have

$$\boldsymbol{w} = \begin{bmatrix} \boldsymbol{w}_a \\ \boldsymbol{w}_b \end{bmatrix}, \qquad \boldsymbol{w}^{\pi} = \begin{bmatrix} \boldsymbol{w}_a^{\pi} \\ \boldsymbol{w}_b^{\pi} \end{bmatrix}, \qquad \boldsymbol{\pi}^T = \begin{bmatrix} \boldsymbol{\pi}_a \\ \boldsymbol{\pi}_b \end{bmatrix}.$$

Let Q be the revised portfolio strategy consisting of a selected stocks according to the stock picking rule $\boldsymbol{\pi}$ and the risk-free asset r_f. This implies that the stocks in $\boldsymbol{R}_a^{\boldsymbol{\pi}}$ are the "selected" stocks and the stocks in $\boldsymbol{R}_b^{\boldsymbol{\pi}}$ are the "non-selected" stocks. By denoting R_Q as the return random variable of the revised portfolio strategy with the corresponding weight $\boldsymbol{w}_Q^{\boldsymbol{\pi}}$ for the selected stocks, we have

$$R_Q = (\boldsymbol{w}_Q^{\boldsymbol{\pi}})^T \boldsymbol{R}_a^{\boldsymbol{\pi}} + [1 - (\boldsymbol{w}_Q^{\boldsymbol{\pi}})^T \mathbf{1}_a] r_f, \tag{6.16}$$

where $\mathbf{1}_a$ is a $a \times 1$ vector of ones and $\boldsymbol{w}_Q^{\boldsymbol{\pi}}$ is a column vector of a dimension.

To fully specify our proposed revised portfolio Q we have yet to determine the portfolio

weight $\boldsymbol{w}_Q^{\boldsymbol{\pi}}$. We propose to construct $\boldsymbol{w}_Q^{\boldsymbol{\pi}}$ based on the following formula:

$$\boldsymbol{w}_Q^{\boldsymbol{\pi}} = \boldsymbol{w}_a^{\boldsymbol{\pi}} + (\boldsymbol{A}_{ba}^{\pi}(\boldsymbol{A}_{aa}^{\pi})^{-1})^T \boldsymbol{w}_b^{\boldsymbol{\pi}}. \tag{6.17}$$

It is important to point out that the portfolio Q depends on the original portfolio strategy $\boldsymbol{w}$, the stock picking rule $\boldsymbol{\pi}$, as well as the decomposition matrix $\boldsymbol{A}$.

The adjusted portfolio weight based on (6.17) can be justified from a linear regression model. To see this, let $\tilde{R}_{aP}^{\boldsymbol{\pi}}$ be the excess portfolio return of the selected stocks with excess return vector $\tilde{\boldsymbol{R}}_a^{\boldsymbol{\pi}}$ and portfolio weight $\boldsymbol{w}_a^{\pi}$. Analogously, let $\tilde{R}_{bP}^{\boldsymbol{\pi}}$ be the excess portfolio return of the non-selected stocks with excess return vector $\tilde{\boldsymbol{R}}_b^{\boldsymbol{\pi}}$ and portfolio weight $\boldsymbol{w}_b^{\pi}$. Now it is possible to assume that the selected stocks and the non-selected stocks are related as follows:

$$\tilde{R}_{bP}^{\boldsymbol{\pi}} = (\boldsymbol{w}_b^{\pi})^T \tilde{\boldsymbol{R}}_b^{\boldsymbol{\pi}} = \alpha_b^{\boldsymbol{\pi}} + (\boldsymbol{\beta}_b^{\boldsymbol{\pi}})^T \tilde{\boldsymbol{R}}_a^{\boldsymbol{\pi}} + \epsilon_b^{\pi}, \tag{6.18}$$

where $\alpha_y^{\boldsymbol{\pi}}$ is a drift term, $\boldsymbol{\beta}_a^{\boldsymbol{\pi}}$ is a a by 1 coefficient vector, ϵ_b^{π} is an error term with mean 0, variance Ω_b^{π}, and is independent of $\tilde{\boldsymbol{R}}_a^{\boldsymbol{\pi}}$. The coefficient vector $\boldsymbol{\beta}_b^{\boldsymbol{\pi}}$ can be derived as the least square estimator of the standard multiple regression model as

$$(\boldsymbol{\beta}_b^{\pi})^T = Cov(\tilde{R}_{bP}^{\boldsymbol{\pi}}, \tilde{\boldsymbol{R}}_a^{\boldsymbol{\pi}})(\boldsymbol{\Sigma}_{aa}^{\pi})^{-1} = (\boldsymbol{w}_b^{\pi})^T \boldsymbol{\Sigma}_{ba}^{\pi}(\boldsymbol{\Sigma}_{aa}^{\pi})^{-1} = (\boldsymbol{w}_b^{\pi})^T \boldsymbol{A}_{ba}^{\pi}(\boldsymbol{A}_{aa}^{\pi})^{-1}. \tag{6.19}$$

For the proof of $\boldsymbol{\Sigma}_{ba}^{\pi}(\boldsymbol{\Sigma}_{aa}^{\pi})^{-1} = \boldsymbol{A}_{ba}^{\pi}(\boldsymbol{A}_{aa}^{\pi})^{-1}$, see Appendix *A*.

Consequently the portfolio strategy (6.17) corresponds to the sum of the two weights

(1) original weight of the selected stocks ($\boldsymbol{w}_a^{\boldsymbol{\pi}}$), and

(2) weight given by the beta coefficient $\boldsymbol{\beta}_b^{\boldsymbol{\pi}}$ (6.19).

In other words,

$$\boldsymbol{w}_Q^{\pi} = \boldsymbol{w}_a^{\boldsymbol{\pi}} + (\boldsymbol{A}_{ba}^{\pi}(\boldsymbol{A}_{aa}^{\pi})^{-1})^T \boldsymbol{w}_b^{\boldsymbol{\pi}} = \boldsymbol{w}_a^{\pi} + \boldsymbol{\beta}_b^{\pi}. \tag{6.20}$$

The revised portfolio Q based on (6.20) can also be interpreted in the following way. The portfolio Q is similar to the target portfolio except that the non-selected assets of the target portfolio are being linearly replaced by the selected assets. As a result, the proportion of investment in the selected stocks is increased by $\boldsymbol{\beta}_b^\pi$.

An additional insight on the return of the revised portfolio Q can be gleaned by considering the following decomposition, which follows from (6.20):

$$\tilde{\mu}_Q = \mathbb{E}[\tilde{R}_Q] = (\boldsymbol{w}_Q^\pi)^T \tilde{\boldsymbol{\mu}}_a^\pi = (\boldsymbol{w}_a^\pi + \boldsymbol{\beta}_b^\pi)^T \tilde{\boldsymbol{\mu}}_a^\pi = (\boldsymbol{w}_a^\pi)^T \tilde{\boldsymbol{\mu}}_a^\pi + (\boldsymbol{\beta}_b^\pi)^T \tilde{\boldsymbol{\mu}}_a^\pi. \tag{6.21}$$

The above decomposition demonstrates that the excess expected return of the portfolio Q arises from two sources. The first source is attributed to the excess expected return produced by the "selected" stocks of the target portfolio ($(\boldsymbol{w}_a^\pi)^T \tilde{\boldsymbol{\mu}}_a^\pi$). The second source is the incremental excess expected return for having additional exposure in the selected stocks or by replacing the "non-selected" stocks with the "selected" stocks ($(\boldsymbol{\beta}_b^\pi)^T \tilde{\boldsymbol{\mu}}_a^\pi$).

An even more important insight is the relationship between the revised portfolio Q and the Jensen's alpha. In our context the Jensen's alpha is defined as the difference between the excess expected return of a portfolio and the excess expected return of the target portfolio.[2] Hence we have

$$\begin{aligned}
\text{Jensen's Alpha} &= \tilde{\mu}_Q - \tilde{\mu}_{TP} \\
&= (\boldsymbol{w}_Q^\pi)^T \tilde{\boldsymbol{\mu}}_a^\pi - \boldsymbol{w}^T \tilde{\boldsymbol{\mu}} \\
&= (\boldsymbol{w}_a^\pi + \boldsymbol{\beta}_b^\pi)^T \tilde{\boldsymbol{\mu}}_a^\pi - \boldsymbol{w}^T \tilde{\boldsymbol{\mu}} \\
&= (\boldsymbol{\beta}_b^\pi)^T \tilde{\boldsymbol{\mu}}_a^\pi - (\boldsymbol{w}_b^\pi)^T \tilde{\boldsymbol{\mu}}_b^\pi \\
&= (\boldsymbol{w}_b^\pi)^T \boldsymbol{\Sigma}_{ba}^\pi (\boldsymbol{\Sigma}_{aa}^\pi)^{-1} \tilde{\boldsymbol{\mu}}_a^\pi - (\boldsymbol{w}_b^\pi)^T \tilde{\boldsymbol{\mu}}_b^\pi \\
&= (\boldsymbol{w}_b^\pi)^T \left[\boldsymbol{\Sigma}_{ba}^\pi (\boldsymbol{\Sigma}_{aa}^\pi)^{-1} \tilde{\boldsymbol{\mu}}_a^\pi - \tilde{\boldsymbol{\mu}}_b^\pi \right] \\
&= -\alpha_b^{\boldsymbol{\pi}},
\end{aligned} \tag{6.22}$$

[2] It should be pointed out that the original definition of the Jensen's alpha assumes CAPM so that the target portfolio is the market portfolio.

where $\alpha_b^{\boldsymbol{\pi}} = (\boldsymbol{w}_b^{\pi})^T \tilde{\boldsymbol{\mu}}_b^{\pi} - (\boldsymbol{\beta}_b^{\boldsymbol{\pi}})^T \tilde{\boldsymbol{\mu}}_a^{\boldsymbol{\pi}} = (\boldsymbol{w}_b^{\pi})^T \left[\tilde{\boldsymbol{\mu}}_b^{\pi} - \boldsymbol{\Sigma}_{ba}^{\pi}(\boldsymbol{\Sigma}_{aa}^{\pi})^{-1}\tilde{\boldsymbol{\mu}}_a^{\pi}\right]$ is obtained from (6.18) after taking expectation. Interestingly, we can prove that the effective portfolio could achieve smaller variance and higher expected return when the selected and non-selected stocks have different correlation and the number of selected stocks is not too small. (See Appendix *E*)

This result provides useful guidance on how the stocks should be selected to the portfolio Q so that the resulting Jensen's alpha is as positive as possible. Ideally, the selected stocks should possess the following properties:

1. Partition the selected and non-selected stocks in such a way that the intercept (i.e. $\alpha_b^{\boldsymbol{\pi}}$) in the multiple regression model (6.18) is as negative as possible.
2. The excess expected return induced by the beta portfolio is larger than the corresponding excess expected return of the non-selected stocks of the target portfolio (i.e. $(\boldsymbol{\beta}_b^{\pi})^T \tilde{\boldsymbol{\mu}}_a^{\pi} > (\boldsymbol{w}_b^{\pi})^T \tilde{\boldsymbol{\mu}}_b^{\pi}$)
3. For positive portfolio weight, the selected stocks should have low variance and/or high expected returns while the non-selected stocks should have low expected returns. This follows from wishing to have $\boldsymbol{\Sigma}_{ba}^{\pi}(\boldsymbol{\Sigma}_{aa}^{\pi})^{-1}\tilde{\boldsymbol{\mu}}_a^{\pi} - \tilde{\boldsymbol{\mu}}_b^{\pi}$ as positive as possible.

By integrating some of these recommendations, Section 6.2.4 will propose some algorithms for constructing the revised portfolio Q.

We make the following two remarks before concluding this subsection.

Remark 6.2.1. *According the model of Kan and Wong (2016), our constructed Jensen's alpha is the negative value of weighted value of alphas of non-selected assets with respect to the selected assets. This can be seen by noting that* $\boldsymbol{\alpha}$ *from* (6.18) *is equivalent to the "unweighted" negative* $\alpha_b^{\boldsymbol{\pi}}$ *of* (6.22).

Remark 6.2.2. *Our discussions so far have assumed that* $\boldsymbol{R}$ *is a collection of* N *stocks that are investable. Our results still apply if each of* $R_j, j = 1, \ldots, N$ *is treated as factor. In particular, by setting* $k = a$ *and* $f_j = (\tilde{\boldsymbol{R}}_a^{\boldsymbol{\pi}})_j$*, according to* (6.9), (6.18) *corresponds to*

a factor model where the factors are the selected assets. Hence the selection of stocks is transformed to the selection of factors. For this reason, we can equivalently view the stocks as the stock factors. An important difference these two interpretations is that if it is a stock factor, then it is not investable, except when the factor is a market portfolio and we assume that the stock index, which is a proxy for the market portfolio, is investable.

6.2.3 Effective Portfolio Dimension (EPD)

So far we have argued that by an appropriate choice of stocks selected from a list of N stocks, it is possible to produce a portfolio strategy Q that outperforms the target portfolio with a higher Jensen's alpha. As also pointed out in the introduction that one of the key issues associated with the sophisticated portfolio strategies is the need to estimate their parameters for implementing these strategies in practice. Hence these strategies are subject to estimation errors. Because of the estimation errors, their empirical out-of-sample performance may not be as efficient as compared to the theoretical efficiency. The estimation errors become more severe on portfolio with more underlying stocks.

This results in a conflicting dilemma in the sense that if one is interested in reducing estimation error, then the portfolio should contain as few stocks as possible (such as investing in the "familiar assets") but this increases the so-called concentration risk. On the other hand, Markowitz advocates diversification. This implies that the portfolio should include as many stocks as possible but at the expense of estimation errors. More discussion on the tradeoff between the concentration (i.e. familiarity) and diversification, see Boyle, et al. (2012).

Likewise our proposed strategy faces the similar tradeoff between concentration and diversification. We begin with a target portfolio of N stocks. Our revised strategy, however, comprises of a stocks that are selected from the given N stocks. What should the appropriate value of a be? If a is close to N, then the revised portfolio Q will be close to the target portfolio but if $a \ll N$, then it subjects to concentration risk, though the estimation

errors also reduced substantially.

To resolve this tradeoff, we relate an optimal choice of a to a concept known as the *effective portfolio dimension (EPD)*. Let $D^{\boldsymbol{\pi}}_{\boldsymbol{w}}$ denote a dimension random variable of a portfolio strategy with a portfolio weight and a stock picking rule $\boldsymbol{w}$. Its cumulative mass function, $F_{\boldsymbol{D}^{\pi}_{\boldsymbol{w}}}(j)$, is defined as follows:

$$F_{\boldsymbol{D}^{\pi}_{\boldsymbol{w}}}(j) = \Pr(D^{\boldsymbol{\pi}}_{\boldsymbol{w}} \leq j) = \frac{(\boldsymbol{w}^{\boldsymbol{\pi}})^T \boldsymbol{A}^{\boldsymbol{\pi}}_{\cdot 1:j} (\boldsymbol{A}^{\boldsymbol{\pi}}_{\cdot 1:j})^T (\boldsymbol{w}^{\boldsymbol{\pi}})^T}{\boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}}, \qquad j = 1, \ldots N, \tag{6.23}$$

where $\boldsymbol{A}^{\boldsymbol{\pi}}_{\cdot 1:j}$ is the first j columns of $\boldsymbol{A}^{\boldsymbol{\pi}}$. Intuitively, $F_{\boldsymbol{D}^{\pi}_{\boldsymbol{w}}}(j)$ captures the relative proportion of variance contributed by the first j stocks selected for a given portfolio rule $\boldsymbol{w}$ and a permutation vector $\boldsymbol{\pi}$. It is easy to see that $F_{\boldsymbol{D}^{\pi}_{\boldsymbol{w}}}(j)$ is a non-decreasing function in $j \in [1, N]$.

The dimension distribution of $D^{\boldsymbol{\pi}}_{\boldsymbol{w}}$ enables us to define the effective portfolio dimension of a portfolio, as shown below. Note the similarity between the effective portfolio dimension and the delta dimension defined in Chapter 2.

Definition 6.2.1. *Given a portfolio weight $\boldsymbol{w}$ and a stock picking rule $\boldsymbol{\pi}$, the effective portfolio dimension (EPD) of the portfolio at a confidence level $\beta^* \in (0, 1)$, denoted by $EPD^{\boldsymbol{\pi}}_{\beta^*}$, is defined as*

$$EPD^{\boldsymbol{\pi}}_{\beta^*} = \inf\{k \in [1, \cdots, N] : F_{\boldsymbol{D}^{\pi}_{\boldsymbol{w}}}(j) \leq k) \geq \beta^*\}. \tag{6.24}$$

Typically β^* is close to 1. Intuitively, $EPD^{\boldsymbol{\pi}}_{\beta^*}$ measures the smallest number of stocks that captures at least β^* proportion of the total portfolio variance. If $EPD^{\boldsymbol{\pi}}_{\beta^*} \ll N$, then the portfolio is effectively dominated by $EPD^{\boldsymbol{\pi}}_{\beta^*}$ stocks even though the portfolio contains a large number of stocks. For example, suppose $N = 100$ and $EPD^{\boldsymbol{\pi}}_{\beta^*} = 5$ at $\beta^* = 99\%$. Then a portfolio constructed based on $\boldsymbol{w}$ and $\boldsymbol{\pi}$ implies that five of the stocks captures at least 99% of the total portfolio variance. This also suggests that the stocking picking rule $\boldsymbol{\pi}$ is extremely effective at reducing the dimension of the portfolio from 100 stocks to

effectively only 5 stocks. For this reason an optimal a can be chosen as one that corresponds to $EPD^{\boldsymbol{\pi}}_{\beta^*}$ for β^* close to 1. We then relabel the resulting portfolio from Q to effective portfolio (EP).

Suppose β_Q is the beta of the portfolio Q relative to the target portfolio TP, then analogous to the CAPM, we have

$$\beta_Q = \frac{Cov(R_Q, R_{TP})}{Var(R_{TP})}, \tag{6.25}$$

where R_{TP} is the return random variable of the target portfolio. The following theorem establishes an important connection between β_Q and $F_{\boldsymbol{D}^{\pi}_{\boldsymbol{w}}}(j)$.

Theorem 6.2.1. *Let R_{TP} be the return random variable of a target portfolio with portfolio weight $\boldsymbol{w}$. For a given a and a stock picking rule $\boldsymbol{\pi}$, let R_Q be the return random variable of a revised portfolio constructed from the target portfolio according to (6.20). Then we have*

$$F_{\boldsymbol{D}^{\pi}_{\boldsymbol{w}}}(a) = \beta_Q, \tag{6.26}$$

where the cumulative mass function $F_{\boldsymbol{D}^{\pi}_{\boldsymbol{w}}}(\cdot)$ and β_Q are defined in (6.23) and (6.25), respectively.

Proof: For arbitrary $\boldsymbol{w}$, according to (6.15),

$$R_{TP} = \boldsymbol{w}^T \boldsymbol{R} + \left(1 - \mathbf{1}^T \boldsymbol{w}\right) r_f = (\boldsymbol{w}^{\pi})^T \begin{bmatrix} \boldsymbol{R}^{\pi}_a \\ \boldsymbol{A}^{\pi}_{ba}(A^{\pi}_{aa})^{-1}\boldsymbol{R}^{\pi}_a + \theta^{\pi} \end{bmatrix} + \left(1 - \mathbf{1}^T \boldsymbol{w}^{\pi}\right) r_f.$$

Similarly, according to (6.17),

$$R_Q = (\boldsymbol{w}^{\pi})^T \begin{bmatrix} \boldsymbol{R}^{\pi}_a \\ \boldsymbol{A}^{\pi}_{ba}(A^{\pi}_{aa})^{-1}\boldsymbol{R}^{\pi}_a \end{bmatrix} + \left(1 - \mathbf{1}^T \begin{bmatrix} \boldsymbol{w}^{\pi}_a \\ (\boldsymbol{A}^{\pi}_{ba}(\boldsymbol{A}^{\pi}_{aa})^{-1})^T \boldsymbol{w}^{\pi}_b \end{bmatrix}\right) r_f,$$

where $\boldsymbol{R}^{\pi}_a$ is independent of θ^{π} and the risk-free rate r_f is considered to be a constant.

Hence,

$$Cov(R_Q, R_{TP}) = Var(R_Q) = (\boldsymbol{w}^\pi)^T \begin{bmatrix} \boldsymbol{A}^\pi_{aa} \\ \boldsymbol{A}^\pi_{ba} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}^\pi_{aa} \\ \boldsymbol{A}^\pi_{ba} \end{bmatrix}^T \boldsymbol{w}^\pi = (\boldsymbol{w}^\pi)^T \boldsymbol{A}^{\boldsymbol{\pi}}_{\cdot 1:a} (\boldsymbol{A}^{\boldsymbol{\pi}}_{\cdot 1:a})^T \boldsymbol{w}^\pi,$$

where $(\boldsymbol{w}^\pi)\boldsymbol{A}^{\boldsymbol{\pi}}_{\cdot 1:a}(\boldsymbol{A}^{\boldsymbol{\pi}}_{\cdot 1:a})^T\boldsymbol{w}^\pi$ is the numerator of (6.23) and $Var(R_{TP})$ is the denominator of (6.23). Hence, the cumulative distribution function in (6.23) could be reconstructed as

$$\Pr(D^{\boldsymbol{\pi}}_{\boldsymbol{w}} \leq a) = \frac{Cov(R_Q, R_{TP})}{Var(R_{TP})}. \tag{6.27}$$

Note that $\Pr(D^{\boldsymbol{\pi}}_{\boldsymbol{w}} \leq a)$ equals to the market beta if $\boldsymbol{w}$ is equally weighted. i.e. the target portfolio is R_M.

6.2.4 Determining Stock Picking Rule

The preceding subsection provided a way of determining a via the notion of effective portfolio dimension. This measure depends on a given stock picking rule; i.e. permutation vector $\boldsymbol{\pi}$. Hence to complete the description of the effective portfolio strategy, we need to have a way of identifying $\boldsymbol{\pi}$. Recall that $\boldsymbol{\pi}$ represents the order of preference among the N stocks. This can depend on investor's risk attitude, or the stocks' various attributes such as the expected returns, variances, etc. In other words, there is no universal optimal way of determining $\boldsymbol{\pi}$. Ideally the stocks that to be selected should lead to a portfolio with high Jensen's alpha. In fact from (6.22) we have already provided some guidelines on appropriate choices of $\boldsymbol{\pi}$. In this section, we provide a specific procedure of determining $\boldsymbol{\pi}$. The empirical evidence to be provided in the next section indicates that our proposed algorithm is quite effective.

Assume that we are given an initial set of N stocks. The following algorithm provides a way of determining $\boldsymbol{\pi}$

Algorithm 6.2.1. *(Goal: Constructing stock picking rule $\boldsymbol{\pi}$)*

Step 1: *Pre-screening N stocks to produce a good candidate set of N^* stocks. Let ζ denotes the resulting candidate set.*

Step 2: *Sort the stocks in ζ according to the variance from smallest to largest. Assign $\boldsymbol{\pi}$ so that π_j-th stock has the j-th smallest variance among all the stocks in ζ.*

We now provide some rationale for the above algorithm 6.2.1. Let us first note that the above algorithm determines $\boldsymbol{\pi}$ via a two-step procedure. The first step narrows down the feasible set of stocks from N to N^*. After pre-screening, $\boldsymbol{\pi}$ is then defined explicitly based on the magnitude of the variances. In theory these two steps can obviously be combined into a single step so that the constructed $\boldsymbol{\pi}$ reflects both selection criteria in Steps 1 and 2. Nevertheless we choose to present the algorithm involving a two-step procedure to simplify the explanation.

The objective of the pre-screening step is to ensure that the stocks in ζ at least have some desirable properties. For example, if the pre-screening criterion corresponds to the expected return of at least 8%, then the stocks in ζ are guarantee to have 8% expected return. Then in Step 2 we are only allowed to define π from the smaller set of N^* stocks, instead from the original list of N stocks.

In our empirical studies, ζ corresponds to the top 70% of the momentum stocks. This is supported by the numerous empirical evidences that stocks that had relatively high returns over the past years tend to give above average returns over the next relative short period. This is attributed to the concept of momentum stocks (see Jegadeesh and Titman. 1993; Li, et al., 2008, 2009; Beker and Espino, 2013) Another reason for considering top 70% is due to Appendix *E*, i.e. the inequality will hold if N^* is not too small. Also, some literature pointed out the poor performance of bottom 30% momentum stocks.

Let us now consider Step 2. Conditioned on ζ, stocks with low variance have high priority to be selected. This can be justified from (6.22) that high Jensen's alpha can be expected from selecting stocks with low variance. An alternate justification is from the regression model. Recall that the regression model (6.18) which relates the "non-selected" stocks to

the "selected" stocks. The coefficient of determination (R^2) is defined

$$R^2(\tilde{R}^{\pi}_{bP}) := R^2(w^{\pi}_b R^{\pi}_b) = \frac{(\boldsymbol{w}^{\pi}_b)^T \boldsymbol{A}^{\pi}_{ba}(\boldsymbol{A}^{\pi}_{ba})^T w^{\pi}_b}{(w^{\pi}_b)^T \boldsymbol{A}^{\pi}_{ba}(\boldsymbol{A}^{\pi}_{ba})^T \boldsymbol{w}^{\pi}_b + \Omega^{\pi}_b},$$

where Ω^{π}_b is the residual variance for non-selected stocks with respect to selected stocks defined in (6.18). The denominator is the total variance of weighted average of the non-selected stocks in the target portfolio, and the numerator is the variance explained by the selected stocks. Hence, we would like Ω^{π}_b to be as small as possible or $(\boldsymbol{w}^{\pi}_b)^T \boldsymbol{A}^{\pi}_{ba}(\boldsymbol{A}^{\pi}_{ba})^T \boldsymbol{w}^{\pi}_b$ to be as large as possible so that $R^2(\boldsymbol{w}^{\pi}_b R^{\pi}_b)$ approaches to 1.

By our model set-up, we have the following property

Proposition 6.2.1. (coefficient of determination $R^2(\boldsymbol{w}^{\pi}_b R^{\pi}_b)$) *Under effective portfolio, maximizing the coefficient of determination is equivalent to choosing non-selected stocks with higher variance; i.e. selected stocks with smaller variance.*

The proof is provided in Appendix B.

The above proposition implies that to produce a portfolio that has as high the lower bound of the coefficient of determination as possible, an optimal strategy is to ensure that the variance of the selected stocks is as small as possible.

Once $\boldsymbol{\pi}$ is determined from Algorithm 6.2.1, then for a given confidence level β^*, the stocks that are selected are given by

$$\{R_{\pi_1}, R_{\pi_2}, \ldots, R_{\pi_{EPD}}\}$$

where $EPD = EPD^{\boldsymbol{\pi}}_{\beta^*}$ and is defined from (6.24). This is our proposed EP effective portfolio strategy and its return random variable is denoted by R_{EP}. From the definition of EPD, it is easy to see that

$$\beta^* Var(R_{TP}) \;\leq\; Var(R_{EP}) \;\leq\; Var(R_{TP}). \tag{6.28}$$

6.3 Empirical Studies

In this section we provide an extensive empirical studies to demonstrate the efficiency of our proposed EP portfolio strategies. Subsection 6.3.1 describes the setup of our empirical studies, including the portfolio strategies that we will benchmark against. Subsection 6.3.2 then describes the measures that will be used in gauging the efficiency of the portfolio strategies. The empirical results will be presented in Subsection **??**.

6.3.1 Empirical Studies Setup and Methodology

Following DeMiguel et al. (2009), our goal in this section is to empirically assess the relative effectiveness of the various asset allocation strategies by applying a variety of datasets. These datasets have been considered extensively in the literature. Our analysis relies on a "rolling-sample" approach. Specifically, given a T-month-long dataset of asset returns, we choose an estimation window of length $M = 10 \times 12$ months with out-of-sample being 1 month. In each month t, starting from $t = M + 1$, we use the data in the previous M months to estimate the parameters needed to implement a particular strategy. These estimated parameters are then used to determine the relative portfolio weights in the portfolio. We then use these weights to compute the return in month $t + 1$. This process is continued by adding the return for the next period in the dataset and dropping the earliest return, until the end of the dataset is reached. The outcome of this rolling-window approach is a series of $T - M$ monthly out-of-sample returns generated by each of the portfolio strategies we are investigating.

For the estimation of mean and variance-covariance matrix of monthly returns, we use in-sample mean and variance-covariance matrix. For robust check, we have also included the in-sample robust estimation of 120 data points using Jorion (1986) and Ledoit and Wolf (2003), and in-sample mean and covariance matrix using 240 data points demonstrated in Appendix C. The risk-free rate is according to the 90-day T-bill rate download from Ken

French Website.

As a brief review of portfolio investment methodology, we use R_t^m to denote the m-vector of excess returns (over the risk-free assets) on m risky assets available for investment at date t. The number of stocks $m = N$ is called as nominal dimension, and $m = EPD$ is called as the effective portfolio dimension. The m-dimensional vectors $\boldsymbol{\mu}_t^m$ and $\boldsymbol{\mu}_t^m$ is used to denote the expected returns on the risky asset in excess of the risk-free rate, and $\boldsymbol{\Sigma}_t^m$ is denoted as the corresponding $m \times m$ variance-covariance matrix of returns, with their sample counterparts given by $\hat{\mu}_t^m$ and $\hat{\Sigma}_t^m$, respectively. Let M be denoted as the length over which these moments are estimated and T as the total length of the data series. We use $\mathbf{1}_m$ to define an m-dimensional vector of ones, and $\boldsymbol{I}_m$ for the $m \times m$ identity matrix. Furthermore, x_t^m is the vector of portfolio weights invested in the m risky assets, with $1 - \mathbf{1}_m^T x_t^m$ invested in the risk-free assets. The vector of relative weights in the portfolio with only-risky assets is

$$w_t^m = \frac{x_t^m}{|\mathbf{1}_m^T x_t^m|},$$

where the normalization by the absolute value of the sum of the portfolio weights, $|\mathbf{1}_m^T x_t^m|$, guarantees that the direction of the portfolio position is preserved in the few cases where the sum of the weights on the risky assets is negative.

In general, when $m = N$, we consider an investor whose preferences are fully described by the mean and variance of a chosen portfolio, x_t. At each time t, the decision-maker selects x_t to maximize expected utility:

$$\max_{x_t} x_t^T \mu_t - \frac{\gamma}{2} x_t^T \Sigma_t x_t, \tag{6.29}$$

where γ is interpreted as the investor's risk aversion. The solution of the above optimization is $x_t = (1/\gamma)\Sigma_t^{-1}\mu_t$.

We now describe the portfolio strategies that will be used in our empirical studies. From these strategies, we will then construct the effective portfolio strategies.

Naive portfolio

The naive portfolio or equally weighted portfolio ('1/N' or 'EWP') that we consider involves holding a portfolio weight $w_t^{ewp} = 1/N$ in each of the N risky assets. This strategy does not involve any optimization or estimation and completely ignores the data.

Sample-based mean-variance portfolio

In the mean-variance ('MV') model of Markowitz (1952), the investor optimizes the tradeoff between the mean and variance of portfolio returns. To implement this model, we follow the classic 'plug-in' approach; that is, we solve the problem in Equation (6.29) with the mean and covariance matrix of asset returns replaced by their sample counterparts $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$, respectively. Note that this portfolio strategy completely ignores the possibility of estimation error. According to Jagannathan and Ma (2003), we use non-negative constraint on MV portfolio to enhance the performance.

Sample-based minimum-variance portfolio

In the minimum-variance ('MIN') model,we choose the portfolio of risky assets that minimizes the variance of returns; i.e.

$$\min_{w_t^{min}} (w_t^{min})^T \boldsymbol{\Sigma}_t w_t^{min}, \tag{6.30}$$

such that $\mathbf{1}_N^T w_t^{min} = 1$. This method uses only the estimate of the covariance matrix of asset returns (the sample covariance matrix) and completely ignores the estimates of the expected returns. According to Jagannathan and Ma (2003), we use non-negative constraint on MIN portfolio to enhance the performance.

Sample-based Optimal Constrained portfolio

We follow Kirby and Ostdiek (2012) for the sample-based optimal constrained ('OC') portfolio. Under the standard approach to conditional mean-variance optimization, the investor's objective in period t is to choose the $N \times 1$ vector of risky assets weights w_{pt}

that maximizes the quadratic objective function

$$Q(w_{pt}) = w_{pt}^T \tilde{\mu}_t - \frac{\gamma}{2} w_{pt}^T \tilde{\Sigma}_t w_{pt},$$

where $\tilde{\mu}_t = E_t(r_{t+1})$ is the conditional mean vector of the excess risky-asset returns, $\tilde{\Sigma}_t$ is the conditional variance-covariance matrix of the excess risky-asset returns, and γ denotes the investor's coefficient of relative risk aversion. The weight in the risk-free asset is determined implicitly by $1 - w_{pt}^T \mathbf{1}_N$. This problem has a straightforward and well-known solution: $w_{pt} = \tilde{\Sigma}_t^{-1} \tilde{\mu}_t / \gamma$. The solution implies that, in general, the investor divides his wealth between the risk-free asset and a tangency portfolio (TP) of risky assets with weights $w_t^{TP} = \tilde{\Sigma}_t^{-1} \tilde{\mu}_t / \mathbf{1}_N \tilde{\Sigma}_t^{-1} \tilde{\mu}_t$. Because there is a 1 to 1 correspondence between γ and $\tilde{\mu}_{pt} = w_{pt}^T \tilde{\mu}_t$ for each t, we can express the vector of optimal weights as

$$w_{pt} = \tilde{\mu}_{pt} \left(\frac{\tilde{\Sigma}_t^{-1} \tilde{\mu}_t}{\tilde{\mu}_t^T \tilde{\Sigma}_t^{-1} \tilde{\mu}_t} \right). \tag{6.31}$$

If the objective is to study MV portfolios that exclude the risk-free asset, an alternative to considering the TP is to solve the investor's portfolio problem subject to the constraint $w_{pt}^T \mathbf{1}_N = 1$. The 1st-order condition for this problem is

$$\tilde{\mu}_t + \delta_t \mathbf{1}_N - \gamma \tilde{\Sigma}_t w_{pt} = 0,$$

where δ_t is the Lagrange multiplier associated with the constraint. Hence the optimal vector of constrained portfolio weights is

$$w_{pt} = \frac{1}{\gamma} \tilde{\Sigma}_t^{-1} \tilde{\mu}_t + \frac{\delta_t}{\gamma} \tilde{\Sigma}_t^{-1} \mathbf{1}_N. \tag{6.32}$$

Solving for δ_t and substitute the resulting expression, Kirby and Ostdiek (2012) obtained

$$w_t^{oc} = \left(\frac{\tilde{\mu}_{pt} - \tilde{\mu}_t^{mv}}{\tilde{\mu}_t^{TP} - \tilde{\mu}_t^{mv}} \right) \left(\frac{\tilde{\Sigma}_t^{-1} \tilde{\mu}_t}{\mathbf{1}_N^T \tilde{\Sigma}_t^{-1} \tilde{\mu}_t} \right) + \left(1 - \frac{\tilde{\mu}_{pt} - \tilde{\mu}_t^{mv}}{\tilde{\mu}_t^{TP} - \tilde{\mu}_t^{mv}} \right) \left(\frac{\tilde{\Sigma}_t^{-1} \mathbf{1}_N}{\mathbf{1}_N^T \tilde{\Sigma}_t^{-1} \mathbf{1}_N} \right). \tag{6.33}$$

According to Jagannathan and Ma (2003), we use non-negative constraint on OC portfolio to enhance the performance.

Sample-based Volatility timing portfolio

According to Fleming et al. (2001,2003) and Kirby and Ostdiek (2012), the volatility timing ('VT') strategy represents how faster the investor will review their portfolios and rebalance it when volatility changes. It has the weights of the form

$$\hat{w}_{it}^{vt} = \frac{(1/\hat{\sigma}_{it}^2)^\eta}{\sum_{i=1}^{N}(1/\hat{\sigma}_{it}^2)^\eta}, \tag{6.34}$$

for $i = 1, 2, \cdots, N$, where $\eta \geq 0$ and $\hat{\sigma}_{it}$ is the estimation of excess volatility at time t for asset i. The tuning parameter η is a measure of timing aggressiveness. i.e. it determines how aggressively we adjust the portfolio weights in response to volatility changes. When $\eta = 0$, the VT strategy reduces to the 1/N strategy. When η goes to ∞, the weight of the asset with less variance will tend to 1. In the numerical result, we let $\eta = 1$ for simplicity.

Sample-based Reward-to-Risk timing portfolio

According to Kirby and Ostdiek (2012), the reward-to-risk timing ('RRT') strategy has the weights of the form

$$\hat{w}_{it}^{rrt} = \frac{(\hat{\mu_{it}^+}/\hat{\sigma}_{it}^2)^\eta}{\sum_{i=1}^{N}(\hat{\mu_{it}^+}/\hat{\sigma}_{it}^2)^\eta}, \tag{6.35}$$

for $i = 1, 2, \cdots, N$, where $\mu_{it}^+ = \max(\hat{\mu}_{it}, 0)$, $\hat{\mu}_{it}$ is the estimation of mean and $\hat{\sigma}_{it}$ is the estimation of volatility at time t for asset i, and $\eta \geq 0$. In the numerical result, we let $\eta = 1$ for simplicity.

6.3.2 Performance Measures

Using the "rolling-sample" described in the last subsection, each strategy produces a time series of monthly out-of-sample returns. To compare and evaluate the relative performance of the portfolio strategies, we consider the following three statistics:

1. **Sharpe ratio:**

   The out-of-sample Sharpe ratio of strategy k, defined as the sample mean of out-of-sample excess returns (over the risk-free asset), $\hat{\mu}_k$, divided by their sample standard deviation, $\hat{\sigma}_k$:

   $$\hat{SR}_k = \frac{\hat{\mu}_k}{\hat{\sigma}_k}. \tag{6.36}$$

   Note that $\hat{\mu}_k$ and $\hat{\sigma}_k$ are estimated from the time series of monthly out-of-sample returns corresponding to the respective portfolio strategy. The more effective the portfolio strategy, the larger the Sharpe ratio.

2. **Certainty-equivalent (CEQ) Return:**

   The CEQ of a portfolio strategy is defined as the risk-free rate that an investor is willing to accept rather than adopting a particular risky portfolio strategy. For a given risk aversion parameter γ, which is 3 in the numerical results, CEQ is calculated as

   $$\hat{CEQ}_k = \hat{\mu}_k - \frac{\gamma/2}{\hat{\sigma}_k^2}. \tag{6.37}$$

   Therefore the higher the CEQ, the more effective the portfolio strategy.

3. **Turnover**

   To get a sense of the amount of trading required to implement each portfolio strategy, we compute the portfolio turnover, defined as the average sum of the absolute value

of the trades across all the available assets:

$$Turnover = \frac{1}{T-M}\sum_{t=1}^{T-M}\sum_{j=1}^{N}(|\hat{w}_{k,j,t+1} - \hat{w}_{k,j,t^+}|), \tag{6.38}$$

where $\hat{w}_k, j, t$ is the portfolio weight in asset j at time t under strategy k, $\hat{w}_{k,j,t^+}$ is the portfolio weight before rebalancing at $t+1$, and $\hat{w}_{k,j,t+1}$ is the desired portfolio weight at time $t+1$ after rebalancing.

In addition to producing the above three performance measures, we are also interested in testing the statistical significance of the two strategies by computing the p-value of the difference. Using the approach suggested by DeMiguel et. al. (2009) and Jobson and Korkie (1981) (after making the correction pointed out in Memmel (2003)), the p-value of the difference between the Sharpe ratio of each strategy and that of the $1/N$ benchmark is computed using as follows. Given two portfolios i and N, with $\hat{\mu}_i$,$\hat{\mu}_N$,$\hat{\sigma}_i$, $\hat{\sigma}_N$,$\hat{\sigma}_{i,N}$ as their estimated means, variances, and covariances over a sample of size $T-M$, the test of the hypothesis $H_0 : \hat{\mu}_i/\hat{\sigma}_i - \hat{\mu}_N/\hat{\sigma}_N = 0$ is obtained via the test statistic $\hat{z}_{JK}$, which is asymptotically distributed as a standard normal:

$$\hat{Z}_{JK} = \frac{\hat{\sigma}_N\hat{\mu}_i - \hat{\sigma}_i\hat{\mu}_N}{\sqrt{\hat{\theta}}},$$

where $\hat{\theta} = \frac{1}{T-M}(2\hat{\sigma}_i^2\hat{\sigma}_N^2 - 2\hat{\sigma}_i\hat{\sigma}_N\hat{\sigma}_{i,N} + \frac{1}{2}\hat{\mu}_i^2\hat{\sigma}_N^2 + \frac{1}{2}\hat{\mu}_N^2\hat{\sigma}_i^2 - \frac{\hat{\mu}_i\hat{\mu}_N}{\hat{\sigma}_i\hat{\sigma}_N}\hat{\sigma}_{i,N}^2)$.

The above statistical test can similarly be carried for the CEQ returns. The p-value of the difference between the CEQ of each strategy and that of the $1/N$ benchmark is computed using the method shown in Greene (2002) and DeMiguel et al(2009). Specifically, if we denote ν as the vector of moments $\nu = (\mu_i, \mu_n, \sigma_i^2, \sigma_n^2)$, $\hat{\nu}$ as its empirical counterpart obtained from a sample of size $T-M$, and $f(\nu) = (\mu_i - \frac{\gamma}{2}\sigma_i^2) - (\mu_n - \frac{\gamma}{2}\sigma_n^2)$ as the difference in the certainty equivalent of two strategies i and n, the asymptotic distribution of $f(\nu)$ is

$\sqrt{T}(f(\hat{\nu} - f(\nu)))$ has normal distribution with mean 0 and variance $\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\nu}}^T \boldsymbol{\Theta} \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\nu}}$, where

$$\boldsymbol{\Theta} = \begin{pmatrix} \sigma_i^2 & \sigma_{i,n} & 0 & 0 \\ \sigma_{i,n} & \sigma_n^2 & 0 & 0 \\ 0 & 0 & 2\sigma_i^4 & 2\sigma_{i,n}^2 \\ 0 & 0 & 2\sigma_{i,n}^2 & 2\sigma_n^4 \end{pmatrix}.$$

Other than reporting the raw turnover for each strategy, we report an economic measure of this by reporting how transactions costs generated by this turnover affect the returns from a particular strategy; i.e. by denoting $W_{k,t}$ as the wealth of strategy k at time t, and define

$$W_{k,t+1} = W_{k,t}(1 + R_{k,p})(1 - \psi_t), \tag{6.39}$$

with the return net of transactions costs given by $\frac{W_{k,t+1}}{W_{k,t}} - 1$. Considering the case with a proportional transaction cost (PTC) and fixed transaction cost (FTC), then

$$\psi_t = PTC \times \text{Turnover} + FTC \times (\text{Number of Trading Stocks}).$$

For the research on optimal portfolio selection with transaction costs, see Lobo et al (2006) and Boyle and Lin (1997).

6.3.3 Empirical Results

The empirical results are collected in this subsection. The first set of comparison merely compares the efficiency of both $1/N$ and the EP portfolio by assuming $1/N$ as the target portfolio. This is depicted in Figure 6.1 which plots the accumulation values over time from both the EP portfolio (red curve) and the $1/N$ strategy (black curve). These results are based on the Industry 10 data (second dataset in Table 6.2). It is reassuring that the accumulation values from the EP portfolio are always greater than the corresponding $1/N$

strategy, thus indicating the superiority of the EP portfolio.

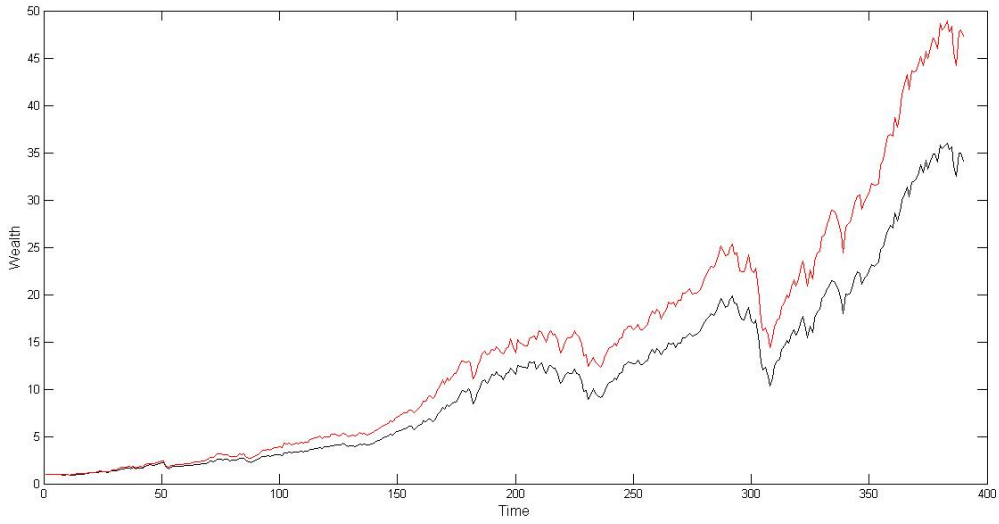

Figure 6.1: Accumulation value under both EPD portfolio and $1/N$ strategy, using Industry 10 dataset

We now provide more comprehensive comparison by considering an exhaustive list of strategies as depicted in Table 6.1. Reference that corresponds to each strategy is provided. For the EP-based strategy, we consider $\beta^* = 0.95$. In the Appendix C, we have included the results for $\beta^* = 0.90$, for robustness check. It is still an open problem determining the optimal number of trading stocks or optimal β. However, from the investor's perspective, we would prefer $\beta < 1$, i.e. the investment is less volatile, but the number of trading stocks is not reduced too much as diversification is still valuable, see Samuelson (1967), and suboptimal, see Ibragimov et al. (2011).

Following George and Hwang (2004) and Jegadeesh and Titman (1993), the holding period for momentum stocks is 6 months. However, our portfolio rebalances monthly. For all sophisticated strategies, we compare their performance based on all stocks in the market and top 70% momentum stocks.

Table 6.1: List of sophisticated strategies considered

#	Description	Abbreviation	Reference
1	Naive Strategy	EWP	DeMiguel, et al. (2009)
2	Mean Variance Portfolio with normalized non-negative weights	MV	Markowitz (1952)
3	Minimum Variance Portfolio with normalized non-negative weights	MIN	Markowitz (1952)
4	Optimal Constrained Portfolio with normalized non-negative weights	OC	Kirby and Ostdiek (2012)
5	Volatility Timing Portfolio	VT	Fleming et al. (2001,2003)
6	Reward-to-Risk Timing Portfolio	RRT	Kirby and Ostdiek (2012)
7	EWP on top 70% momentum stocks	EWP2	
8	MV on top 70% momentum stocks	MV2	
9	MIN on top 70% momentum stocks	MIN2	
10	OC on top 70% momentum stocks	OC2	
11	VT on top 70% momentum stocks	VT2	
12	RRT on top 70% momentum stocks	RRT2	
13	effective portfolio on EWP	EWP + EP	
14	effective portfolio on MV	MV + EP	
15	effective portfolio on MIN	MIN + EP	
16	effective portfolio on OC	OC + EP	
17	effective portfolio on VT	VT + EP	
18	effective portfolio on RRT	RRT + EP	

Note that we also denote "nn" as non-negative constraints, i.e. no-short selling constraints, and "nr" as normalized weights based on non-negative constraints, i.e. no risk-free assets.

Table 6.2: List of datasets considered (source: Ken French)

#	Dataset and source	N	Time period	Abbreviation
1	25 Portfolios Formed on Size and Book-to-Market	25	07/1963-12/2015	BM-25
2	100 Portfolios Formed on Size and Book-to-Market	100	07/1963-12/2015	BM-100
3	25 Portfolios Formed on Size and Operating Profitability	25	07/1963-12/2015	OP-25
4	100 Portfolios Formed on Size and Operating Profitability	100	07/1963-12/2015	OP-100
5	25 Portfolios Formed on Size and Investment	25	07/1963-12/2015	INV-25
6	100 Portfolios Formed on Size and Investment	100	07/1963-12/2015	INV-100
7	10 Industry Portfolios	10	07/1963-12/2015	IND-10
8	49 Industry Portfolios	49	07/1963-12/2015	IND-49

Remark 6.3.1. *Note: the data starts from 07/1963, we use the starting date of 07/1983 for the robust check of 120 and 240 in sample monthly data points, respectively.*

Table 6.3: Sharpe ratios for empirical data (Zero Transaction Cost)

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.1464	0.1475	0.1392	0.1456	0.1480	0.1517	0.1639	0.1413
EWP2	0.1566	0.1538	0.1532	0.1607	0.1637	0.1626	0.1778	0.1673
	(0.1201)	(0.2897)	(0.0086)	(0.0010)	(0.0054)	(0.0187)	(0.2385)	(0.0028)
EWP + EP	0.1912	0.2005	0.1794	0.1735	0.1835	0.2037	0.1870	0.1643
	(0.0000)	(0.0000)	(0.0001)	(0.0231)	(0.0037)	(0.0001)	(0.0554)	(0.1034)
EWP + EP nn	0.1934	0.1980	0.1789	0.1777	0.1814	0.1944	0.1868	0.1714
	(0.0000)	(0.0000)	(0.0001)	(0.0127)	(0.0061)	(0.0012)	(0.0574)	(0.0359)
EWP + EP nr	0.1914	0.1943	0.1825	0.1748	0.1861	0.1994	0.1829	0.1637
	(0.0000)	(0.0001)	(0.0000)	(0.0190)	(0.0017)	(0.0002)	(0.1245)	(0.1261)
MV	0.1773	0.1646	0.1527	0.1511	0.1430	0.1443	0.1407	0.1298
MV2	0.1855	0.1650	0.1623	0.1580	0.1500	0.1425	0.1466	0.1311
	(0.2413)	(0.9613)	(0.0402)	(0.2647)	(0.2847)	(0.7172)	(0.5665)	(0.9202)
MV + EP	0.1902	0.1769	0.1609	0.1643	0.1548	0.1801	0.1516	0.1122
	(0.2019)	(0.3489)	(0.3030)	(0.3113)	(0.2061)	(0.0065)	(0.2347)	(0.1603)
MV + EP nn	0.1749	0.1590	0.1588	0.1464	0.1585	0.1774	0.1511	0.1438
	(0.8655)	(0.7758)	(0.6105)	(0.7901)	(0.2936)	(0.0225)	(0.2954)	(0.4864)
MV + EP nr	0.1772	0.1785	0.1701	0.1707	0.1565	0.1792	0.1550	0.1200
	(0.9888)	(0.2671)	(0.0297)	(0.1166)	(0.1968)	(0.0077)	(0.1450)	(0.4346)
MIN	0.1644	0.1671	0.1630	0.1644	0.1848	0.1823	0.1918	0.1676
MIN2	0.1791	0.1801	0.1676	0.1788	0.1867	0.1814	0.1895	0.1845
	(0.0989)	(0.0783)	(0.6997)	(0.1271)	(0.8905)	(0.8775)	(0.8881)	(0.2813)
MIN + EP	0.1642	0.1866	0.1480	0.1676	0.1862	0.2056	0.1920	0.1897
	(0.9962)	(0.1019)	(0.1899)	(0.7856)	(0.8973)	(0.0576)	(0.9897)	(0.1680)
MIN + EP nn	0.1823	0.1872	0.1680	0.1687	0.1956	0.2058	0.1899	0.1695
	(0.3345)	(0.0922)	(0.8227)	(0.7088)	(0.6602)	(0.0558)	(0.9178)	(0.9473)
MIN + EP nr	0.1686	0.1842	0.1643	0.1706	0.1869	0.2053	0.1850	0.1782
	(0.7314)	(0.1491)	(0.9332)	(0.5879)	(0.8874)	(0.0623)	(0.7037)	(0.5585)
OC	0.1787	0.1609	0.1528	0.1378	0.1411	0.1420	0.1384	0.1251
OC2	0.1853	0.1626	0.1633	0.1497	0.1536	0.1357	0.1467	0.1344
	(0.3269)	(0.8392)	(0.0250)	(0.1051)	(0.0441)	(0.4672)	(0.4858)	(0.4943)
OC + EP	0.1888	0.1668	0.1591	0.1480	0.1531	0.1759	0.1484	0.1148
	(0.2713)	(0.6414)	(0.4090)	(0.4368)	(0.1999)	(0.0137)	(0.2623)	(0.3903)
OC + EP nn	0.1796	0.1606	0.1573	0.1416	0.1599	0.1770	0.1501	0.1424
	(0.9358)	(0.9899)	(0.7007)	(0.8488)	(0.1887)	(0.0854)	(0.2239)	(0.3688)
OC + EP nr	0.1760	0.1697	0.1681	0.1540	0.1577	0.1730	0.1529	0.1164
	(0.7748)	(0.4556)	(0.0441)	(0.2086)	(0.1027)	(0.0290)	(0.1203)	(0.4780)
VT	0.1601	0.1606	0.1503	0.1564	0.1621	0.1651	0.1815	0.1567
VT2	0.1667	0.1640	0.1617	0.1679	0.1717	0.1710	0.1923	0.1785
	(0.2296)	(0.4841)	(0.0182)	(0.0028)	(0.0638)	(0.1249)	(0.3394)	(0.0039)
VT + EP	0.1843	0.2052	0.1831	0.1670	0.1877	0.2039	0.2141	0.1773
	(0.0188)	(0.0002)	(0.0003)	(0.3865)	(0.0207)	(0.0017)	(0.0087)	(0.1564)
VT + EP nn	0.1874	0.2027	0.1833	0.1695	0.1877	0.1992	0.2141	0.1797
	(0.0102)	(0.0004)	(0.0003)	(0.2901)	(0.0211)	(0.0060)	(0.0087)	(0.1154)
VT + EP nr	0.1867	0.2004	0.1855	0.1675	0.1897	0.2020	0.2113	0.1758
	(0.0117)	(0.0008)	(0.0001)	(0.3653)	(0.0121)	(0.0028)	(0.0176)	(0.1933)
RRT	0.1625	0.1619	0.1527	0.1568	0.1624	0.1658	0.1795	0.1551
RRT2	0.1675	0.1649	0.1611	0.1658	0.1720	0.1706	0.1857	0.1724
	(0.2600)	(0.4527)	(0.0192)	(0.0022)	(0.0293)	(0.1310)	(0.6118)	(0.0424)
RRT + EP	0.1868	0.2076	0.1763	0.1759	0.1837	0.2015	0.2062	0.1713
	(0.0106)	(0.0001)	(0.0130)	(0.1058)	(0.0480)	(0.0036)	(0.0330)	(0.2522)
RRT + EP nn	0.1904	0.2050	0.1753	0.1800	0.1795	0.2002	0.2067	0.1749
	(0.0047)	(0.0002)	(0.0194)	(0.0649)	(0.1051)	(0.0053)	(0.0305)	(0.1623)
RRT + EP nr	0.1871	0.2031	0.1777	0.1773	0.1837	0.2003	0.2045	0.1732
	(0.0101)	(0.0004)	(0.0085)	(0.0838)	(0.0436)	(0.0047)	(0.0541)	(0.2171)

Based on the Table 6.3, we make the following remarks:

- The sophisticated strategies under effective portfolio have been highlighted as they are constructed according to the algorithm 6.2.1.
- The effective portfolios consistently outperform $1/N$ strategy under zero transaction cost.
- The effective portfolios with non-negative constraints and normalized weights do not affect the performance too much under zero transaction cost.
- In most cases, the sophisticated strategies based on top 70 % momentum stocks could outperform themselves based on all stocks in the market. This verifies that momentum effects happen everywhere. It is important to incorporate economic factors in terms of investment. See George and Hwang (2004) and Asness et al. (2013).
- Our methods could outperform any sophisticated portfolio based on all stocks and most of sophisticated portfolios based on top 70 % momentum stocks under zero transaction cost. This means if the transaction cost is zero, reducing number of trading stocks is very important.
- In terms of p-value, the sophisticated strategies constructed under effective portfolio could enhance the performance over $1/N$.

Table 6.4: Sharpe ratios for empirical data (50 bps Transaction Cost)

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.1443	0.1450	0.1373	0.1431	0.1461	0.1492	0.1607	0.1376
EWP2	0.1485	0.1446	0.1457	0.1515	0.1560	0.1534	0.1665	0.1567
	(0.5268)	(0.9516)	(0.1196)	(0.0747)	(0.0837)	(0.3750)	(0.6221)	(0.0284)
EWP + EP	0.1696	0.1764	0.1589	0.1497	0.1651	0.1823	0.1694	0.1432
	(0.0190)	(0.0067)	(0.0351)	(0.6028)	(0.1298)	(0.0142)	(0.4683)	(0.6862)
EWP + EP nn	0.1733	0.1764	0.1592	0.1556	0.1635	0.1751	0.1694	0.1521
	(0.0089)	(0.0077)	(0.0369)	(0.3387)	(0.1627)	(0.0515)	(0.4707)	(0.3121)
EWP + EP nr	0.1714	0.1732	0.1626	0.1527	0.1683	0.1799	0.1656	0.1445
	(0.0132)	(0.0166)	(0.0145)	(0.4511)	(0.0739)	(0.0190)	(0.6899)	(0.6353)
MV	0.1551	0.1456	0.1313	0.1334	0.1197	0.1235	0.1212	0.1094
MV2	0.1590	0.1374	0.1381	0.1297	0.1208	0.1133	0.1232	0.1054
	(0.5658)	(0.2454)	(0.1403)	(0.5602)	(0.8540)	(0.0418)	(0.8508)	(0.7366)
MV + EP	0.1561	0.1271	0.1309	0.1247	0.1173	0.1435	0.1262	0.0668
	(0.9231)	(0.1788)	(0.9711)	(0.5184)	(0.8226)	(0.1340)	(0.5821)	(0.0009)
MV + EP nn	0.1483	0.1316	0.1347	0.1191	0.1296	0.1479	0.1282	0.1138
	(0.6302)	(0.4671)	(0.7658)	(0.4099)	(0.4939)	(0.0948)	(0.4824)	(0.8287)
MV + EP nr	0.1506	0.1489	0.1461	0.1427	0.1279	0.1493	0.1320	0.0931
	(0.6647)	(0.7874)	(0.0585)	(0.4614)	(0.4326)	(0.0504)	(0.2722)	(0.1960)
MIN	0.1554	0.1590	0.1554	0.1551	0.1775	0.1728	0.1846	0.1572
MIN2	0.1580	0.1576	0.1461	0.1543	0.1646	0.1582	0.1692	0.1607
	(0.7798)	(0.8455)	(0.4121)	(0.9299)	(0.2950)	(0.0198)	(0.3884)	(0.8344)
MIN + EP	0.1367	0.1615	0.1148	0.1403	0.1511	0.1824	0.1685	0.1504
	(0.1250)	(0.8449)	(0.0011)	(0.2070)	(0.0472)	(0.4406)	(0.3361)	(0.6688)
MIN + EP nn	0.1625	0.1622	0.1458	0.1421	0.1751	0.1826	0.1701	0.1439
	(0.7108)	(0.7973)	(0.6055)	(0.2640)	(0.8949)	(0.4293)	(0.4286)	(0.5302)
MIN + EP nr	0.1506	0.1593	0.1437	0.1437	0.1676	0.1821	0.1647	0.1523
	(0.6860)	(0.9867)	(0.3451)	(0.3229)	(0.4741)	(0.4587)	(0.2800)	(0.7701)
OC	0.1572	0.1316	0.1331	0.1110	0.1151	0.1008	0.1195	0.0970
OC2	0.1591	0.1334	0.1407	0.1194	0.1222	0.0996	0.1248	0.1049
	(0.7747)	(0.8311)	(0.0949)	(0.2739)	(0.2445)	(0.8875)	(0.6599)	(0.5651)
OC + EP	0.1559	0.0974	0.1310	0.0877	0.1097	0.1103	0.1227	0.0626
	(0.8962)	(0.0141)	(0.8064)	(0.0920)	(0.5928)	(0.4983)	(0.7128)	(0.0044)
OC + EP nn	0.1537	0.1286	0.1348	0.1066	0.1267	0.1320	0.1278	0.1067
	(0.7797)	(0.9086)	(0.8698)	(0.8328)	(0.4063)	(0.1261)	(0.3894)	(0.6116)
OC + EP nr	0.1506	0.1322	0.1456	0.1185	0.1250	0.1284	0.1309	0.0832
	(0.4948)	(0.9503)	(0.0942)	(0.5625)	(0.3304)	(0.0554)	(0.2186)	(0.2605)
VT	0.1579	0.1579	0.1483	0.1538	0.1600	0.1625	0.1780	0.1529
VT2	0.1580	0.1544	0.1535	0.1583	0.1631	0.1611	0.1797	0.1670
	(0.9885)	(0.4845)	(0.2899)	(0.2616)	(0.5486)	(0.7379)	(0.8813)	(0.0624)
VT + EP	0.1653	0.1826	0.1646	0.1436	0.1702	0.1845	0.1961	0.1566
	(0.4781)	(0.0333)	(0.0763)	(0.4049)	(0.3669)	(0.0774)	(0.1417)	(0.7906)
VT + EP nn	0.1690	0.1815	0.1650	0.1468	0.1703	0.1805	0.1961	0.1595
	(0.3035)	(0.0431)	(0.0719)	(0.5749)	(0.3625)	(0.1502)	(0.1417)	(0.6472)
VT + EP nr	0.1682	0.1794	0.1670	0.1449	0.1723	0.1832	0.1934	0.1554
	(0.3378)	(0.0657)	(0.0454)	(0.4634)	(0.2722)	(0.0954)	(0.2148)	(0.8549)
RRT	0.1580	0.1574	0.1485	0.1525	0.1581	0.1614	0.1716	0.1467
RRT2	0.1579	0.1543	0.1524	0.1554	0.1627	0.1599	0.1670	0.1570
	(0.9694)	(0.4435)	(0.2908)	(0.3347)	(0.2999)	(0.6385)	(0.7208)	(0.2271)
RRT + EP	0.1670	0.1849	0.1555	0.1528	0.1659	0.1821	0.1859	0.1472
	(0.3585)	(0.0167)	(0.4615)	(0.9763)	(0.4789)	(0.0960)	(0.2454)	(0.9705)
RRT + EP nn	0.1712	0.1839	0.1551	0.1585	0.1622	0.1818	0.1869	0.1525
	(0.1913)	(0.0215)	(0.4977)	(0.6301)	(0.6996)	(0.1013)	(0.2174)	(0.6850)
RRT + EP nr	0.1679	0.1823	0.1572	0.1558	0.1665	0.1818	0.1849	0.1507
	(0.3127)	(0.0298)	(0.3638)	(0.7798)	(0.4340)	(0.0980)	(0.2994)	(0.7847)

Based on the Table 6.4, we make the following remarks:

- we already know from Table 6.3 that the effective portfolios with or without non-negative and normalized constraints do not affect the performance too much. As all sophisticated strategies having no-short selling and no risk-free assets constraints, it is reasonable to compare the sophisticated strategies under effective portfolio and these constraints.

- The effective portfolios consistently outperform $1/N$, OC and RRT strategies under 50 bps proportional transaction cost. However, they may perform as good as constrained MV, constrained MIN and VT strategies, which belong to strategies reducing turnover rates. This demonstrates that when the proportional transaction cost is high, reducing both dimension and turnover rate is very important.

- With proportional transaction cost, the sophisticated strategies based on top 70% momentum stocks could not consistently outperform themselves based on all stocks in the market.

- In terms of p-value, the sophisticated strategies constructed under effective portfolio could enhance the performance in most cases. However, the advantage becomes less comparing to Table 6.3. This means that the effective portfolio could not reduce the turnover rate.

Table 6.5: Sharpe ratios for empirical data (50 bps Transaction Cost + 1 bps fixed Transaction Cost on each trading stock)

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.0926	-0.0518	0.0862	-0.0591	0.0947	-0.0545	0.1366	0.0355
EWP2	0.1097	-0.0002	0.1078	0.0028	0.1176	0.0044	0.1489	0.0818
	(0.0095)	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.2963)	(0.0000)
EWP + EP	0.1593	0.1537	0.1494	0.1299	0.1555	0.1610	0.1555	0.1138
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.1153)	(0.0000)
EWP + EP nn	0.1644	0.1573	0.1506	0.1383	0.1546	0.1574	0.1557	0.1263
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.1126)	(0.0000)
EWP + EP nr	0.1613	0.1496	0.1527	0.1315	0.1579	0.1575	0.1520	0.1150
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2102)	(0.0000)
MV	0.1037	-0.0568	0.0799	-0.0691	0.0644	-0.0854	0.1008	0.0268
MV2	0.1214	-0.0074	0.1000	-0.0150	0.0802	-0.0362	0.1084	0.0457
	(0.0100)	(0.0000)	(0.0000)	(0.0000)	(0.0133)	(0.0000)	(0.4624)	(0.1081)
MV + EP	0.1424	0.0846	0.1173	0.0915	0.1029	0.1153	0.1162	0.0302
	(0.0003)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0927)	(0.7941)
MV + EP nn	0.1387	0.1094	0.1251	0.0978	0.1196	0.1267	0.1199	0.0906
	(0.0111)	(0.0000)	(0.0002)	(0.0000)	(0.0002)	(0.0000)	(0.0553)	(0.0011)
MV + EP nr	0.1399	0.1141	0.1352	0.1138	0.1168	0.1232	0.1236	0.0651
	(0.0004)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0199)	(0.0026)
MIN	0.0954	-0.0631	0.0937	-0.0747	0.1140	-0.0674	0.1558	0.0247
MIN2	0.1138	-0.0041	0.1020	-0.0146	0.1188	-0.0150	0.1495	0.0690
	(0.0425)	(0.0000)	(0.4726)	(0.0000)	(0.7054)	(0.0000)	(0.7196)	(0.0047)
MIN + EP	0.1229	0.1470	0.0974	0.1248	0.1328	0.1693	0.1534	0.1052
	(0.0239)	(0.0000)	(0.7669)	(0.0000)	(0.1530)	(0.0000)	(0.8877)	(0.0000)
MIN + EP nn	0.1535	0.1481	0.1363	0.1273	0.1658	0.1697	0.1582	0.1165
	(0.0013)	(0.0000)	(0.0310)	(0.0000)	(0.0240)	(0.0000)	(0.8951)	(0.0000)
MIN + EP nr	0.1416	0.1443	0.1337	0.1279	0.1573	0.1683	0.1533	0.1215
	(0.0002)	(0.0000)	(0.0016)	(0.0000)	(0.0019)	(0.0000)	(0.8852)	(0.0000)
OC	0.1508	0.0772	0.1271	0.0477	0.1087	0.0139	0.1153	0.0785
OC2	0.1536	0.0886	0.1354	0.0731	0.1164	0.0449	0.1212	0.0937
	(0.6871)	(0.1634)	(0.0682)	(0.0024)	(0.2109)	(0.0006)	(0.6203)	(0.2690)
OC + EP	0.1436	0.0406	0.1194	0.0397	0.0961	0.0704	0.1151	0.0322
	(0.4634)	(0.0134)	(0.3282)	(0.5790)	(0.2046)	(0.0001)	(0.9922)	(0.0002)
OC + EP nn	0.1455	0.1040	0.1273	0.0822	0.1179	0.1069	0.1221	0.0881
	(0.6689)	(0.3039)	(0.9720)	(0.0901)	(0.5081)	(0.0000)	(0.4812)	(0.6116)
OC + EP nr	0.1417	0.0896	0.1371	0.0827	0.1151	0.0959	0.1251	0.0610
	(0.3390)	(0.3163)	(0.1802)	(0.0091)	(0.5241)	(0.0000)	(0.2892)	(0.1578)
VT	0.1039	-0.0476	0.0951	-0.0564	0.1056	-0.0522	0.1519	0.0436
VT2	0.1179	0.0046	0.1143	0.0051	0.1231	0.0051	0.1611	0.0876
	(0.0106)	(0.0000)	(0.0001)	(0.0000)	(0.0008)	(0.0000)	(0.4164)	(0.0000)
VT + EP	0.1563	0.1627	0.1557	0.1264	0.1617	0.1664	0.1831	0.1310
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0121)	(0.0000)
VT + EP nn	0.1606	0.1635	0.1564	0.1309	0.1619	0.1635	0.1831	0.1350
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0121)	(0.0000)
VT + EP nr	0.1591	0.1584	0.1576	0.1264	0.1630	0.1630	0.1806	0.1287
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0217)	(0.0000)
RRT	0.1078	-0.0346	0.0992	-0.0421	0.1070	-0.0394	0.1480	0.0555
RRT2	0.1198	0.0109	0.1152	0.0103	0.1245	0.0110	0.1497	0.0875
	(0.0092)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.8848)	(0.0002)
RRT + EP	0.1581	0.1654	0.1469	0.1342	0.1571	0.1639	0.1738	0.1204
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0378)	(0.0000)
RRT + EP nn	0.1631	0.1668	0.1469	0.1420	0.1538	0.1654	0.1752	0.1282
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0290)	(0.0000)
RRT + EP nr	0.1591	0.1624	0.1483	0.1364	0.1573	0.1624	0.1733	0.1239
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0496)	(0.0000)

Based on the Table 6.5, we make the following remarks:

- Dedicated to the contribution of dimension reduction, the effective portfolios has significant impact on the fixed transaction cost.
- With fixed transaction cost, the "1/N" or "EWP" strategy performs the worst among all sophisticated strategies. This is dedicated to its low return and high transaction cost in this case.
- Due to the dimension reduction, the sophisticated strategies based on top 70% momentum stocks could consistently outperform themselves based on all stocks in the market with fixed transaction cost.
- Our methods could still consistently outperform all sophisticated portfolios based on either all stocks or top 70% momentum stocks under fixed transaction cost. This means if there is certain fixed transaction cost, reducing number of trading stocks is given priority.
- In terms of p-value, the sophisticated strategies constructed under effective portfolio perform could enhance the performance over $1/N$ in all cases.

Table 6.6: CEQ (Risk-aversion=3) for empirical data (Zero Transaction Cost)

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.0036	0.0036	0.0032	0.0035	0.0037	0.0038	0.0042	0.0033
EWP2	0.0041	0.0039	0.0039	0.0043	0.0044	0.0044	0.0048	0.0045
	(0.1090)	(0.2854)	(0.0028)	(0.0006)	(0.0019)	(0.0087)	(0.2078)	(0.0011)
EWP + EP	0.0058	0.0064	0.0053	0.0050	0.0055	0.0065	0.0051	0.0044
	(0.0000)	(0.0000)	(0.0000)	(0.0019)	(0.0001)	(0.0000)	(0.0748)	(0.0485)
EWP + EP nn	0.0059	0.0063	0.0052	0.0052	0.0054	0.0061	0.0051	0.0048
	(0.0000)	(0.0000)	(0.0000)	(0.0004)	(0.0002)	(0.0000)	(0.0745)	(0.0063)
EWP + EP nr	0.0056	0.0057	0.0052	0.0048	0.0053	0.0058	0.0050	0.0042
	(0.0004)	(0.0036)	(0.0001)	(0.1923)	(0.0141)	(0.0240)	(0.1241)	(0.3795)
MV	0.0051	0.0044	0.0039	0.0038	0.0034	0.0035	0.0033	0.0025
MV2	0.0055	0.0045	0.0043	0.0041	0.0037	0.0034	0.0036	0.0026
	(0.1045)	(0.6547)	(0.0156)	(0.1122)	(0.1720)	(0.9763)	(0.5196)	(0.9070)
MV + EP	0.0058	0.0051	0.0043	0.0045	0.0039	0.0052	0.0038	0.0015
	(0.0730)	(0.1333)	(0.1549)	(0.1377)	(0.1115)	(0.0007)	(0.3241)	(0.1000)
MV + EP nn	0.0051	0.0037	0.0042	0.0033	0.0042	0.0052	0.0038	0.0029
	(0.3434)	(0.1518)	(0.1235)	(0.3621)	(0.0465)	(0.0007)	(0.2669)	(0.1735)
MV + EP nr	0.0051	0.0050	0.0047	0.0047	0.0040	0.0050	0.0040	0.0021
	(0.9866)	(0.5085)	(0.0276)	(0.3329)	(0.1302)	(0.0372)	(0.1321)	(0.2511)
MIN	0.0043	0.0044	0.0042	0.0043	0.0050	0.0050	0.0049	0.0041
MIN2	0.0049	0.0050	0.0044	0.0049	0.0051	0.0050	0.0050	0.0048
	(0.0426)	(0.0335)	(0.3799)	(0.1189)	(0.6034)	(0.9677)	(0.6577)	(0.0861)
MIN + EP	0.0042	0.0054	0.0035	0.0045	0.0049	0.0061	0.0047	0.0048
	(0.7274)	(0.0185)	(0.0661)	(0.4387)	(0.7512)	(0.0062)	(0.6113)	(0.1511)
MIN + EP nn	0.0052	0.0054	0.0046	0.0046	0.0060	0.0061	0.0049	0.0045
	(0.0560)	(0.0151)	(0.2097)	(0.3414)	(0.0439)	(0.0058)	(0.9152)	(0.1857)
MIN + EP nr	0.0045	0.0051	0.0043	0.0045	0.0052	0.0059	0.0048	0.0047
	(0.4403)	(0.1513)	(0.6458)	(0.7212)	(0.4551)	(0.0774)	(0.8389)	(0.1606)
OC	0.0052	0.0043	0.0039	0.0031	0.0033	0.0034	0.0031	0.0022
OC2	0.0055	0.0044	0.0044	0.0037	0.0039	0.0030	0.0036	0.0027
	(0.1621)	(0.7870)	(0.0090)	(0.0484)	(0.0159)	(0.8910)	(0.3590)	(0.4417)
OC + EP	0.0057	0.0046	0.0042	0.0036	0.0039	0.0050	0.0037	0.0016
	(0.1218)	(0.4948)	(0.2205)	(0.3092)	(0.1199)	(0.0061)	(0.3587)	(0.3043)
OC + EP nn	0.0053	0.0030	0.0041	0.0026	0.0043	0.0052	0.0038	0.0028
	(0.2188)	(0.1426)	(0.1609)	(0.2076)	(0.0232)	(0.0051)	(0.1717)	(0.1285)
OC + EP nr	0.0050	0.0047	0.0046	0.0039	0.0041	0.0047	0.0039	0.0018
	(0.8260)	(0.8360)	(0.0387)	(0.4685)	(0.0694)	(0.0939)	(0.1143)	(0.2880)
VT	0.0042	0.0042	0.0038	0.0040	0.0043	0.0044	0.0048	0.0040
VT2	0.0045	0.0044	0.0043	0.0046	0.0047	0.0047	0.0052	0.0049
	(0.1670)	(0.3935)	(0.0037)	(0.0009)	(0.0174)	(0.0597)	(0.1621)	(0.0010)
VT + EP	0.0054	0.0065	0.0054	0.0046	0.0055	0.0064	0.0060	0.0049
	(0.0027)	(0.0000)	(0.0000)	(0.1105)	(0.0021)	(0.0000)	(0.0037)	(0.0716)
VT + EP nn	0.0055	0.0064	0.0054	0.0047	0.0055	0.0062	0.0060	0.0050
	(0.0007)	(0.0000)	(0.0000)	(0.0559)	(0.0021)	(0.0001)	(0.0037)	(0.0409)
VT + EP nr	0.0053	0.0059	0.0053	0.0045	0.0054	0.0058	0.0060	0.0047
	(0.0280)	(0.0068)	(0.0003)	(0.8565)	(0.0546)	(0.0556)	(0.0033)	(0.4035)
RRT	0.0043	0.0043	0.0039	0.0041	0.0043	0.0045	0.0048	0.0039
RRT2	0.0046	0.0045	0.0043	0.0045	0.0048	0.0047	0.0050	0.0047
	(0.1833)	(0.3135)	(0.0037)	(0.0004)	(0.0067)	(0.0445)	(0.5037)	(0.0290)
RRT + EP	0.0055	0.0066	0.0050	0.0051	0.0054	0.0063	0.0058	0.0047
	(0.0010)	(0.0000)	(0.0014)	(0.0148)	(0.0081)	(0.0001)	(0.0275)	(0.2030)
RRT + EP nn	0.0057	0.0065	0.0050	0.0053	0.0052	0.0063	0.0058	0.0049
	(0.0002)	(0.0000)	(0.0017)	(0.0036)	(0.0183)	(0.0001)	(0.0181)	(0.0784)
RRT + EP nr	0.0054	0.0060	0.0050	0.0049	0.0052	0.0058	0.0058	0.0046
	(0.0200)	(0.0049)	(0.0234)	(0.2875)	(0.1386)	(0.0728)	(0.0274)	(0.5519)

Based on the Table 6.6, we make the following remarks:

- In terms of CEQ, out effective portfolios could consistently outperform all sophisticated strategies under effective portfolios with zero transaction cost. This means that our effective portfolios could reduce both variance and dimensions while maximizing the returns at the same time.
- There are many cases in red colour showing that he sophisticated strategies based on top 70% momentum stocks could not outperform themselves based on all stocks.
- In terms of p-value, the sophisticated strategies constructed under effective portfolio perform could enhance the performance in all cases.

Table 6.7: CEQ (Risk-aversion=3) for empirical data (50 bps Transaction Cost + 1 bps fixed Transaction Cost on each trading stock)

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.0010	-0.0061	0.0006	-0.0064	0.0011	-0.0062	0.0031	-0.0016
EWP2	0.0018	-0.0036	0.0017	-0.0033	0.0022	-0.0033	0.0036	0.0006
	(0.0038)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2600)	(0.0000)
EWP + EP	0.0042	0.0039	0.0037	0.0027	0.0040	0.0043	0.0038	0.0020
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.1362)	(0.0000)
EWP + EP nn	0.0045	0.0041	0.0038	0.0031	0.0040	0.0041	0.0038	0.0026
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.1291)	(0.0000)
EWP + EP nr	0.0042	0.0037	0.0039	0.0029	0.0041	0.0040	0.0037	0.0022
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2037)	(0.0000)
MV	0.0015	-0.0061	0.0003	-0.0069	-0.0002	-0.0075	0.0013	-0.0034
MV2	0.0023	-0.0039	0.0013	-0.0044	0.0005	-0.0053	0.0017	-0.0023
	(0.0019)	(0.0000)	(0.0000)	(0.0000)	(0.0039)	(0.0000)	(0.3993)	(0.0663)
MV + EP	0.0034	0.0005	0.0021	0.0008	0.0016	0.0020	0.0021	-0.0032
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0886)	(0.7731)
MV + EP nn	0.0031	0.0002	0.0024	0.0003	0.0022	0.0025	0.0023	-0.0006
	(0.0009)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0309)	(0.0005)
MV + EP nr	0.0033	0.0021	0.0030	0.0021	0.0022	0.0025	0.0025	-0.0010
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0099)	(0.0005)
MIN	0.0014	-0.0056	0.0013	-0.0059	0.0022	-0.0054	0.0036	-0.0010
MIN2	0.0021	-0.0030	0.0016	-0.0033	0.0024	-0.0033	0.0035	0.0005
	(0.0164)	(0.0000)	(0.3103)	(0.0000)	(0.5215)	(0.0000)	(0.9646)	(0.0012)
MIN + EP	0.0025	0.0036	0.0015	0.0026	0.0029	0.0045	0.0034	0.0018
	(0.0135)	(0.0000)	(0.8181)	(0.0000)	(0.1350)	(0.0000)	(0.5764)	(0.0000)
MIN + EP nn	0.0039	0.0036	0.0031	0.0027	0.0045	0.0045	0.0037	0.0022
	(0.0000)	(0.0000)	(0.0036)	(0.0000)	(0.0008)	(0.0000)	(0.7445)	(0.0000)
MIN + EP nr	0.0033	0.0034	0.0030	0.0027	0.0040	0.0044	0.0036	0.0025
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.7412)	(0.0000)
OC	0.0038	0.0001	0.0026	-0.0015	0.0018	-0.0028	0.0020	-0.0006
OC2	0.0039	0.0007	0.0030	-0.0002	0.0022	-0.0016	0.0022	0.0003
	(0.4967)	(0.1086)	(0.0348)	(0.0005)	(0.1313)	(0.0002)	(0.5129)	(0.2078)
OC + EP	0.0034	-0.0018	0.0022	-0.0020	0.0012	0.0000	0.0020	-0.0033
	(0.5091)	(0.0047)	(0.3513)	(0.5175)	(0.1599)	(0.0000)	(0.8005)	(0.0000)
OC + EP nn	0.0035	-0.0015	0.0024	-0.0015	0.0021	0.0012	0.0023	-0.0008
	(0.6933)	(0.2084)	(0.4448)	(0.1021)	(0.2105)	(0.0000)	(0.4290)	(0.4802)
OC + EP nr	0.0033	0.0009	0.0031	0.0006	0.0021	0.0013	0.0025	-0.0013
	(0.3120)	(0.3470)	(0.1857)	(0.0056)	(0.4938)	(0.0000)	(0.2844)	(0.0999)
VT	0.0016	-0.0055	0.0012	-0.0059	0.0017	-0.0056	0.0036	-0.0009
VT2	0.0022	-0.0031	0.0020	-0.0030	0.0025	-0.0030	0.0040	0.0010
	(0.0031)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.2406)	(0.0000)
VT + EP	0.0040	0.0044	0.0040	0.0026	0.0043	0.0046	0.0048	0.0029
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0050)	(0.0000)
VT + EP nn	0.0043	0.0044	0.0041	0.0028	0.0043	0.0044	0.0048	0.0030
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0050)	(0.0000)
VT + EP nr	0.0041	0.0041	0.0041	0.0027	0.0042	0.0042	0.0048	0.0028
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0048)	(0.0000)
RRT	0.0018	-0.0049	0.0013	-0.0053	0.0017	-0.0050	0.0035	-0.0005
RRT2	0.0023	-0.0028	0.0021	-0.0029	0.0025	-0.0028	0.0036	0.0009
	(0.0020)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.8473)	(0.0000)
RRT + EP	0.0041	0.0045	0.0036	0.0030	0.0041	0.0044	0.0045	0.0024
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0267)	(0.0000)
RRT + EP nn	0.0044	0.0046	0.0036	0.0033	0.0039	0.0045	0.0046	0.0027
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0148)	(0.0000)
RRT + EP nr	0.0041	0.0042	0.0036	0.0031	0.0040	0.0042	0.0045	0.0026
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0231)	(0.0000)

Based on the Table 6.7, we make the following remarks:

- With fixed transaction cost, the sophisticated strategies based on small number of stocks, such as BM-25, OP-25, INV-25, IND-10, outperform those based on large number of stocks, such as BM-100, OP-100, INV-100, IND-49. This is dedicated to the transaction cost penalty on the number of trading stocks.
- With fixed transaction cost, it is obvious that our effective portfolios could outperform any sophisticated strategy in terms of CEQ. This is again dedicated to the contribution of dimension reduction.
- With fixed transaction cost, any sophisticated strategy under effective portfolio could outperform itself. This is because dimension reduction can reduce fixed transaction cost.

Table 6.8: Effective Dimension for empirical data ($\beta = 0.95$)

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP + EP	4.8000	11.1077	4.5538	9.5846	4.6308	10.3692	5.3231	13.1231
MV + EP	6.3846	19.9846	6.3538	15.8308	6.2308	13.0462	4.5231	19.8615
MIN + EP	5.1231	6.0615	6.3077	6.4308	6.4308	5.2923	4.6769	14.5385
OC + EP	6.5846	27.2769	6.8462	23.9385	6.4769	17.7692	4.6154	19.4308
VT + EP	4.0308	9.3231	4.0769	8.0462	3.8308	8.3538	4.7077	10.6923
RRT + EP	4.0308	9.0154	3.9846	8.7385	3.9385	8.3692	4.5077	11.3692

Based on the Table 6.8, we make the following remarks:

- The higher the total dimension is, the more we could reduce. For example, the reduction ration with respect to 1/N strategy for INV-25 is (25-4.6308)/25=81.48% and for INV-100 is (100-10.3692)/100=89.63%
- The largest proportion of stocks the effective portfolio suggests to invest under these datasets is around 50 %. This means only a small number of stocks could dominant the whole market in terms of variance contribution.
- The smallest proportion of stocks the effective portfolio suggests to invest under these datasets is around 5 %. This means our procedure do not suggest to invest a very small number, such as 1 % or 2 %.

6.4 Conclusion

By incorporating the idea of dimension reduction, the effective portfolio strategy achieves better efficiency by improving on any given target portfolio in terms of investing smaller number of assets, positive alpha and beta < 1.

Key Drawbacks of this chapter
1. The performance of effective portfolio may still rely on the in-sample estimation.
2. The effective portfolio performs better if the targeting portfolio is a more diversified portfolio strategy.

Chapter 7

Conclusions & Further Research

7.1 Conclusions

Effective dimension describes the efficiency of the QMC in estimating high-dimensional integrals. In general, it is hard and computationally inefficient to compute the effective dimension for an arbitrary function with large dimensionality by ANOVA decomposition. Our proposed Delta dimension distribution method provides a simple way to compute aspects of the effective dimension for any arbitrary test function. The Delta method provides one method to decompose a function into the summary of an analytical tractable function and a remainder whose variance is close to zero. This also provides a further research opportunity on seeking better QMC-based algorithms for evaluating high-dimensional integral. We also proved that the conditional tailed dimension is a better measurement than the effective dimension.

For past decade of years, researchers about Quasi-Monte Carlo in the area of numerical finance try to seek optimal PGM by considering the functions of interest under both non-differentiable and high-dimensional functions. We proved a multivariate Delta approach to reduce the nominal dimension for a close function. Under our new designed Delta control approach, we could easily found optimal PGM by combining sophisticated methods on the

function covariance matrix. We also give the new definition of effective dimensions under Delta approach, and discussed that our DC methods intend to handle high dimensionality and multiple non-differentiability at the same time. Furthermore, we showed that the existing PGMs could be covered by our DC method if we let $f_i = x_i$, for $i = 1, \cdots, d$. For the numerical results, we could see that our new DC methods are consistently competitive on reducing effective dimension, and especially it could handle multiple non-differentiability and high dimensionality. Furthermore, our severity control could test the importance of various sub-functions, which may involve discontinuous functions. For further work, we could intend to generalize the dimension reduction with more general dynamics of asset prices, consider mixture of different PGMs, and even design the dimension reduction for rare events.

In the area of insurance, we extend the idea of dimension reduction in the portfolio of insurance contracts. In the past few decades significant advances have been made in the pricing and hedging of VAs. These algorithms are typically able to price a single VA policy with high precision but they are not scalable to large portfolio of VAs valuation. This is unfortunate as in practice, the VA providers are concerned with pricing and hedging of hundreds of thousands of VA policies. In this aspect, the VA providers are willing to sacrifice the high precision of a valuation method that applies to a single VA policy, but resort to a compromised algorithm that has the capability of valuing a large portfolio of VAs at an acceptable level of accuracy. This is precisely the motivation for our proposed green mesh method. As demonstrated in the paper, our proposed method relies on the greater uniformity of the (randomized) low discrepancy sequence of QMC. It has several appealing features, including its simplicity, easy to implement and portability. Its efficiency and its real-time application were also highlighted in our numerical examples. While our experiments focused on the example provided in Gan (2013), for further work we could generalize our algorithm so that it can handle nested simulation, such as that considered in Gan and Lin (2015) and Xu, et al. (2016).

In the area of finance, investors believe that a portfolio with market $beta < 1$ and portfolio

$alpha > 0$ is a better investment strategy. Motivated by dimension reduction in quasi-monte carlo simulation in derivative pricing, this thesis shows how to construct better portfolio alpha and beta trade-off via dimension reduction, i.e. by choosing smaller number of stocks. We showed that the effective portfolio corresponds to a factor model where the factors are the selected assets, and only small portion of stocks or factors, denoted as the effective portfolio dimension (EPD), dominate the whole market. The EPD depends on a predetermined beta value, and it is very flexible so that the stocks with higher alphas could be selected. Due to the estimation error in the practical situation, we show that our effective portfolio could be combined with mispricing factors, i.e. momentum and value, to construct a portfolio with $beta < 1$ and portfolio $alpha > 0$ in a robust way. For further work, the pre-screening of stocks could be incorporated by other informations, such as familiarity, see Boyle et al. (2012); option pricing linked stock prices, see Shristoffersen and Pan (2014); clustering, see Harris (1992); risk measures (drawdown, drawup), see Chekhlov et al. (2005); and nonlinear support vector machine (kernel function), see Huerta et al. (2013). The effective portfolio method should be applied because higher returns increase the potential for higher risks.

All in all, controlling the effective dimensionality is related to control the time complexity, risk and reward trade-off, model and parameter uncertainty. This will become the crucial research in the area of quantitative finance and insurance.

7.2 Further Research on QMC

The past three decades has sparked the development of a large body of theory concerning multivariate probability distributions. Most studies about multivariate asset models are based on Brownian motions due to their simple structure. However, since the work of Mandelbrot and Taylor (1967) and Clark (1973), it has been widely recognized the presence of significant skewness and excess kurtosis in empirical asset return distributions; that is, the returns are non-normally distributed. To allow for both kurtosis and skewness for the

multivariate probability distribution of assets returns, multivariate Lévy processes are used as a tractable model for asset returns.

Definition 7.2.1. *(Multivariate Affine Transformation) Let* $X = (X_1, \cdots, X_N)'$ *be an* N*-dimensional random vector distributed with mean vector* $M \in R^N$ *and covariance matrix* $\boldsymbol{\Sigma} \in R^{N \times N}$ *is a square matrix such that* $\boldsymbol{AA'} = \boldsymbol{\Sigma}$*. We defined the multivariate affine transformation of* $\boldsymbol{X}$ *is* $\boldsymbol{AZ}$*, where* $\boldsymbol{Z}$ *is a vector of* N*-dimensional independent and identically distributed (i.i.d) random variables with mean* 0 *and variance* 1*, which is called standard distribution. If* $\boldsymbol{Z}$ *is a vector* N*-dimensional i.i.d standard general hyperbolic (GH) random variables, then* $\boldsymbol{X}$ *is said to have multivariate affine general hyperbolic (MAGH) distribution. If* $\boldsymbol{Z}$ *is a vector* N*-dimensional i.i.d standard jump diffusion (JD) random variables, then* $\boldsymbol{X}$ *is said to have multivariate affine jump diffusion (MAJD) distribution.*

As Schmidt et al. (2006) and Fajardo and Farias (2010) pointed out, this definition is responsible for simplifying the estimation procedure and allows us to model more leptokurtic data. Following the similar algorithm used by Schmidt et al. (2006) and Fajardo and Farias (2009, 2010), we use the following steps to estimate the parameter of MAGH and MAJD distributions.

Step 1: Get $\boldsymbol{Z} = \boldsymbol{B}(\boldsymbol{X} - \boldsymbol{M})$, where $\boldsymbol{B}$ is the inverse square matrix of $\boldsymbol{A}$ such that $\boldsymbol{AA'} = \boldsymbol{\Sigma}$. Then $\boldsymbol{Z}$ has mean $\boldsymbol{0}$ and variance covariance matrix $\boldsymbol{I}$, with parameters $\boldsymbol{\omega}$.

Step 2: Estimate the parameters of univariate standard random variables Z_i by using maximum likelihood estimation, for $i = 1, \cdots, N$.

The procedure leads to a simplification on the parameter estimation and allows us to estimate d one-dimensional standard distributions rather than the simultaneous estimation. Note that $\boldsymbol{A}$ is a key generation of MAGH and MAJD distributions, and hence we refer to $\boldsymbol{A}$ as a generating matrix.

The two additional advantages of MAGH or MAJD distributions are that the joint density and characteristic functions exist in closed-form. Specifically, according to Definitions (7.2.1), the density function of $\boldsymbol{X}$ adhering to an MAGH or MAJD density can be repre-

sented as

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = |\boldsymbol{A}|^{-1} \prod_{j=1}^{N} f_{Z_j}(z_j), \tag{7.1}$$

where $|\boldsymbol{A}|$ is the determinant of $\boldsymbol{A}$, $\boldsymbol{x} = (x_1, \cdots, x_N)'$ and $\boldsymbol{z} = (z_1, \cdots, z_N)' = \boldsymbol{B}(\boldsymbol{x} - \boldsymbol{M})$. In addition, the characteristic function (CF) of $\boldsymbol{X}$ is given by

$$\phi_{\boldsymbol{X}}(\boldsymbol{\omega}) = E(\exp(i\boldsymbol{\omega}'\boldsymbol{X})) = \exp(i\boldsymbol{\omega}'\boldsymbol{X}) \prod_{k=1}^{N} \phi_{Z_k}(\psi_k), \tag{7.2}$$

where $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_N)'$ and $\boldsymbol{\psi}' = (\psi_1, \cdots, \psi_N) = \boldsymbol{\omega}'\boldsymbol{A}$.

Now, we introduce the MAGH and MAJD models for the asset returns. The risky asset returns over a small time interval are defined as follows:

$$R_j(t) = log(S_j(t)) - log(S_j(t-1)), j = 1, \cdots, N, t = 1, \cdots, T, \tag{7.3}$$

where $S_j(t)$ is the j-th asset price at time t. The return on the risk-free asset over the same time interval equals r_f. In this line, we construct the assets returns using the MAGH or MAJD distributions based on the standard GH or JD margins; that is

$$\boldsymbol{R(t)} = \begin{bmatrix} M_1(t) \\ \ddots \\ M_N(t) \end{bmatrix} + \begin{bmatrix} A_{11}(t) & \cdots & A_{1N}(t) \\ \vdots & \ddots & \vdots \\ A_{n1}(t) & \cdots & A_{NN}(t) \end{bmatrix} \begin{bmatrix} Z_1 \\ \ddots \\ Z_N \end{bmatrix} = \boldsymbol{M}(t) + \boldsymbol{A}(t)\boldsymbol{Z}, \tag{7.4}$$

where $M_j(t)$ is mean of the asset return $R_j(t)$; Z_j adheres to a standard GH or JD distribution; and $\boldsymbol{A}(t)$ is a generating matrix such that the covariance matrix $\boldsymbol{\Sigma}(t)$ at time t, i.e. $\boldsymbol{A}(t)\boldsymbol{A}(t)^T = \boldsymbol{\Sigma}(t)$. Note in a single period model, we could simply define $\boldsymbol{R} = \boldsymbol{M} + \boldsymbol{Z}$, where $\boldsymbol{Z}$ is a vector of N-dimensional standard GH or JD random variables.

For Derivative pricing, we have to construct a risk-neutral measure to ensure that there is no arbitrage opportunities in the market described by the model (see Harrison and Kreps,

1979; Harrison and Pliska, 1981, 1983). Gerber and Shiu (1994) first employ the use of the Esscher transform for option valuation in an incomplete market. Employing Esscher transform to multidimensional derivative pricing, Fajardo and Farias (2010) provide the risk-neutral probability density function of MAGH distribution.

The choice of a generating matrix $\boldsymbol{A}$ such that $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{\Sigma}$ plays a indispensable role in pricing derivatives under Black-Scholes model and can be generalized to the MAGH or MAJD models. Different specification of A generates different simulated paths of MAGH or MAJD distributions, which in turn affects pricing efficiency of Quasi Monte Carlo simulation.

7.3 Further Research on Portfolio Selection

An investor is concerned with allocating his or her initial wealth among the above $N+1$ assets. Let $\boldsymbol{w} := (w_1, \cdots, w_N)'$ denote the portfolio strategy for the N risky assets so that $w_i, i = 1, \ldots, N$, captures the portfolio weight that is invested in the i-th risky asset. Then $1 - \boldsymbol{w}'\mathbf{1}$ gives the weight invested in the risk-free asset, where $\mathbf{1} = (1, 1, \ldots, 1)' \in \mathbb{R}^N$. Without loss of generality, we assume that the investor's initial wealth at time 0 is \$1.

Let $W^{\boldsymbol{w}}$ denote the investor's wealth random variable at time 1 corresponding to the portfolio weight $\boldsymbol{w}$. Then we have

$$W^{\boldsymbol{w}} = (1 + r_f) + \boldsymbol{w}'(\boldsymbol{R} - r_f\mathbf{1}). \tag{7.5}$$

Under the utility-based approach to portfolio selection, the optimal portfolio strategy $\boldsymbol{w}$ is chosen as one that maximizes the expected utility of the terminal wealth. Formally, this boils down to solving the following optimization problem:

$$\max_{\boldsymbol{w}} \; E[U(W^{\boldsymbol{w}})], \tag{7.6}$$

where $U(\cdot)$ corresponds to the investor's utility function. In our case, $U(\cdot)$ is selected to

be the exponential utility function

$$U(W) = -e^{-\lambda W}, \quad W \in \mathbb{R}, \tag{7.7}$$

where $\lambda > 0$ quantifies the risk aversion of the investor. Using (7.5) and (7.7), the objective function in (7.6) can be expressed explicitly as

$$E[U(W^{\boldsymbol{w}})] = -e^{-\lambda(1+r_f)} E[\exp(-\lambda \boldsymbol{w}'(\boldsymbol{R} - r_f \mathbf{1}))]. \tag{7.8}$$

Hence the tractability of the expected utility optimization problem (7.6) in turn depends on the complexity of the distributional assumption of the multivariate random vector $\boldsymbol{R}$.

Assuming the assets returns follow the MAGH or MAJD distribution, then $\boldsymbol{R} = \boldsymbol{M} + \boldsymbol{AZ}$ according to (7.2.1). We could solve the optimal portfolio strategy according to(7.6). Then the expectation in (7.8) becomes

$$E[\exp(-\lambda \boldsymbol{w}'(\boldsymbol{R} - r_f \mathbf{1}))] = E[\exp(-\lambda \boldsymbol{w}'(\boldsymbol{M} - r_f \mathbf{1}) - \lambda \boldsymbol{w}' \boldsymbol{AZ})].$$

Therefore, the exponential utility-based portfolio maximization problem (7.6) is equivalent to

$$\max_{\boldsymbol{w}} \left\{ \lambda \boldsymbol{w}'(\boldsymbol{M} - r_f \mathbf{1}) - \ln \phi_{GH}(i\lambda \tilde{\boldsymbol{w}}) \right\}. \tag{7.9}$$

We could see that the value of utility function also depends on the choice of $\boldsymbol{A}$ even incorporating skewness and kurtosis, and hence, the dimension reduction technique is important to determine the optimal matrix $\boldsymbol{A}$.

Bibliography

[1] Acworth, P., Broadie, M. and Glasserman, P. (1998). A comparison of some Monte Carlo and quasi-Monte Carlo techniques for option pricing. In Monte Carlo and Quasi-Monte Carlo Methods 1996 (H. Niederreiter, P. Hellekalek, G. Larcher & P. Zinterhof eds). New York: Springer, 1-18.

[2] Akesson, F. and Lehoczky, J.P. (1998). Discrete eigenfunction expansion of the multi-dimensional Brownian motion and the ornstein-Uhlenbeck process. *Technical Report*, Carnegie-Mellon University.

[3] Aloise, D., Deshpande A., Hansen, P. and Popat, P. (2009). NP-hardness of euclidean sum-of-squares clustering. *Machine Learning*, **75**(2): 245-248.

[4] Asness, C.S., Moskowitz, T.J. and Pedersen L.H. (2013). Value and Momentum Everywhere. *The Journal of Finance*, **LXVIII**(3).

[5] Azimzadeh, P., and Forsyth P.A. (2015). The existence of optimal bang-bang controls for GMxB contracts. *SIAM Journal on Financial Mathematics*, **6**(3): 117-139.

[6] Bacinello, A.R., Millossovich P. and Montealegre, A. (2016). The valuation of GMWB variable annuities under alternative fund distributions and policyholder behaviours. *Scandinavian Actuarial Journal* - Issue 5.

[7] Bauer, D., Kling, A. and Russ, J. (2008). A universal pricing framework for guaranteed minimum benefits in variable annuities. *ASTIN Bulletin*, **38**(2): 621-651.

[8] Bauer, D., Reuss A. and Singer, D. (2012). On the calculation of the solvency capital requirement based on nested simulations. *ASTIN Bulletin*, **42**(2): 453-499.

[9] Bawa, V.S., Brown, S. and Klein, R. (1979). Estimation Risk and Optimal Portfolio Choice. *North Holland*, Amsterdam.

[10] Beker, F., and Espino, E. (2013). Short-Term Momentum and Long-Term Reversal in General Equilibrium. *Working paper*.

[11] Black, F. and Litterman, R. (1990). Asset Allocation: Combining Investors Views with Market Equilibrium. *Fixed Income Research*, Goldman, Sachs & Company, September.

[12] Boyle, P.P. (1977). Options: A Monte Carlo Approach. *J. Financial Economics*, **4**(3): 323-338.

[13] Boyle, P., Broadie, M. and Glasserman, P. (1997). Monte Carlo methods for security pricing. *J. Econom. Dynam. Control*, **21**: 1267-1321.

[14] Boyle, P.P., Garlappi L., Uppal R. and Wang T. (2012). Keynes Meets Markowitz: The Trade-off Between Familiarity and Diversification *Management Science*, **58**(2): 253-272.

[15] Boyle, P.P., Kolkiewicz, A.W. and Tan, K.S. (2001). Valuation of the reset options embedded in some equity-linked insurance. *North American Actuarial Journal*, 5(3),1-18.

[16] Boyle, P.P., Kolkiewicz, A.W. and Tan, K.S. (2003). Pricing American Style Options Using Low Discrepancy Mesh Method. *Quantitative Finance*, 6, 841-860.

[17] Boyle, P., and Lin, X. (1997). Optimal Portfolio Selection with Transaction Costs *North Amer. Act. J.*, 27-39.

[18] Boyle, P.P. and Tian, W. (2008) The Design of Equity-Indexed Annuities. *Insurance: Mathematics and Economics*, **43**,303-315.

[19] Bratley, P. and Fox, B.L. (1988). Algorithm 659 Implementing Sobol's Quasirandom Sequence Generator. *ACM Transactions on Mathematical Software*, **14**(1): 88-100.

[20] Bratley, P., Fox, B.L. and Niederreiter, H. (1992). Implementation and tests of low-discrepancy sequences. *ACM Transactions on Modeling and Computer Simulation*, **2**: 195-213.

[21] Brown, S.J. (1976). Optimal Portfolio Choice under Uncertainty. *Ph.D. Diss.*, University of Chicago.

[22] Carhar, M.M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, **52**(1): 57-82.

[23] Caflisch, R., Morokoff, W. and Owen, A. (1997). Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension. *J. Comp. Finance*, **1**: 27-46.

[24] Chekhlov, A., Uryasev, S. and Zabarankin M. (2005). Drawdown Measure In Portfolio Optimization. *International Journal of Theoretical and Applied Finance*, **8**(1): 13-58.

[25] Christoffersen, P. and Pan, X. (2014). Equity Portfolio Management Using Option Price Information. *Rotman School of Management Working Paper*.

[26] Dai, T.S., Yang S.S. and Liu, L.C. (2015). Pricing guaranteed minimum/lifetime withdrawal benefits with various provisions under investment, interest rate and mortality risks. *Insurance: Mathematics and Economics*, **64**, 364-379.

[27] Dasgupta, S. (2007). The hardness of k-means clustering. Technical Report CS2007-0890. *University of California*, San Diego.

[28] Dybvig, P.H., and Ross, S.A. (1985). Differential Information and Performance Measurement Using a Security Market Line. *The Journal of Finance*, **40**: 383-399.

[29] DeMiguel, V., Garlappi L., and Uppal R. (2009). Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *The Review of Financial Studies*, **22**(5):

i4.

[30] Fajardo, J., Farias, A. (2009). Multivariate affine generalized hyperbolic distributions: an empirical investigation. *International Review of Financial Analysis* 18(4), 174-184.

[31] Fajardo, J., Farias, A. (2010). Derivative pricing using multivariate affine generalized hyperbolic distributions. *Journal of Banking & Finance* 34(7), 1607-1617.

[32] Fama, E.F., and French K.R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33(3).

[33] Fama, E.F., and French K.R. (2015). A five-factor asset pricing model. *Journal of Financial Economics* 116, 1-22.

[34] Fan, K., Shen, Y., Siu, T.K. and Wang, R. (2015). Pricing annuity guarantees under a double regime-switching model. *Insurance: Mathematics and Economics* **52**: 62-78.

[35] Faure, H. (1982). Discrépance de suites associées à un système de numération (en dimensions). *Acta Arith.*, **41**: 337-351.

[36] Fleming J., Kirby C. and Ostdiek B. (2001). The Economic Value of Volatility Timing *The Journal of Finance*, **56**(1): 329-352.

[37] Fleming J., Kirby C. and Ostdiek B. (2003). The Economic Value of Volatility Timing Using Realized Volatility *Journal of Financial Economics*, **67**: 473-509.

[38] Forsyth, P.A. and Vetzal, K.R. (2014). An optimal stochastic control framework for determining the cost of hedging of variable annuities. *Journal of Economic Dynamics and Control*, **44**: 29-53.

[39] Frost, P.A. and Savarino, J. E. (1986). An Empirical Bayes Approach to Efficient Portfolio Selection. *The Journal of Financial and Quantitative Analysis*, **21**(3): 293-305.

[40] Frost, P.A. and Savarino, J. E. (1988). For better performance: Constrain portfolio weights. *J. Portfolio Management*, **15**(1): 29-24.

[41] Gan, G. (2013). Application of data clustering and machine learning in variable annuity valuation. *Insurance: Mathematics and Economics*, **53**(3): 795-801.

[42] Gan, G. and Lin, X.S. (2015). Valuation of large variable annuity portfolios under nested simulation: A functional data approach. *Insurance: Mathematics and Economics*, **62**, 138-150.

[43] Gan, G. and Lin X.S. (2016). Efficient greek calculation of variable annuity portfolios for dynamic hedging: a two-level metamodeling approach. *North American Actuarial Journal*, In Press.

[44] Gan, G. and Valdez, E.A. (2016). Regression Modeling for the Valuation of Large Variable Annuity Portfolios. *Working paper*.

[45] Garlappi, L., Uppal, R. and Wang, T. (2007). Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach. *Review of Financial Studies*, **20**(1): 41-81.

[46] George, T.J. and Hwang, C.Y. (2004). The 52-Week High and Momentum Investing. *The Journal of Finance*, **59**(5): 2145-2176.

[47] Gerber, H. and Shiu, E. (1994). Option Pricing by Esscher Transforms. *Transactions of the Society of Actuaries*, **46**: 99-191.

[48] Glasserman, P. (2004). Monte Carlo methods in financial engineering. Springer-Verlag, New York.

[49] Goldfarb, D. and Iyengar, G. (2003). Robust portfolio selection problems. *Math. Oper. Res.*, **28**: 1-38.

[50] Greene, W.H. (2002). Econometric Analysis. *New York: Prentice-Hall*.

[51] Griebel, M (2006). Sparse grids and related approximation schemes for higher dimensional problems, in: L. Pardo, A. Pinkus, E. Suli, M. Todd (Eds.). *Foundations of Computational Mathematics (FoCM05), Santander, Cambridge University Press*, pp. 106-161.

[52] Griebel, M. and Holtz, M. (2010). Dimension-wise integration of high-dimensional functions with applications to finance. *Journal of Complexity*, **26**: 455-489.

[53] Grinold, R. and Kahn, R. (2000). Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Selecting Superior Returns and Controlling Risk. *2nd ed., McGraw-Hill, New York*.

[54] Halldórsson, B.Y and Tütüncü, R.H. (2000). An Interior-Point Method for a Class of Saddle-Point Problems. *J. Optim. Theory Appl.*, **116**(3): 559-590.

[55] Halton, J.H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, **2**: 84-90.

[56] Hardy, M. (2000). Hedging and reserving for single-premium segregated fund contracts. *North American Actuarial Journal*, **4**(2): 63-74.

[57] Hardy, M. (2003). Investment Guarantees: Modeling and Risk Management for Equity-Linked Life Insurance. *John Wiley & Sons*, Inc., Hoboken, New Jersey.

[58] Harris, L. (1991). Stock price clustering and discreteness. *Review of Financial Studies*, **4**(3): 389-415.

[59] Harrison, J. M. and Kreps, D. M. (1979). Martingales and Arbitrage in Multiperiod Securities Markets. *Journal of Economic Theory*, **20**: 381-408.

[60] Harrison, J. M. and Pliska, S. R. (1981). Martingales and Stochastic Integrals in the Theory of Continuous Trading. *Stochastic Processes and Their Applications*, **11**: 215-280.

[61] Harrison, J.M. and Pliska, S. R. (1983). A Stochastic Calculus Model of Continuous Trading: Complete Markets. *Stochastic Processes and Their Applications*, **15**: 313-316.

[62] Hejazi, S.A. and Jackson, K.R. (2016). A neural network approach to efficient valuation of large portfolios of variable annuities. *Insurance: Mathematics and Economics*, **70**: 169-181.

[63] Hejazi, S.A., Jackson, K.R. and Gan, G. (2015). A Spatial Interpolation Framework for Efficient Valuation of Large Portfolios of Variable Annuities. *Working paper*.

[64] Hong, H.S. and Hickernell, F.J. (2003). Algorithm 823: Implementing Scrambled Digital Sequences. *ACM Transactions on Mathematical Software*, **29**(2): 95-109.

[65] Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, **2**(3): 283-304.

[66] Huang, Y.T. and Kwok, Y.K. (2016). Regression-based Monte Carlo methods for stochastic control models: variable annuities with lifelong guarantees. *Quantitative Finance*, **16**(6): 905-928.

[67] Huang, J.Z., Ng, M.K., Rong, H. and Li, X. (2005). Automated variable weighting in k-means type clustering. *IEEE Trans. Pattern Anal. Mach. Intell*, **27**(5): 657-668.

[68] Huerta, R., Elkan, C. and Corbacho, F. (2013). Nonlinear support vector machines can systematically identify stocks with high and low future returns. *Algorithmic Finance*, **2**(1): 45-58.

[69] Hyndman, C. and Wenger, M.. (2014). Pricing and hedging GMWB riders in a binomial framework. *Working paper*.

[70] Ibragimov, R., Jaffee, D. and Walden, J. (2011). Diversification disasters. *Journal of Financial Economics*, **99**: 333-348.

[71] Ignatieva, K., Song, A. and Ziveyi, J. (2016). Pricing and hedging of guaranteed minimum benefits under regime-switching and stochastic mortality. *Insurance: Mathematics and Economics* 70: 286-300.

[72] Imai, J. and Tan, K. S. (2006). A general dimension reduction technique for derivative pricing. *J. Computational Finance*, **10**(2): 129-155.

[73] Imai, J. and Tan, K. S. (2009). An accelerating quasi-Monte Carlo method for option pricing under the generalized hyperbolic Lévy process. *SIAM Journal on Scientific*

Computing, **31**(3): 2282-2302.

[74] Imai, J. and Tan, K.S. (2014). Pricing Derivative Securities Using Integrated Quasi-Monte Carlo Methods with Dimension Reduction and Discontinuity Realignment. *SIAM Journal on Scientific Computing*, **36**(4): A2101-A2121.

[75] Isaaks, E. and Srivastava, R. (1990). An Introduction to Applied Geostatistics. *Oxford University Press*, Oxford, UK.

[76] Jagannathan, R. and Ma, T. (2003). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *The Journal of Finance*, **58**(4): 1651-1684.

[77] Jegadeesh, N. and Titman, S. (1993). Returns to Bying Winners and Selling Losers: Implications for Stock Market Efficiency. *Technical Report*, 1-17.

[78] Jensen, M.C. (1969). Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios. *Journal of Business*, **42**(2): 167-247.

[79] Jiang, S. and Chang, M. (2010). Variable annuity with guarantees: valuation and simulation. *Journal of Money, Investment and Banking*, **14**: 74-83.

[80] Jobson, J.D. and Korkie, R. (1981). Performance Hypothesis Testing with the Sharpe and Treynor Measures. *The Journal of Finance*, **36**(4): 889-908.

[81] Jorion P. (1986). Bayes-Stein Estimation for Portfolio Analysis. *Journal of Financial and Quantitative Analysis*, **21**(3): 279-292.

[82] Joy, C., Boyle, P. and Tang, K. S. (1996). Quasi-Monte Carlo methods in numerical finance. *Management Science*, **42**(6): 926-938.

[83] Kan, R. and Wang, X. (2016). On the Economic Value of Alphas. *29th Australasian Finance and Banking Conference 2016*.

[84] Kan, R. and Zhou, G. (2007). Optimal Portfolio Choice with Parameter Uncertainty. *Journal of Financial and Quantitative Analysis*, **42**(3): 621-656.

[85] Kirby, C., and Ostdiek B. (2012). Its All in the Timing: Simple Active Portfolio Strategies that Outperform Naive Diversification. *Journal of Financial and Quantitative Analysis*, **47**(2): 437-467.

[86] Klein, R.W. and Bawa, V.S. (1976). The Effect of Estimation Risk on Optimal Portfolio Choice. *Journal of Financial Economics*, **3**: 215-231.

[87] Ledoit O., and Wolf M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection *Journal of Empirical Finance*, **10**: 603-621.

[88] Ledoit O., and Wolf M. (2004). Honey, I Shrunk the Sample Covariance Matrix *Journal of Portfolio Management*, **30**(4): 110-119.

[89] Ledoit O., and Wolf M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices *Annals of Statistics*, **40**(2): 1024-1060.

[90] Lemieux, C. (2009). Monte Carlo and Quasi-Monte Carlo Sampling. *Springer Series in Statistics*, New York.

[91] Leobacher, G. (2012). Fast orthogonal transforms and generation of Brownian paths. *J. Complexity*, **28**: 278-302.

[92] Kritzman, M., Page, S., and Turkington, D. (2010). In Defense of Optimization: The Fallacy of 1/N. *Financial Analysts Journal*, **66**(2).

[93] Li, X., Miffre, J., Brooks, C., and O'Sullivan, N. (2008). Momentum profits and time-varying unsystematic risk. *J. Banking & Finance*, **32**(4): 541-558.

[94] Li, X., Brooks, C., and Miffre, J. (2009). Low-cost momentum strategies. *J. Asset Management*, **9**(6): 366-379.

[95] Lin, X.S., Tan, K.S. and Yang, H. (2009). Pricing annuity guarantees under a regime switching model. *North American Actuarial Journal*, **13**: 316-338.

[96] Lin, J., and Wang, X. (2008). New Brownian bridge construction in quasi-Monte Carlo methods for computational finance. *J. Complexity*, **24**: 109-133.

[97] Lobo, M., Fazel, M. and Boyd, S. (2007). Portfolio Optimization with Linear and Fixed Transaction Costs. *Annuals of Operations Research*, **152**(1): 341-365.

[98] Luo, X., and Shevchenko, P.V. (2015). Valuation of variable annuities with guaranteed minimum withdrawal and death benefits via stochastic control optimization. *Insurance: Mathematics and Economics*, **62**, 5-15.

[99] MacKinlay, A.C. and Pástor, L. (2000). Asset Pricing Models: Implications for Expected Returns and Portfolio Selection. *Reveiw of Financial Studies*, **13**(4): 883-916.

[100] Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance* 7(1), 77-91.

[101] Matousek, J. (1998). On the L2-Discrepancy for Anchored Boxed. *Journal of Complexity*, **14**(4): 527-556.

[102] McKay, M. D., Conover, W. J. and Beckman, R. J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, **21**: 239-245.

[103] McLeish, D.L. (2004). Monte Carlo Simulation and Finance. *John Wiley and Sons Inc.*, New Jerscy.

[104] Memmel, C. (2003). Performance Hypothesis Testing with the Sharpe Ratio. *Finance Letters*, **1**: 21-23.

[105] Michaud, R.O. (1989). Efficient Asset Management. *Harvard Business School Press*, Boston.

[106] Moore, E.H. (1920). On the Reciprocal of the General Algebraic Matrix. *Bulletin ofthe American Mathematical Society*, **26**: 394-395.

[107] Moskowitz, B. and Caflisch, R. E. (1996). Smoothness and Dimension Reduction in Quasi-Monte Carlo Methods. *Journal of Mathematical and Computer Modeling*, **23**(8): 37-54.

[108] Ng, A.C.Y. and Li, J.S.H. (2011). Valuing variable annuity guarantees with the multivariate Esscher transform. *Insurance: Mathematics and Economics*, **49**(3): 393-400.

[109] Niederreiter, H. (1992). Random number generation and quasi-Monte Carlo methods. *Volume 63 of SIAM CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia.

[110] Oehlert, Gary W. (1992). A Note on the Delta Method. *The American Statistician*, **46**(1): 27-29.

[111] Olivares-Nadal, A.V. and DeMiguel, V. (2015). A Robust Perspective on Transaction Costs in Portfolio Optimization. *Working paper*.

[112] Owen, A.B. (1997). Monte Carlo variance of scrambled equidistribution quadrature. *SIAM Journal of Numerical Analysis*, **34**(5): 1884-1910.

[113] Owen, A.B. (2003). The dimension distribution and quadrature test functions. *Statist. Sinica*, **13**: 1-17.

[114] Ökten, G. and Estman, W. (2004). Randomized quasi-Monte Carlo methods in pricing securities. *Journal of Economic Dynamics & Control*, **28**: 2399-2426.

[115] Ökten, G. and Willyard, M. (2010). Parameterization based on randomized quasi-Monte Carlo methods. *Parallel Computing*, **36**: 415-422.

[116] Papageorgiou, A. (2002). The Brownian bridge does not offer a consistent advantage in quasi-Monte Carlo integration. *J. Complexity*, **18**: 171-186.

[117] Pástor, L. (2000). Portfolio Selection and Asset Pricing Models. *The Journal of Finance*, **55**: 179-223.

[118] Pástor, L. and Stambaugh, R.F. (2000). Comparing Asset Pricing Models: An Investment Perspective. *Journal of Financial Economics*, **56**: 335-381.

[119] Paskov, S. H. and Traub, J. F. (1995). Faster evaluation of financial derivatives. *J. Portfolio Management*, **22**(1): 113-120.

[120] Samuelson, P.A. (1967). General proof that diversification pays. *The Journal of Financial and Quantitative Analysis*, **2**: 1-13.

[121] Scheicher, K. (2007). Complexity and effective dimension of discrete Lévy areas. *J. Complexity*, **23**(2): 152-168.

[122] Schmidt, R., Hrycej, T. and Stutzle, E. (2006). Multivariate Distribution Models with Generalized Hyperbolic Margins. *Computational Statistics and Data Analysis*, **50**: 2065-2096.

[123] Sobol', I.M. and Kucherenko, S.S. (2005). Global Sensitivity Indices for Nonlinear Mathematical Models. *Review, Wilmott*, **1**: 56-61.

[124] Stambaugh, R.F. and Yuan, Y. (2017). Mispricing Factors. *Review of Financial Studies*, **30**: 1270-1315.

[125] Steinortha, P. and Mitchell, O.S. (2015). Valuing variable annuities with guaranteed minimum lifetime withdrawal benefits. *Insurance: Mathematics and Economics*, **65**: 246-258.

[126] Tan, K.S. and Boyle, P.P. (2000). Applications of randomized low discrepancy sequences to the valuation of complex securities. *Journal of Economic Dynamics and Control*, **24**: 1747–1782.

[127] Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B(Methodological)*, **58**(1): 267-288.

[128] Tiong, S. (2013). Pricing inflation-linked variable annuities under stochastic interest rates. *Insurance: Mathematics and Economics*, **52**(1): 77-86.

[129] Treynor, J.L. (1965). How to Rate Management of Investment Funds. *Harvard Business Review*, **43**: 63-75.

[130] Tu, J., and Zhou G. (2004). Data-Generating Process Uncertainty: What Difference Does It Make in Portfolio Decisions. *Journal of Financial Economics* **72**(2), 385-421.

[131] Tu, C., and Zhou G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics* 99, 204-215.

[132] Tütüncü, R.H. and Koenig, M. (2004). Robust Asset Allocation. *Ann. Oper. Res.*, **132**: 157-187.

[133] Uppal, R. and Zaffaroni, P. (2017). Portfolio Choice with Model Misspecification: A Foundation for Alpha and Beta Portfolios *Working paper.*

[134] Wang, X. and Sloan, I.H. (2011). Quasi-Monte Carlo methods in financial engineering: an equivalent principle and dimension reduction. *Oper. Res.*, **59**(1): 80-95.

[135] Wang, X., and Tan, K.S. (2012). How do path generation methods affect the accuracy of quasi-Monte Carlo methods for problems in finance? *J. Complexity*, **28**(2): 250-277.

[136] Wang, X. and Tan, K.S. (2013). Pricing and hedging with discontinuous functions: quasi-Monte Carlo methods and dimension reduction. *Management Science*, **59**(2): 376-389.

[137] Xu, W., Chen, X. and Coleman T.F. (2016). The Efficient Application of Automatic Differentiation for Computing Gradients in Financial Applications. *J. Comput. Finance*, **19**(3).

[138] Xu, W., Chen, Y., Coleman, C. and Coleman, T.F. (2016). Efficient Machine Learning Methods for Risk Management of Large Variable Annuity Portfolios. *GRI TECHNICAL REPORT NO. 01.*

[139] Xu, L., and Ökten, G. (2015). High Performance Financial Simulation Using Randomized Quasi-Monte Carlo Methods. *Quantitative Finance*, **15**(8): 1425-1436.

[140] Yang, S.S. and Dai, T.S. (2013). A flexible tree for evaluating guaranteed minimum withdrawal benefits under deferred life annuity contracts with various provisions. *Insurance: Mathematics and Economics* **52**(2), 231-242.

[141] Zineb, T.B. and Gobet, E. (2016). Analytical approximation of variable annuities for small volatility and small withdrawal. *Teor. Veroyatnost. i Primenen.* 61(1), 5-25.

Appendix A

Proof of Equality 6.19

This appendix gives the proof of the Equality (6.19). If the matrix $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{aa} & 0 \\ \boldsymbol{A}_{ba} & \boldsymbol{A}_{bb} \end{bmatrix}$, then $\boldsymbol{A}_{ba}\boldsymbol{A}_{aa}^{-1} = \boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}$. i.e. Let $\boldsymbol{x} = \boldsymbol{A}_{ba}\boldsymbol{A}_{aa}^{-1}$, and $(\boldsymbol{\beta}_b^{\pi})^T = (w_b)^T\boldsymbol{x} = (w_b)^T\boldsymbol{A}_{ba}\boldsymbol{A}_{aa}^{-1}$, then for $i = 1, \cdots, a, j = a+1, \cdots, d$, the solution of solving $\boldsymbol{x}$ such that

$$
\begin{aligned}
&x_{j-a,1}A(1,1) + x_{j-a,2}A(2,1) + \cdots + x_{j-a,a}A(a,1) = A(j,1) \qquad \text{(A.1)}\\
&x_{j-a,1}A(1,2) + x_{j-a,2}A(2,2) + \cdots + x_{j-a,a}A(a,2) = A(j,2)\\
&\cdots\\
&x_{j-a,1}A(1,a) + x_{j-a,2}A(2,a) + \cdots + x_{j-a,a}A(a,a) = A(j,a),
\end{aligned}
$$

i.e. $[x_{j-a,1}x_{j-a,2}\cdots x_{j-a,a}]\boldsymbol{A}_{aa} = [A(j,1)A(j,2)\cdots A(j,a)]$, is equivalent to solve $\boldsymbol{x}$ such that

$$
\begin{aligned}
&x_{j-a,1}\Sigma(1,1) + x_{j-a,2}\Sigma(2,1) + \cdots + x_{j-a,a}\Sigma(a,1) = \Sigma(j,1) \qquad \text{(A.2)}\\
&x_{j-a,1}\Sigma(1,2) + x_{j-a,2}\Sigma(2,2) + \cdots + x_{j-a,a}\Sigma(a,2) = \Sigma(j,2)\\
&\cdots\\
&x_{j-a,1}\Sigma(1,a) + x_{j-a,2}\Sigma(2,a) + \cdots + x_{j-a,a}\Sigma(a,a) = \Sigma(j,a),
\end{aligned}
$$

i.e. $[x_{j-a,1}x_{j-a,2}\cdots x_{j-a,a}]\boldsymbol{\Sigma}_{aa} = [\Sigma(j,1)\Sigma(j,2)\cdots\Sigma(j,a)]$ **Proof:** For $j = a+1,\cdots,d$, $\boldsymbol{x} = \boldsymbol{A}_{ba}\boldsymbol{A}_{aa}^{-1}$ is the solution of

$$[x_{j-a,1}x_{j-a,2}\cdots x_{j-a,a}]\boldsymbol{A}_{aa} = [A(j,1)A(j,2)\cdots A(j,a)] \tag{A.3}$$

i.e.

$$[x_{j-a,1}x_{j-a,2}\cdots x_{j-a,a}]\boldsymbol{A}_{aa}\boldsymbol{A}_{aa}^T = [A(j,1)A(j,2)\cdots A(j,a)]A_{aa}^T,$$

where $\boldsymbol{A}_{aa}\boldsymbol{A}_{aa}^T = \boldsymbol{\Sigma}_{aa}$. Furthermore, we let $\boldsymbol{A}_{ad} = \begin{bmatrix} \boldsymbol{A}_{aa} & \boldsymbol{0} \end{bmatrix}$, then $\boldsymbol{A}_{ad}^T = \begin{bmatrix} \boldsymbol{A}_{aa} \\ \boldsymbol{0} \end{bmatrix}$. i.e.

$$\begin{aligned}
[A(j,1)A(j,2)\cdots A(j,a)]\boldsymbol{A}_{aa}^T &= [A(j,1)A(j,2)\cdots A(j,d)]\boldsymbol{A}_{ad}^T \\
&\quad \text{Due to zeros, there is no change on the value} \\
&= [\Sigma(j,1)\Sigma(j,2)\cdots\Sigma(j,a)] \\
&\quad \text{by definition } \Sigma(j,i) = \boldsymbol{A}_{j\cdot}\boldsymbol{A}_{\cdot i}
\end{aligned}$$

for $j = a+1,\cdots,d$, where $\boldsymbol{A}_{j\cdot}$ denotes the j^{th} row of $\boldsymbol{A}$ and $\boldsymbol{A}_{\cdot i}$ denotes the i^{th} column of $\boldsymbol{A}$

Hence, by substituting Equation (A.3) into the above formula, we have

$$[\Sigma(j,1)\Sigma(j,2)\cdots\Sigma(j,a)] = [x_{j-a,1}x_{j-a,2}\cdots x_{j-a,a}]A_{aa}^T = [x_{j-a,1}x_{j-a,2}\cdots x_{j-a,a}]\boldsymbol{A}_{aa}\boldsymbol{A}_{aa}^T,$$

where $\boldsymbol{A}_{aa}\boldsymbol{A}_{aa}^T = \boldsymbol{\Sigma}_{aa}$. i.e.

$$[x_{j-a,1}x_{j-a,2}\cdots x_{j-a,a}] = [\Sigma(j,1)\Sigma(j,2)\cdots\Sigma(j,a)]\Sigma_{aa}^{-1} = \boldsymbol{\Sigma}_{ba}\Sigma_{aa}^{-1},$$

for $j = a+1,\cdots,d$.

Thus,

$$\boldsymbol{x} = \boldsymbol{\Sigma_{ba}\Sigma_{aa}}^{-1}.$$

By definition, $\boldsymbol{A}_{ba}\boldsymbol{A}_{aa}^{-1} = \boldsymbol{x} = \boldsymbol{\Sigma_{ba}\Sigma_{aa}}^{-1}$

Appendix B

Proof of Selecting Stocks with Smaller Variance

$$R^2(\tilde{R}^{\pi}_{bP}) = \frac{(w^{\pi}_b)^T \boldsymbol{A}^{\pi}_{ba}(\boldsymbol{A}^{\pi}_{ba})^T w^{\pi}_b}{(w^{\pi}_b)^T \boldsymbol{A}^{\pi}_{ba}(\boldsymbol{A}^{\pi}_{ba})^T w^{\pi}_b + \Omega^{\pi}_b} = \left(\frac{(w^{\pi}_b)^T \boldsymbol{A}^{\pi}_{ba}(\boldsymbol{A}^{\pi}_{ba})^T w^{\pi}_b + \Omega^{\pi}_b}{(w^{\pi}_b)^T \boldsymbol{A}^{\pi}_{ba}(\boldsymbol{A}^{\pi}_{ba})^T w^{\pi}_b} \right)^{-1} \quad \text{(B.1)}$$

$$= \left(1 + \frac{\Omega^{\pi}_b}{(w^{\pi}_b)^T \boldsymbol{A}^{\pi}_{ba}(\boldsymbol{A}^{\pi}_{ba})^T w^{\pi}_b} \right)^{-1} \geq \left(1 + \frac{(1-\alpha)Var(\tilde{R}_P)}{(w^{\pi}_b)^T \boldsymbol{A}^{\pi}_{ba}(\boldsymbol{A}^{\pi}_{ba})^T w^{\pi}_b} \right)^{-1}. \quad \text{(B.2)}$$

Let $\boldsymbol{y_1} = (w^{\pi}_a)^T A^{\pi}_{aa}$, $\boldsymbol{y_2} = (w^{\pi}_b)^T A^{\pi}_{ba}$, $\boldsymbol{y_3} = (w^{\pi}_b)^T A^{\pi}_{bb}$. We have by triangle inequality

$$(||\boldsymbol{y_1}|| + ||\boldsymbol{y_2}||)^2 \geq Var(\tilde{R}^{EPD}_P) = ||\boldsymbol{y_1} + \boldsymbol{y_2}||^2 \geq \alpha Var(\tilde{R}_P),$$

where $||\boldsymbol{x}||$ denotes by the Euclidean norm of vector $\boldsymbol{x}$. Because $||\boldsymbol{y_1}|| = \sqrt{Var(\tilde{R}^{\pi}_{aP})}$, we have

$$||\boldsymbol{y_2}||^2 = (w^{\pi}_b)^T \boldsymbol{A}^{\pi}_{ba}(\boldsymbol{A}^{\pi}_{ba})^T w^{\pi}_b \geq \left(\sqrt{\alpha Var(\tilde{R}_P)} - \sqrt{Var(\tilde{R}^{\pi}_{aP})} \right)^2. \quad \text{(B.3)}$$

Substituting this into Equation (B.1), we want $(w_b^\pi)^T \boldsymbol{A}_{ba}^\pi (\boldsymbol{A}_{ba}^\pi)^T w_b^\pi$ as large as possible. One feasible way is to make the lower bound, i.e. $\left(\sqrt{\alpha Var(\tilde{R}_P)} - \sqrt{Var(\tilde{R}_{aP}^\pi)}\right)^2$, to be as large as possible, or equivalently, $\sqrt{Var(\tilde{R}_{aP}^\pi)}$ to be as small as possible. This completes the proof of Appendix A.

Appendix C

Robust Check

In terms of robust check, we only consider the cases with zero transaction cost or 50 bps proportional transaction cost following DeMiguel at al. (2009), and the holding period for momentum stocks is 6 months following George and Hwang (2004).

C.1 Robust Check For $\beta = 0.90$

Table C.1: Sharpe ratios for empirical data (Zero Transaction Cost, In-Sample = 120, $\beta = 0.90$)

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.1464	0.1475	0.1392	0.1456	0.1480	0.1517	0.1639	0.1413
EWP2	0.1566	0.1538	0.1532	0.1607	0.1637	0.1626	0.1778	0.1673
	(0.1201)	(0.2897)	(0.0086)	(0.0010)	(0.0054)	(0.0187)	(0.2385)	(0.0028)
EWP + EP	0.1923	0.2082	0.1859	0.1638	0.1785	0.2037	0.2070	0.1802
	(0.0005)	(0.0003)	(0.0001)	(0.2519)	(0.0500)	(0.0024)	(0.0090)	(0.0433)
EWP + EP nn	0.1962	0.2078	0.1853	0.1657	0.1784	0.2040	0.2065	0.1893
	(0.0002)	(0.0003)	(0.0002)	(0.2117)	(0.0490)	(0.0024)	(0.0096)	(0.0147)
EWP + EP nr	0.1955	0.2032	0.1892	0.1658	0.1814	0.2051	0.2005	0.1813
	(0.0003)	(0.0009)	(0.0001)	(0.2021)	(0.0298)	(0.0021)	(0.0270)	(0.0427)
MV	0.1773	0.1646	0.1527	0.1511	0.1430	0.1443	0.1407	0.1298
MV2	0.1855	0.1650	0.1623	0.1580	0.1500	0.1425	0.1466	0.1311
	(0.2413)	(0.9613)	(0.0402)	(0.2647)	(0.2847)	(0.7172)	(0.5665)	(0.9202)
MV + EP	0.1982	0.1906	0.1702	0.1512	0.1624	0.1900	0.1485	0.1233
	(0.0956)	(0.1288)	(0.1892)	(0.9982)	(0.1304)	(0.0046)	(0.4188)	(0.7009)
MV + EP nn	0.1948	0.1953	0.1672	0.1427	0.1560	0.1907	0.1470	0.1439
	(0.2326)	(0.0932)	(0.3496)	(0.6680)	(0.4410)	(0.0042)	(0.5446)	(0.5629)
MV + EP nr	0.1869	0.1900	0.1742	0.1568	0.1621	0.1868	0.1509	0.1333
	(0.4547)	(0.1330)	(0.0956)	(0.7451)	(0.1470)	(0.0072)	(0.3214)	(0.8581)
MIN	0.1644	0.1671	0.1630	0.1644	0.1848	0.1823	0.1918	0.1676
MIN2	0.1791	0.1801	0.1676	0.1788	0.1867	0.1814	0.1895	0.1845
	(0.0989)	(0.0783)	(0.6997)	(0.1271)	(0.8905)	(0.8775)	(0.8881)	(0.2813)
MIN + EP	0.1711	0.1942	0.1339	0.1762	0.1793	0.2006	0.1917	0.1839
	(0.6066)	(0.0776)	(0.0369)	(0.4271)	(0.6806)	(0.2233)	(0.9979)	(0.3594)
MIN + EP nn	0.1885	0.1942	0.1498	0.1763	0.1905	0.2006	0.1884	0.1550
	(0.1956)	(0.0776)	(0.4796)	(0.4218)	(0.7993)	(0.2233)	(0.8628)	(0.5788)
MIN + EP nr	0.1712	0.1878	0.1402	0.1767	0.1794	0.2001	0.1834	0.1697
	(0.6243)	(0.1754)	(0.1093)	(0.4013)	(0.6903)	(0.2355)	(0.6543)	(0.9228)
OC	0.1787	0.1609	0.1528	0.1378	0.1411	0.1420	0.1384	0.1251
OC2	0.1853	0.1626	0.1633	0.1497	0.1536	0.1357	0.1467	0.1344
	(0.3269)	(0.8392)	(0.0250)	(0.1051)	(0.0441)	(0.4672)	(0.4858)	(0.4943)
OC + EP	0.1991	0.1837	0.1714	0.1433	0.1653	0.1895	0.1459	0.1195
	(0.1013)	(0.1885)	(0.1618)	(0.7682)	(0.0516)	(0.0082)	(0.4138)	(0.7353)
OC + EP nn	0.1971	0.1634	0.1692	0.1266	0.1644	0.1894	0.1477	0.1441
	(0.2079)	(0.9280)	(0.2945)	(0.6223)	(0.1567)	(0.0178)	(0.3444)	(0.4264)
OC + EP nr	0.1885	0.1828	0.1751	0.1537	0.1686	0.1828	0.1498	0.1280
	(0.4396)	(0.1954)	(0.0853)	(0.3747)	(0.0295)	(0.0225)	(0.2352)	(0.8789)
VT	0.1601	0.1606	0.1503	0.1564	0.1621	0.1651	0.1815	0.1567
VT2	0.1667	0.1640	0.1617	0.1679	0.1717	0.1710	0.1923	0.1785
	(0.2296)	(0.4841)	(0.0182)	(0.0028)	(0.0638)	(0.1249)	(0.3394)	(0.0039)
VT + EP	0.1892	0.1944	0.1827	0.1657	0.1691	0.2029	0.2197	0.1922
	(0.0199)	(0.0329)	(0.0106)	(0.5468)	(0.6208)	(0.0165)	(0.0188)	(0.0571)
VT + EP nn	0.1898	0.1954	0.1831	0.1687	0.1691	0.2031	0.2197	0.1947
	(0.0174)	(0.0285)	(0.0097)	(0.4303)	(0.6193)	(0.0160)	(0.0188)	(0.0420)
VT + EP nr	0.1898	0.1898	0.1845	0.1678	0.1697	0.2020	0.2150	0.1908
	(0.0185)	(0.0674)	(0.0073)	(0.4572)	(0.5941)	(0.0200)	(0.0418)	(0.0716)
RRT	0.1625	0.1619	0.1527	0.1568	0.1624	0.1658	0.1795	0.1551
RRT2	0.1675	0.1649	0.1611	0.1658	0.1720	0.1706	0.1857	0.1724
	(0.2600)	(0.4527)	(0.0192)	(0.0022)	(0.0293)	(0.1310)	(0.6118)	(0.0424)
RRT + EP	0.1743	0.1948	0.1853	0.1559	0.1715	0.2014	0.2061	0.1878
	(0.3529)	(0.0343)	(0.0138)	(0.9505)	(0.5111)	(0.0222)	(0.0735)	(0.0692)
RRT + EP nn	0.1762	0.1956	0.1833	0.1592	0.1689	0.2019	0.2067	0.1905
	(0.2956)	(0.0302)	(0.0220)	(0.8841)	(0.6355)	(0.0201)	(0.0679)	(0.0481)
RRT + EP nr	0.1749	0.1911	0.1857	0.1577	0.1710	0.2012	0.2039	0.1897
	(0.3335)	(0.0614)	(0.0122)	(0.9588)	(0.5312)	(0.0230)	(0.1092)	(0.0654)

Table C.2: Sharpe ratios for empirical data (50 bps Transaction Cost + 1 bps fixed Transaction Cost on each trading stock, In-Sample = 120, $\beta = 0.90$)

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.0926	-0.0518	0.0862	-0.0591	0.0947	-0.0545	0.1366	0.0355
EWP2	0.1097	-0.0002	0.1078	0.0028	0.1176	0.0044	0.1489	0.0818
	(0.0095)	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.2963)	(0.0000)
EWP + EP	0.1676	0.1718	0.1576	0.1265	0.1545	0.1733	0.1771	0.1360
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0002)	(0.0000)	(0.0146)	(0.0000)
EWP + EP nn	0.1729	0.1730	0.1578	0.1304	0.1554	0.1745	0.1770	0.1493
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0148)	(0.0000)
EWP + EP nr	0.1715	0.1664	0.1606	0.1280	0.1580	0.1732	0.1710	0.1381
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0385)	(0.0000)
MV	0.1037	-0.0568	0.0799	-0.0691	0.0644	-0.0854	0.1008	0.0268
MV2	0.1214	-0.0074	0.1000	-0.0150	0.0802	-0.0362	0.1084	0.0457
	(0.0100)	(0.0000)	(0.0000)	(0.0000)	(0.0133)	(0.0000)	(0.4624)	(0.1081)
MV + EP	0.1574	0.1389	0.1328	0.0969	0.1159	0.1473	0.1143	0.0446
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.1689)	(0.3076)
MV + EP nn	0.1610	0.1549	0.1367	0.1010	0.1218	0.1514	0.1164	0.0932
	(0.0001)	(0.0000)	(0.0002)	(0.0000)	(0.0008)	(0.0000)	(0.1382)	(0.0052)
MV + EP nr	0.1529	0.1460	0.1424	0.1102	0.1268	0.1453	0.1202	0.0798
	(0.0002)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0616)	(0.0060)
MIN	0.0954	-0.0631	0.0937	-0.0747	0.1140	-0.0674	0.1558	0.0247
MIN2	0.1138	-0.0041	0.1020	-0.0146	0.1188	-0.0150	0.1495	0.0690
	(0.0425)	(0.0000)	(0.4726)	(0.0000)	(0.7054)	(0.0000)	(0.7196)	(0.0047)
MIN + EP	0.1412	0.1629	0.0967	0.1427	0.1426	0.1711	0.1576	0.1210
	(0.0009)	(0.0000)	(0.8382)	(0.0000)	(0.0372)	(0.0000)	(0.9187)	(0.0000)
MIN + EP nn	0.1646	0.1629	0.1230	0.1431	0.1639	0.1711	0.1602	0.1130
	(0.0002)	(0.0000)	(0.1402)	(0.0000)	(0.0143)	(0.0000)	(0.8191)	(0.0002)
MIN + EP nr	0.1487	0.1563	0.1150	0.1429	0.1528	0.1700	0.1556	0.1256
	(0.0001)	(0.0000)	(0.1501)	(0.0000)	(0.0058)	(0.0000)	(0.9820)	(0.0000)
OC	0.1508	0.0772	0.1271	0.0477	0.1087	0.0139	0.1153	0.0785
OC2	0.1536	0.0886	0.1354	0.0731	0.1164	0.0449	0.1212	0.0937
	(0.6871)	(0.1634)	(0.0682)	(0.0024)	(0.2109)	(0.0006)	(0.6203)	(0.2690)
OC + EP	0.1588	0.0884	0.1365	0.0684	0.1173	0.1213	0.1146	0.0369
	(0.5279)	(0.5369)	(0.4701)	(0.2742)	(0.4952)	(0.0000)	(0.9440)	(0.0146)
OC + EP nn	0.1639	0.1201	0.1418	0.0811	0.1301	0.1358	0.1207	0.0903
	(0.3767)	(0.1110)	(0.3398)	(0.1370)	(0.1883)	(0.0000)	(0.5843)	(0.6162)
OC + EP nr	0.1550	0.1219	0.1462	0.0993	0.1340	0.1256	0.1230	0.0713
	(0.7449)	(0.0091)	(0.1328)	(0.0042)	(0.0449)	(0.0000)	(0.4213)	(0.7046)
VT	0.1039	-0.0476	0.0951	-0.0564	0.1056	-0.0522	0.1519	0.0436
VT2	0.1179	0.0046	0.1143	0.0051	0.1231	0.0051	0.1611	0.0876
	(0.0106)	(0.0000)	(0.0001)	(0.0000)	(0.0008)	(0.0000)	(0.4164)	(0.0000)
VT + EP	0.1640	0.1589	0.1571	0.1303	0.1473	0.1738	0.1938	0.1490
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0044)	(0.0000)	(0.0101)	(0.0000)
VT + EP nn	0.1648	0.1604	0.1577	0.1346	0.1474	0.1741	0.1938	0.1529
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0044)	(0.0000)	(0.0101)	(0.0000)
VT + EP nr	0.1645	0.1534	0.1586	0.1318	0.1477	0.1714	0.1894	0.1468
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0042)	(0.0000)	(0.0230)	(0.0000)
RRT	0.1078	-0.0346	0.0992	-0.0421	0.1070	-0.0394	0.1480	0.0555
RRT2	0.1198	0.0109	0.1152	0.0103	0.1245	0.0110	0.1497	0.0875
	(0.0092)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.8848)	(0.0002)
RRT + EP	0.1495	0.1615	0.1589	0.1193	0.1481	0.1711	0.1781	0.1441
	(0.0013)	(0.0000)	(0.0000)	(0.0000)	(0.0041)	(0.0000)	(0.0427)	(0.0000)
RRT + EP nn	0.1520	0.1631	0.1577	0.1250	0.1463	0.1721	0.1795	0.1489
	(0.0008)	(0.0000)	(0.0000)	(0.0000)	(0.0053)	(0.0000)	(0.0341)	(0.0000)
RRT + EP nr	0.1505	0.1573	0.1593	0.1218	0.1482	0.1697	0.1770	0.1458
	(0.0010)	(0.0000)	(0.0000)	(0.0000)	(0.0037)	(0.0000)	(0.0567)	(0.0000)

Table C.3: CEQ (Risk-aversion=3) for empirical data (Zero Transaction Cost, In-Sample = 120, $\beta = 0.90$)

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.0036	0.0036	0.0032	0.0035	0.0037	0.0038	0.0042	0.0033
EWP2	0.0041	0.0039	0.0039	0.0043	0.0044	0.0044	0.0048	0.0045
	(0.1090)	(0.2854)	(0.0028)	(0.0006)	(0.0019)	(0.0087)	(0.2078)	(0.0011)
EWP + EP	0.0058	0.0068	0.0056	0.0045	0.0052	0.0065	0.0059	0.0051
	(0.0000)	(0.0000)	(0.0000)	(0.0684)	(0.0156)	(0.0001)	(0.0051)	(0.0170)
EWP + EP nn	0.0061	0.0068	0.0055	0.0046	0.0052	0.0065	0.0059	0.0056
	(0.0000)	(0.0000)	(0.0000)	(0.0412)	(0.0108)	(0.0001)	(0.0052)	(0.0022)
EWP + EP nr	0.0057	0.0060	0.0055	0.0044	0.0050	0.0060	0.0057	0.0049
	(0.0011)	(0.0081)	(0.0004)	(0.5874)	(0.1858)	(0.0528)	(0.0183)	(0.1697)
MV	0.0051	0.0044	0.0039	0.0038	0.0034	0.0035	0.0033	0.0025
MV2	0.0055	0.0045	0.0043	0.0041	0.0037	0.0034	0.0036	0.0026
	(0.1045)	(0.6547)	(0.0156)	(0.1122)	(0.1720)	(0.9763)	(0.5196)	(0.9070)
MV + EP	0.0061	0.0058	0.0047	0.0038	0.0043	0.0057	0.0037	0.0022
	(0.0717)	(0.0289)	(0.1158)	(0.8010)	(0.0692)	(0.0005)	(0.5422)	(0.5457)
MV + EP nn	0.0061	0.0062	0.0047	0.0032	0.0041	0.0058	0.0036	0.0030
	(0.0582)	(0.0035)	(0.0861)	(0.7200)	(0.1031)	(0.0003)	(0.5467)	(0.2743)
MV + EP nr	0.0055	0.0055	0.0049	0.0040	0.0043	0.0053	0.0038	0.0029
	(0.5467)	(0.2997)	(0.1394)	(0.7776)	(0.1197)	(0.0434)	(0.3148)	(0.7552)
MIN	0.0043	0.0044	0.0042	0.0043	0.0050	0.0050	0.0049	0.0041
MIN2	0.0049	0.0050	0.0044	0.0049	0.0051	0.0050	0.0050	0.0048
	(0.0426)	(0.0335)	(0.3799)	(0.1189)	(0.6034)	(0.9677)	(0.6577)	(0.0861)
MIN + EP	0.0045	0.0057	0.0029	0.0049	0.0047	0.0058	0.0047	0.0046
	(0.8278)	(0.0209)	(0.0063)	(0.2152)	(0.3348)	(0.1041)	(0.5746)	(0.3695)
MIN + EP nn	0.0054	0.0057	0.0037	0.0049	0.0055	0.0058	0.0048	0.0039
	(0.0529)	(0.0209)	(0.8531)	(0.2085)	(0.2437)	(0.1041)	(0.9218)	(0.7034)
MIN + EP nr	0.0046	0.0053	0.0033	0.0048	0.0048	0.0056	0.0048	0.0043
	(0.3287)	(0.1392)	(0.1642)	(0.4044)	(0.9420)	(0.3242)	(0.8879)	(0.4143)
OC	0.0052	0.0043	0.0039	0.0031	0.0033	0.0034	0.0031	0.0022
OC2	0.0055	0.0044	0.0044	0.0037	0.0039	0.0030	0.0036	0.0027
	(0.1621)	(0.7870)	(0.0090)	(0.0484)	(0.0159)	(0.8910)	(0.3590)	(0.4417)
OC + EP	0.0062	0.0055	0.0048	0.0034	0.0044	0.0056	0.0035	0.0019
	(0.0775)	(0.0823)	(0.1059)	(0.5861)	(0.0221)	(0.0025)	(0.5420)	(0.5913)
OC + EP nn	0.0062	0.0031	0.0048	0.0016	0.0045	0.0058	0.0036	0.0030
	(0.0496)	(0.1171)	(0.0613)	(0.6731)	(0.0184)	(0.0012)	(0.2971)	(0.1922)
OC + EP nr	0.0056	0.0052	0.0049	0.0039	0.0046	0.0051	0.0037	0.0026
	(0.5664)	(0.4705)	(0.1426)	(0.6376)	(0.0202)	(0.0803)	(0.2353)	(0.7635)
VT	0.0042	0.0042	0.0038	0.0040	0.0043	0.0044	0.0048	0.0040
VT2	0.0045	0.0044	0.0043	0.0046	0.0047	0.0047	0.0052	0.0049
	(0.1670)	(0.3935)	(0.0037)	(0.0009)	(0.0174)	(0.0597)	(0.1621)	(0.0010)
VT + EP	0.0056	0.0059	0.0053	0.0045	0.0047	0.0064	0.0062	0.0055
	(0.0042)	(0.0042)	(0.0017)	(0.2250)	(0.3755)	(0.0010)	(0.0105)	(0.0275)
VT + EP nn	0.0056	0.0060	0.0053	0.0047	0.0047	0.0064	0.0062	0.0056
	(0.0034)	(0.0032)	(0.0015)	(0.1293)	(0.3741)	(0.0009)	(0.0105)	(0.0154)
VT + EP nr	0.0054	0.0054	0.0052	0.0045	0.0046	0.0059	0.0061	0.0052
	(0.0450)	(0.1984)	(0.0166)	(0.8786)	(0.9503)	(0.1065)	(0.0121)	(0.1942)
RRT	0.0043	0.0043	0.0039	0.0041	0.0043	0.0045	0.0048	0.0039
RRT2	0.0046	0.0045	0.0043	0.0045	0.0048	0.0047	0.0050	0.0047
	(0.1833)	(0.3135)	(0.0037)	(0.0004)	(0.0067)	(0.0445)	(0.5037)	(0.0290)
RRT + EP	0.0049	0.0060	0.0055	0.0040	0.0048	0.0063	0.0058	0.0053
	(0.2513)	(0.0043)	(0.0023)	(0.6480)	(0.3409)	(0.0020)	(0.0729)	(0.0702)
RRT + EP nn	0.0050	0.0060	0.0054	0.0042	0.0047	0.0063	0.0058	0.0055
	(0.1720)	(0.0030)	(0.0033)	(0.3936)	(0.3961)	(0.0016)	(0.0521)	(0.0360)
RRT + EP nr	0.0048	0.0055	0.0053	0.0040	0.0046	0.0058	0.0058	0.0052
	(0.6114)	(0.1582)	(0.0238)	(0.6245)	(0.9422)	(0.1357)	(0.0624)	(0.3085)

Table C.4: CEQ (Risk-aversion=3) for empirical data (50 bps Transaction Cost + 1 bps fixed Transaction Cost on each trading stock, In-Sample = 120, $\beta = 0.90$)

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.0010	-0.0061	0.0006	-0.0064	0.0011	-0.0062	0.0031	-0.0016
EWP2	0.0018	-0.0036	0.0017	-0.0033	0.0022	-0.0033	0.0036	0.0006
	(0.0038)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2600)	(0.0000)
EWP + EP	0.0046	0.0049	0.0041	0.0025	0.0040	0.0049	0.0047	0.0031
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0077)	(0.0000)
EWP + EP nn	0.0049	0.0050	0.0041	0.0027	0.0040	0.0050	0.0047	0.0037
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0075)	(0.0000)
EWP + EP nr	0.0047	0.0044	0.0042	0.0027	0.0040	0.0046	0.0045	0.0031
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0002)	(0.0000)	(0.0244)	(0.0000)
MV	0.0015	-0.0061	0.0003	-0.0069	-0.0002	-0.0075	0.0013	-0.0034
MV2	0.0023	-0.0039	0.0013	-0.0044	0.0005	-0.0053	0.0017	-0.0023
	(0.0019)	(0.0000)	(0.0000)	(0.0000)	(0.0039)	(0.0000)	(0.3993)	(0.0663)
MV + EP	0.0041	0.0032	0.0029	0.0010	0.0022	0.0036	0.0020	-0.0022
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.1634)	(0.2312)
MV + EP nn	0.0043	0.0040	0.0030	0.0008	0.0023	0.0038	0.0021	-0.0003
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0973)	(0.0020)
MV + EP nr	0.0039	0.0035	0.0034	0.0019	0.0027	0.0035	0.0023	0.0001
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0379)	(0.0023)
MIN	0.0014	-0.0056	0.0013	-0.0059	0.0022	-0.0054	0.0036	-0.0010
MIN2	0.0021	-0.0030	0.0016	-0.0033	0.0024	-0.0033	0.0035	0.0005
	(0.0164)	(0.0000)	(0.3103)	(0.0000)	(0.5215)	(0.0000)	(0.9646)	(0.0012)
MIN + EP	0.0033	0.0043	0.0015	0.0034	0.0033	0.0046	0.0036	0.0024
	(0.0003)	(0.0000)	(0.9190)	(0.0000)	(0.0290)	(0.0000)	(0.7495)	(0.0000)
MIN + EP nn	0.0043	0.0043	0.0025	0.0034	0.0043	0.0046	0.0038	0.0021
	(0.0000)	(0.0000)	(0.0443)	(0.0000)	(0.0008)	(0.0000)	(0.7345)	(0.0000)
MIN + EP nr	0.0036	0.0040	0.0022	0.0034	0.0037	0.0044	0.0037	0.0026
	(0.0000)	(0.0000)	(0.0587)	(0.0000)	(0.0005)	(0.0000)	(0.6281)	(0.0000)
OC	0.0038	0.0001	0.0026	-0.0015	0.0018	-0.0028	0.0020	-0.0006
OC2	0.0039	0.0007	0.0030	-0.0002	0.0022	-0.0016	0.0022	0.0003
	(0.4967)	(0.1086)	(0.0348)	(0.0005)	(0.1313)	(0.0002)	(0.5129)	(0.2078)
OC + EP	0.0042	0.0006	0.0031	-0.0006	0.0022	0.0024	0.0020	-0.0028
	(0.5235)	(0.4460)	(0.4070)	(0.2078)	(0.4079)	(0.0000)	(0.7661)	(0.0052)
OC + EP nn	0.0045	-0.0004	0.0033	-0.0015	0.0027	0.0030	0.0022	-0.0006
	(0.1568)	(0.0629)	(0.1111)	(0.1352)	(0.0475)	(0.0000)	(0.5460)	(0.4940)
OC + EP nr	0.0040	0.0024	0.0036	0.0014	0.0030	0.0026	0.0024	-0.0004
	(0.8858)	(0.0089)	(0.1863)	(0.0021)	(0.0291)	(0.0000)	(0.4230)	(0.5877)
VT	0.0016	-0.0055	0.0012	-0.0059	0.0017	-0.0056	0.0036	-0.0009
VT2	0.0022	-0.0031	0.0020	-0.0030	0.0025	-0.0030	0.0040	0.0010
	(0.0031)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.2406)	(0.0000)
VT + EP	0.0044	0.0042	0.0041	0.0028	0.0036	0.0049	0.0052	0.0036
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0005)	(0.0000)	(0.0041)	(0.0000)
VT + EP nn	0.0044	0.0043	0.0041	0.0030	0.0036	0.0049	0.0052	0.0038
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0005)	(0.0000)	(0.0041)	(0.0000)
VT + EP nr	0.0043	0.0038	0.0041	0.0029	0.0036	0.0046	0.0051	0.0035
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0034)	(0.0000)	(0.0054)	(0.0000)
RRT	0.0018	-0.0049	0.0013	-0.0053	0.0017	-0.0050	0.0035	-0.0005
RRT2	0.0023	-0.0028	0.0021	-0.0029	0.0025	-0.0028	0.0036	0.0009
	(0.0020)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.8473)	(0.0000)
RRT + EP	0.0037	0.0043	0.0042	0.0022	0.0036	0.0048	0.0047	0.0034
	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0005)	(0.0000)	(0.0320)	(0.0000)
RRT + EP nn	0.0038	0.0044	0.0041	0.0024	0.0036	0.0048	0.0048	0.0037
	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0006)	(0.0000)	(0.0193)	(0.0000)
RRT + EP nr	0.0037	0.0040	0.0041	0.0024	0.0036	0.0045	0.0047	0.0034
	(0.0007)	(0.0000)	(0.0000)	(0.0000)	(0.0038)	(0.0000)	(0.0255)	(0.0000)

C.2 Robust Check Using 240 In-sample Data Points

Table C.5: Sharpe ratios for empirical data (Zero Transaction Cost, In-Sample = 240)

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.1464	0.1475	0.1392	0.1456	0.1480	0.1517	0.1639	0.1413
EWP2	0.1566	0.1538	0.1532	0.1607	0.1637	0.1626	0.1778	0.1673
	(0.1201)	(0.2897)	(0.0086)	(0.0010)	(0.0054)	(0.0187)	(0.2385)	(0.0028)
EWP + EP	0.1844	0.1930	0.1560	0.1962	0.1926	0.2082	0.1828	0.1870
	(0.0003)	(0.0001)	(0.1163)	(0.0003)	(0.0004)	(0.0000)	(0.1244)	(0.0017)
EWP + EP nn	0.1836	0.1863	0.1562	0.1964	0.1927	0.2080	0.1828	0.1889
	(0.0004)	(0.0011)	(0.1096)	(0.0003)	(0.0004)	(0.0000)	(0.1243)	(0.0011)
EWP + EP nr	0.1833	0.1884	0.1595	0.1970	0.1937	0.2096	0.1799	0.1887
	(0.0005)	(0.0006)	(0.0595)	(0.0002)	(0.0003)	(0.0000)	(0.1918)	(0.0011)
MV	0.1651	0.1699	0.1374	0.1490	0.1530	0.1589	0.1653	0.1267
MV2	0.1744	0.1627	0.1425	0.1591	0.1558	0.1559	0.1873	0.1236
	(0.1275)	(0.3952)	(0.3400)	(0.1066)	(0.7870)	(0.4587)	(0.0632)	(0.8166)
MV + EP	0.1822	0.1889	0.1366	0.1769	0.1616	0.1946	0.1871	0.1278
	(0.0485)	(0.0881)	(0.9388)	(0.0244)	(0.2094)	(0.0015)	(0.0690)	(0.9393)
MV + EP nn	0.1693	0.1863	0.1344	0.1741	0.1542	0.1912	0.1890	0.1422
	(0.6834)	(0.3262)	(0.7771)	(0.0553)	(0.9283)	(0.0061)	(0.0528)	(0.3487)
MV + EP nr	0.1721	0.1895	0.1385	0.1775	0.1594	0.1951	0.1826	0.1364
	(0.4032)	(0.0783)	(0.9058)	(0.0206)	(0.4934)	(0.0013)	(0.1671)	(0.5247)
MIN	0.1689	0.1645	0.1533	0.1592	0.1837	0.1690	0.1973	0.1911
MIN2	0.1774	0.1777	0.1584	0.1766	0.1890	0.1703	0.1939	0.1974
	(0.3554)	(0.0758)	(0.6186)	(0.0150)	(0.6976)	(0.8427)	(0.8407)	(0.6660)
MIN + EP	0.1683	0.1810	0.1473	0.1727	0.1839	0.1848	0.1921	0.1967
	(0.9449)	(0.1377)	(0.5953)	(0.2430)	(0.9770)	(0.1883)	(0.7584)	(0.7116)
MIN + EP nn	0.1813	0.1818	0.1675	0.1733	0.1853	0.1844	0.1989	0.1737
	(0.4207)	(0.1191)	(0.4299)	(0.2216)	(0.9510)	(0.2024)	(0.9208)	(0.3366)
MIN + EP nr	0.1725	0.1813	0.1564	0.1771	0.1825	0.1881	0.1905	0.1798
	(0.7717)	(0.1363)	(0.8262)	(0.1140)	(0.9283)	(0.1149)	(0.6934)	(0.4809)
OC	0.1665	0.1683	0.1282	0.1423	0.1542	0.1509	0.1531	0.1216
OC2	0.1733	0.1585	0.1348	0.1592	0.1552	0.1481	0.1732	0.1242
	(0.1992)	(0.3643)	(0.2674)	(0.0191)	(0.9225)	(0.6054)	(0.0613)	(0.8752)
OC + EP	0.1830	0.1770	0.1349	0.1596	0.1630	0.1843	0.1811	0.1192
	(0.0249)	(0.5062)	(0.4701)	(0.1422)	(0.1947)	(0.0034)	(0.0102)	(0.8601)
OC + EP nn	0.1727	0.1643	0.1377	0.1676	0.1500	0.1774	0.1831	0.1362
	(0.4895)	(0.8751)	(0.3770)	(0.1260)	(0.7648)	(0.0417)	(0.0062)	(0.4107)
OC + EP nr	0.1749	0.1753	0.1332	0.1666	0.1582	0.1810	0.1775	0.1315
	(0.2300)	(0.5731)	(0.5810)	(0.0423)	(0.6793)	(0.0094)	(0.0265)	(0.5421)
VT	0.1598	0.1604	0.1489	0.1555	0.1626	0.1654	0.1788	0.1569
VT2	0.1660	0.1638	0.1604	0.1678	0.1732	0.1718	0.1888	0.1781
	(0.2929)	(0.5214)	(0.0212)	(0.0021)	(0.0421)	(0.1072)	(0.3902)	(0.0056)
VT + EP	0.1872	0.2071	0.1594	0.1899	0.1989	0.2113	0.1944	0.2063
	(0.0051)	(0.0000)	(0.2733)	(0.0097)	(0.0012)	(0.0001)	(0.2126)	(0.0010)
VT + EP nn	0.1872	0.2046	0.1594	0.1912	0.1987	0.2122	0.1944	0.2059
	(0.0053)	(0.0001)	(0.2729)	(0.0075)	(0.0013)	(0.0001)	(0.2126)	(0.0011)
VT + EP nr	0.1861	0.2035	0.1624	0.1914	0.1996	0.2115	0.1930	0.2047
	(0.0078)	(0.0001)	(0.1652)	(0.0071)	(0.0009)	(0.0001)	(0.2592)	(0.0014)
RRT	0.1663	0.1662	0.1495	0.1548	0.1662	0.1694	0.1734	0.1589
RRT2	0.1713	0.1679	0.1601	0.1645	0.1742	0.1732	0.1858	0.1731
	(0.2871)	(0.6763)	(0.0052)	(0.0023)	(0.0797)	(0.2587)	(0.2200)	(0.0498)
RRT + EP	0.1847	0.1952	0.1553	0.1963	0.1876	0.2055	0.1963	0.2044
	(0.0514)	(0.0096)	(0.5694)	(0.0029)	(0.0530)	(0.0018)	(0.0594)	(0.0010)
RRT + EP nn	0.1858	0.1949	0.1570	0.1974	0.1876	0.2073	0.1974	0.2048
	(0.0426)	(0.0123)	(0.4663)	(0.0027)	(0.0527)	(0.0014)	(0.0489)	(0.0008)
RRT + EP nr	0.1834	0.1912	0.1588	0.1974	0.1883	0.2060	0.1946	0.2024
	(0.0695)	(0.0244)	(0.3615)	(0.0020)	(0.0431)	(0.0015)	(0.0810)	(0.0017)

Table C.6: Sharpe ratios for empirical data (50 bps Transaction Cost + 1 bps fixed Transaction Cost on each trading stock, In-Sample = 240)

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.0926	-0.0518	0.0862	-0.0591	0.0947	-0.0545	0.1366	0.0355
EWP2	0.1097	-0.0002	0.1078	0.0028	0.1176	0.0044	0.1489	0.0818
	(0.0095)	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.2963)	(0.0000)
EWP + EP	0.1547	0.1487	0.1291	0.1533	0.1672	0.1631	0.1521	0.1357
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.2037)	(0.0000)
EWP + EP nn	0.1565	0.1493	0.1299	0.1569	0.1686	0.1681	0.1522	0.1429
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.2024)	(0.0000)
EWP + EP nr	0.1548	0.1470	0.1321	0.1539	0.1683	0.1650	0.1495	0.1389
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2887)	(0.0000)
MV	0.0991	-0.0450	0.0592	-0.0673	0.0825	-0.0671	0.1263	0.0151
MV2	0.1193	-0.0065	0.0768	-0.0082	0.0940	-0.0184	0.1506	0.0304
	(0.0009)	(0.0000)	(0.0015)	(0.0000)	(0.2630)	(0.0000)	(0.0413)	(0.2814)
MV + EP	0.1431	0.1141	0.0964	0.1197	0.1238	0.1499	0.1570	0.0535
	(0.0000)	(0.0000)	(0.0002)	(0.0000)	(0.0000)	(0.0000)	(0.0102)	(0.0061)
MV + EP nn	0.1383	0.1348	0.1006	0.1273	0.1252	0.1523	0.1618	0.0886
	(0.0002)	(0.0000)	(0.0001)	(0.0000)	(0.0015)	(0.0000)	(0.0035)	(0.0000)
MV + EP nr	0.1404	0.1291	0.1042	0.1267	0.1299	0.1530	0.1565	0.0830
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0151)	(0.0000)
MIN	0.1039	-0.0642	0.0888	-0.0738	0.1168	-0.0689	0.1649	0.0562
MIN2	0.1155	0.0006	0.0956	-0.0076	0.1230	-0.0195	0.1565	0.0867
	(0.2032)	(0.0000)	(0.5179)	(0.0000)	(0.6440)	(0.0000)	(0.6220)	(0.0327)
MIN + EP	0.1324	0.1434	0.1015	0.1327	0.1315	0.1479	0.1580	0.1275
	(0.0093)	(0.0000)	(0.3019)	(0.0000)	(0.2890)	(0.0000)	(0.6816)	(0.0000)
MIN + EP nn	0.1532	0.1449	0.1370	0.1341	0.1538	0.1483	0.1703	0.1275
	(0.0009)	(0.0000)	(0.0065)	(0.0000)	(0.0510)	(0.0000)	(0.7461)	(0.0001)
MIN + EP nr	0.1454	0.1437	0.1274	0.1365	0.1521	0.1504	0.1627	0.1316
	(0.0004)	(0.0000)	(0.0036)	(0.0000)	(0.0152)	(0.0000)	(0.9000)	(0.0000)
OC	0.1449	0.0697	0.0943	0.0263	0.1337	0.0300	0.1295	0.0883
OC2	0.1503	0.0744	0.0966	0.0602	0.1279	0.0470	0.1487	0.0877
	(0.3030)	(0.6635)	(0.7055)	(0.0000)	(0.5692)	(0.0027)	(0.0740)	(0.9615)
OC + EP	0.1482	0.0626	0.0932	0.0691	0.1260	0.1169	0.1497	0.0452
	(0.6353)	(0.6107)	(0.9117)	(0.0012)	(0.2788)	(0.0000)	(0.0594)	(0.0041)
OC + EP nn	0.1462	0.1059	0.1044	0.1125	0.1217	0.1261	0.1550	0.0856
	(0.8700)	(0.1472)	(0.3583)	(0.0000)	(0.3968)	(0.0000)	(0.0181)	(0.8727)
OC + EP nr	0.1481	0.1003	0.0990	0.1023	0.1295	0.1246	0.1503	0.0815
	(0.6343)	(0.0122)	(0.6104)	(0.0000)	(0.6585)	(0.0000)	(0.0543)	(0.6693)
VT	0.1038	-0.0469	0.0939	-0.0567	0.1065	-0.0507	0.1498	0.0453
VT2	0.1174	0.0049	0.1131	0.0052	0.1250	0.0068	0.1581	0.0883
	(0.0207)	(0.0000)	(0.0002)	(0.0000)	(0.0005)	(0.0000)	(0.4734)	(0.0000)
VT + EP	0.1611	0.1674	0.1331	0.1507	0.1753	0.1722	0.1639	0.1584
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.2606)	(0.0000)
VT + EP nn	0.1612	0.1674	0.1331	0.1537	0.1753	0.1754	0.1639	0.1611
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.2606)	(0.0000)
VT + EP nr	0.1595	0.1635	0.1353	0.1513	0.1754	0.1716	0.1629	0.1572
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2999)	(0.0000)
RRT	0.1123	-0.0319	0.0954	-0.0470	0.1116	-0.0363	0.1437	0.0524
RRT2	0.1237	0.0130	0.1139	0.0079	0.1268	0.0137	0.1547	0.0867
	(0.0162)	(0.0000)	(0.0000)	(0.0000)	(0.0011)	(0.0000)	(0.2789)	(0.0000)
RRT + EP	0.1584	0.1547	0.1281	0.1546	0.1640	0.1656	0.1662	0.1567
	(0.0000)	(0.0000)	(0.0019)	(0.0000)	(0.0000)	(0.0000)	(0.0648)	(0.0000)
RRT + EP nn	0.1603	0.1583	0.1307	0.1603	0.1648	0.1707	0.1678	0.1586
	(0.0000)	(0.0000)	(0.0009)	(0.0000)	(0.0000)	(0.0000)	(0.0490)	(0.0000)
RRT + EP nr	0.1573	0.1520	0.1315	0.1573	0.1648	0.1666	0.1654	0.1543
	(0.0000)	(0.0000)	(0.0006)	(0.0000)	(0.0000)	(0.0000)	(0.0763)	(0.0000)

Table C.7: CEQ (Risk-aversion=3) for empirical data (Zero Transaction Cost, In-Sample = 240)

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.0036	0.0036	0.0032	0.0035	0.0037	0.0038	0.0042	0.0033
EWP2	0.0041	0.0039	0.0039	0.0043	0.0044	0.0044	0.0048	0.0045
	(0.0749)	(0.2352)	(0.0009)	(0.0001)	(0.0006)	(0.0036)	(0.1616)	(0.0003)
EWP + EP	0.0055	0.0059	0.0041	0.0061	0.0059	0.0067	0.0049	0.0054
	(0.0000)	(0.0000)	(0.0160)	(0.0000)	(0.0000)	(0.0000)	(0.1368)	(0.0001)
EWP + EP nn	0.0055	0.0057	0.0041	0.0061	0.0059	0.0068	0.0049	0.0056
	(0.0000)	(0.0000)	(0.0134)	(0.0000)	(0.0000)	(0.0000)	(0.1366)	(0.0000)
EWP + EP nr	0.0052	0.0054	0.0042	0.0057	0.0057	0.0062	0.0048	0.0053
	(0.0005)	(0.0045)	(0.0980)	(0.0014)	(0.0007)	(0.0010)	(0.1507)	(0.0027)
MV	0.0045	0.0047	0.0031	0.0037	0.0038	0.0041	0.0042	0.0026
MV2	0.0050	0.0044	0.0034	0.0042	0.0040	0.0040	0.0052	0.0025
	(0.0263)	(0.4281)	(0.1159)	(0.0151)	(0.5207)	(0.6337)	(0.0051)	(0.7105)
MV + EP	0.0054	0.0056	0.0031	0.0051	0.0042	0.0059	0.0051	0.0027
	(0.0079)	(0.0122)	(0.6840)	(0.0015)	(0.1037)	(0.0000)	(0.0294)	(0.8576)
MV + EP nn	0.0048	0.0058	0.0030	0.0050	0.0040	0.0058	0.0052	0.0033
	(0.1620)	(0.0061)	(0.6022)	(0.0011)	(0.3045)	(0.0000)	(0.0101)	(0.1400)
MV + EP nr	0.0048	0.0055	0.0032	0.0049	0.0042	0.0056	0.0050	0.0031
	(0.3547)	(0.1178)	(0.6146)	(0.0344)	(0.1789)	(0.0065)	(0.0205)	(0.5743)
MIN	0.0044	0.0043	0.0038	0.0041	0.0049	0.0045	0.0052	0.0050
MIN2	0.0048	0.0049	0.0040	0.0048	0.0052	0.0045	0.0052	0.0053
	(0.1803)	(0.0079)	(0.2939)	(0.0027)	(0.3497)	(0.8475)	(0.6960)	(0.2961)
MIN + EP	0.0044	0.0051	0.0035	0.0047	0.0048	0.0052	0.0049	0.0051
	(0.6615)	(0.0182)	(0.3420)	(0.0464)	(0.6131)	(0.0451)	(0.4095)	(0.9199)
MIN + EP nn	0.0051	0.0051	0.0045	0.0048	0.0054	0.0052	0.0053	0.0047
	(0.1245)	(0.0131)	(0.0615)	(0.0344)	(0.1197)	(0.0401)	(0.7525)	(0.8523)
MIN + EP nr	0.0046	0.0050	0.0039	0.0048	0.0050	0.0052	0.0052	0.0048
	(0.4627)	(0.0637)	(0.5104)	(0.0530)	(0.5997)	(0.1539)	(0.6659)	(0.9853)
OC	0.0046	0.0047	0.0027	0.0034	0.0039	0.0038	0.0038	0.0023
OC2	0.0049	0.0042	0.0030	0.0042	0.0040	0.0037	0.0047	0.0025
	(0.0621)	(0.3008)	(0.1426)	(0.0006)	(0.6096)	(0.8929)	(0.0034)	(0.9602)
OC + EP	0.0054	0.0051	0.0030	0.0042	0.0043	0.0054	0.0049	0.0022
	(0.0019)	(0.4220)	(0.1655)	(0.0417)	(0.1003)	(0.0002)	(0.0017)	(0.6278)
OC + EP nn	0.0049	0.0041	0.0031	0.0046	0.0037	0.0052	0.0051	0.0030
	(0.0685)	(0.1683)	(0.0591)	(0.0025)	(0.6392)	(0.0005)	(0.0003)	(0.1859)
OC + EP nr	0.0050	0.0049	0.0029	0.0045	0.0041	0.0051	0.0049	0.0028
	(0.1232)	(0.9370)	(0.3774)	(0.0500)	(0.3701)	(0.0333)	(0.0012)	(0.6367)
VT	0.0042	0.0042	0.0037	0.0040	0.0043	0.0045	0.0047	0.0040
VT2	0.0045	0.0044	0.0042	0.0046	0.0048	0.0048	0.0051	0.0049
	(0.1792)	(0.3880)	(0.0018)	(0.0002)	(0.0048)	(0.0302)	(0.1778)	(0.0006)
VT + EP	0.0055	0.0065	0.0042	0.0057	0.0061	0.0067	0.0052	0.0062
	(0.0001)	(0.0000)	(0.0684)	(0.0001)	(0.0000)	(0.0000)	(0.1826)	(0.0000)
VT + EP nn	0.0055	0.0065	0.0042	0.0058	0.0061	0.0068	0.0052	0.0062
	(0.0001)	(0.0000)	(0.0680)	(0.0001)	(0.0000)	(0.0000)	(0.1826)	(0.0000)
VT + EP nr	0.0053	0.0060	0.0043	0.0055	0.0059	0.0063	0.0053	0.0059
	(0.0059)	(0.0005)	(0.2145)	(0.0129)	(0.0015)	(0.0019)	(0.1101)	(0.0015)
RRT	0.0045	0.0045	0.0037	0.0040	0.0045	0.0046	0.0045	0.0040
RRT2	0.0047	0.0046	0.0042	0.0044	0.0049	0.0048	0.0050	0.0047
	(0.1305)	(0.4596)	(0.0001)	(0.0001)	(0.0075)	(0.0722)	(0.0379)	(0.0046)
RRT + EP	0.0054	0.0059	0.0040	0.0060	0.0055	0.0064	0.0053	0.0060
	(0.0027)	(0.0001)	(0.2231)	(0.0000)	(0.0050)	(0.0000)	(0.0256)	(0.0000)
RRT + EP nn	0.0055	0.0060	0.0041	0.0062	0.0055	0.0065	0.0053	0.0060
	(0.0015)	(0.0001)	(0.1340)	(0.0000)	(0.0038)	(0.0000)	(0.0171)	(0.0000)
RRT + EP nr	0.0052	0.0055	0.0041	0.0058	0.0054	0.0060	0.0053	0.0057
	(0.0574)	(0.0699)	(0.4102)	(0.0027)	(0.0591)	(0.0149)	(0.0109)	(0.0009)

Table C.8: CEQ (Risk-aversion=3) for empirical data (50 bps Transaction Cost + 1 bps fixed Transaction Cost on each trading stock, In-Sample = 240)

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.0010	-0.0061	0.0006	-0.0064	0.0011	-0.0062	0.0031	-0.0016
EWP2	0.0018	-0.0036	0.0017	-0.0033	0.0022	-0.0033	0.0036	0.0006
	(0.0013)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2106)	(0.0000)
EWP + EP	0.0040	0.0037	0.0027	0.0039	0.0046	0.0044	0.0037	0.0031
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2076)	(0.0000)
EWP + EP nn	0.0041	0.0037	0.0027	0.0041	0.0047	0.0047	0.0037	0.0034
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2058)	(0.0000)
EWP + EP nr	0.0039	0.0036	0.0029	0.0039	0.0045	0.0043	0.0036	0.0032
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2352)	(0.0000)
MV	0.0012	-0.0054	-0.0003	-0.0066	0.0007	-0.0063	0.0027	-0.0029
MV2	0.0022	-0.0037	0.0004	-0.0039	0.0011	-0.0042	0.0037	-0.0021
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.1192)	(0.0000)	(0.0037)	(0.1654)
MV + EP	0.0034	0.0020	0.0012	0.0023	0.0025	0.0037	0.0039	-0.0009
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0015)	(0.0004)
MV + EP nn	0.0031	0.0028	0.0014	0.0026	0.0025	0.0039	0.0041	0.0006
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)
MV + EP nr	0.0033	0.0028	0.0016	0.0027	0.0028	0.0038	0.0039	0.0005
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0005)	(0.0000)
MIN	0.0017	-0.0055	0.0011	-0.0060	0.0023	-0.0056	0.0040	0.0001
MIN2	0.0022	-0.0029	0.0013	-0.0031	0.0025	-0.0036	0.0038	0.0011
	(0.0876)	(0.0000)	(0.3168)	(0.0000)	(0.3940)	(0.0000)	(0.9001)	(0.0048)
MIN + EP	0.0029	0.0034	0.0016	0.0029	0.0028	0.0036	0.0037	0.0026
	(0.0015)	(0.0000)	(0.2214)	(0.0000)	(0.2377)	(0.0000)	(0.3922)	(0.0000)
MIN + EP nn	0.0038	0.0035	0.0031	0.0030	0.0039	0.0036	0.0042	0.0027
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0011)	(0.0000)	(0.5679)	(0.0000)
MIN + EP nr	0.0035	0.0034	0.0027	0.0031	0.0037	0.0037	0.0041	0.0029
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0004)	(0.0000)	(0.5583)	(0.0000)
OC	0.0035	-0.0003	0.0011	-0.0024	0.0030	-0.0020	0.0028	0.0006
OC2	0.0038	-0.0001	0.0012	-0.0008	0.0027	-0.0013	0.0036	0.0006
	(0.1273)	(0.5631)	(0.5932)	(0.0000)	(0.7115)	(0.0002)	(0.0073)	(0.8982)
OC + EP	0.0037	-0.0006	0.0009	-0.0004	0.0026	0.0022	0.0036	-0.0015
	(0.4167)	(0.5183)	(0.8898)	(0.0000)	(0.1843)	(0.0000)	(0.0233)	(0.0002)
OC + EP nn	0.0035	-0.0001	0.0013	0.0012	0.0023	0.0025	0.0039	0.0002
	(0.3166)	(0.0548)	(0.1141)	(0.0000)	(0.7123)	(0.0000)	(0.0020)	(0.9445)
OC + EP nr	0.0037	0.0014	0.0013	0.0015	0.0028	0.0026	0.0037	0.0004
	(0.5222)	(0.0030)	(0.4395)	(0.0000)	(0.8084)	(0.0000)	(0.0058)	(0.5051)
VT	0.0016	-0.0055	0.0011	-0.0060	0.0017	-0.0056	0.0036	-0.0009
VT2	0.0022	-0.0031	0.0020	-0.0030	0.0026	-0.0030	0.0039	0.0010
	(0.0029)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2646)	(0.0000)
VT + EP	0.0043	0.0046	0.0029	0.0038	0.0049	0.0048	0.0041	0.0041
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2080)	(0.0000)
VT + EP nn	0.0043	0.0046	0.0029	0.0039	0.0049	0.0050	0.0041	0.0042
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2080)	(0.0000)
VT + EP nr	0.0041	0.0043	0.0031	0.0038	0.0048	0.0046	0.0041	0.0039
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.1407)	(0.0000)
RRT	0.0020	-0.0047	0.0012	-0.0055	0.0020	-0.0048	0.0033	-0.0004
RRT2	0.0025	-0.0027	0.0020	-0.0030	0.0027	-0.0026	0.0038	0.0010
	(0.0014)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0754)	(0.0000)
RRT + EP	0.0041	0.0040	0.0027	0.0040	0.0044	0.0045	0.0042	0.0040
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0257)	(0.0000)
RRT + EP nn	0.0042	0.0042	0.0028	0.0043	0.0044	0.0048	0.0042	0.0040
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0154)	(0.0000)
RRT + EP nr	0.0040	0.0038	0.0029	0.0040	0.0043	0.0044	0.0042	0.0038
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0116)	(0.0000)

C.3 Robust Check Using Jorion (1986) and Ledoit and Wolf (2003) (JLW)

Table C.9: Sharpe ratios for empirical data (Zero Transaction Cost, In-Sample = 120 (JLW))

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.1464	0.1475	0.1392	0.1456	0.1480	0.1517	0.1639	0.1413
EWP2	0.1566	0.1538	0.1532	0.1607	0.1637	0.1626	0.1778	0.1673
	(0.1201)	(0.2897)	(0.0086)	(0.0010)	(0.0054)	(0.0187)	(0.2385)	(0.0028)
EWP + EP	0.1948	0.1894	0.1793	0.1860	0.1915	0.1976	0.1990	0.1683
	(0.0036)	(0.0135)	(0.0168)	(0.0357)	(0.0087)	(0.0228)	(0.0034)	(0.0406)
EWP + EP nn	0.1910	0.1874	0.1793	0.1860	0.1911	0.1951	0.1990	0.1698
	(0.0070)	(0.0180)	(0.0168)	(0.0357)	(0.0095)	(0.0303)	(0.0034)	(0.0320)
EWP + EP nr	0.1898	0.1846	0.1789	0.1844	0.1899	0.1931	0.1952	0.1644
	(0.0078)	(0.0258)	(0.0159)	(0.0408)	(0.0106)	(0.0382)	(0.0097)	(0.0862)
MV	0.1747	0.1664	0.1427	0.1546	0.1507	0.1559	0.1523	0.1530
MV2	0.1808	0.1707	0.1537	0.1633	0.1590	0.1584	0.1700	0.1558
	(0.2816)	(0.3618)	(0.0503)	(0.1159)	(0.2076)	(0.5503)	(0.1063)	(0.7887)
MV + EP	0.1933	0.1724	0.1508	0.1800	0.1679	0.1867	0.1545	0.1586
	(0.0605)	(0.6981)	(0.2841)	(0.1291)	(0.1287)	(0.0792)	(0.8362)	(0.6084)
MV + EP nn	0.1838	0.1710	0.1481	0.1797	0.1600	0.1859	0.1550	0.1608
	(0.3816)	(0.7684)	(0.4841)	(0.1337)	(0.4046)	(0.0868)	(0.8001)	(0.5350)
MV + EP nr	0.1881	0.1668	0.1539	0.1795	0.1640	0.1836	0.1591	0.1559
	(0.1885)	(0.9771)	(0.1598)	(0.1304)	(0.2601)	(0.1075)	(0.5463)	(0.8288)
MIN	0.1646	0.1692	0.1605	0.1677	0.1829	0.1818	0.1909	0.1789
MIN2	0.1788	0.1798	0.1686	0.1801	0.1882	0.1853	0.1949	0.1873
	(0.1028)	(0.1071)	(0.4877)	(0.1753)	(0.7079)	(0.6003)	(0.8481)	(0.5926)
MIN + EP	0.1713	0.1832	0.1557	0.1818	0.1798	0.1949	0.1888	0.1969
	(0.5354)	(0.2789)	(0.7000)	(0.3374)	(0.8286)	(0.3251)	(0.8983)	(0.2091)
MIN + EP nn	0.1731	0.1832	0.1551	0.1818	0.1792	0.1949	0.1856	0.1851
	(0.4434)	(0.2782)	(0.6678)	(0.3374)	(0.7979)	(0.3251)	(0.7601)	(0.7136)
MIN + EP nr	0.1761	0.1795	0.1652	0.1800	0.1809	0.1944	0.1878	0.1861
	(0.3139)	(0.4135)	(0.7462)	(0.3928)	(0.8903)	(0.3544)	(0.8622)	(0.6831)
OC	0.1799	0.1635	0.1488	0.1475	0.1471	0.1449	0.1463	0.1468
OC2	0.1838	0.1676	0.1581	0.1602	0.1562	0.1501	0.1565	0.1566
	(0.4840)	(0.4624)	(0.0404)	(0.0571)	(0.1902)	(0.3822)	(0.3640)	(0.4111)
OC + EP	0.1873	0.1757	0.1493	0.1835	0.1630	0.1818	0.1547	0.1517
	(0.3987)	(0.3533)	(0.9427)	(0.0186)	(0.1418)	(0.0551)	(0.3685)	(0.6525)
OC + EP nn	0.1830	0.1644	0.1471	0.1795	0.1562	0.1795	0.1538	0.1597
	(0.7248)	(0.9393)	(0.8361)	(0.0364)	(0.3931)	(0.0690)	(0.4377)	(0.3041)
OC + EP nr	0.1827	0.1708	0.1505	0.1787	0.1604	0.1774	0.1597	0.1542
	(0.7536)	(0.5652)	(0.8200)	(0.0366)	(0.2351)	(0.0819)	(0.1767)	(0.5413)
VT	0.1601	0.1606	0.1503	0.1564	0.1621	0.1651	0.1815	0.1567
VT2	0.1667	0.1640	0.1617	0.1679	0.1717	0.1710	0.1923	0.1785
	(0.2296)	(0.4841)	(0.0182)	(0.0028)	(0.0638)	(0.1249)	(0.3394)	(0.0039)
VT + EP	0.1927	0.1849	0.1725	0.1873	0.1889	0.1943	0.2114	0.1940
	(0.0265)	(0.1283)	(0.1553)	(0.0795)	(0.0612)	(0.1123)	(0.0110)	(0.0040)
VT + EP nn	0.1927	0.1842	0.1725	0.1873	0.1889	0.1943	0.2114	0.1941
	(0.0265)	(0.1398)	(0.1553)	(0.0795)	(0.0612)	(0.1126)	(0.0110)	(0.0039)
VT + EP nr	0.1891	0.1797	0.1711	0.1851	0.1864	0.1910	0.2079	0.1905
	(0.0440)	(0.2207)	(0.1741)	(0.0972)	(0.0838)	(0.1551)	(0.0271)	(0.0100)
RRT	0.1614	0.1615	0.1518	0.1570	0.1625	0.1655	0.1789	0.1560
RRT2	0.1670	0.1648	0.1616	0.1674	0.1723	0.1711	0.1884	0.1746
	(0.2625)	(0.4660)	(0.0217)	(0.0038)	(0.0405)	(0.1116)	(0.3655)	(0.0094)
RRT + EP	0.1916	0.1851	0.1711	0.1860	0.1901	0.1956	0.2105	0.1950
	(0.0367)	(0.1338)	(0.2189)	(0.0973)	(0.0472)	(0.0985)	(0.0069)	(0.0020)
RRT + EP nn	0.1916	0.1842	0.1711	0.1860	0.1901	0.1953	0.2105	0.1950
	(0.0367)	(0.1494)	(0.2189)	(0.0973)	(0.0472)	(0.1006)	(0.0069)	(0.0019)
RRT + EP nr	0.1876	0.1797	0.1698	0.1842	0.1872	0.1919	0.2076	0.1925
	(0.0632)	(0.2357)	(0.2422)	(0.1137)	(0.0693)	(0.1417)	(0.0138)	(0.0043)

Table C.10: Sharpe ratios for empirical data (50 bps Transaction Cost + 1 bps fixed Transaction Cost on each trading stock, In-Sample = 120 (JLW))

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.0926	-0.0518	0.0862	-0.0591	0.0947	-0.0545	0.1366	0.0355
EWP2	0.1097	-0.0002	0.1078	0.0028	0.1176	0.0044	0.1489	0.0818
	(0.0095)	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.2963)	(0.0000)
EWP + EP	0.1681	0.1519	0.1526	0.1517	0.1643	0.1620	0.1692	0.1183
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0064)	(0.0000)
EWP + EP nn	0.1651	0.1513	0.1526	0.1517	0.1641	0.1606	0.1692	0.1219
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0064)	(0.0000)
EWP + EP nr	0.1623	0.1449	0.1509	0.1469	0.1615	0.1543	0.1655	0.1124
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.0169)	(0.0000)
MV	0.0996	-0.0545	0.0694	-0.0684	0.0722	-0.0723	0.1092	0.0317
MV2	0.1161	-0.0029	0.0915	-0.0144	0.0904	-0.0214	0.1286	0.0583
	(0.0029)	(0.0000)	(0.0001)	(0.0000)	(0.0044)	(0.0000)	(0.0787)	(0.0075)
MV + EP	0.1541	0.1232	0.1129	0.1419	0.1249	0.1476	0.1183	0.0866
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.4116)	(0.0000)
MV + EP nn	0.1473	0.1244	0.1120	0.1417	0.1195	0.1471	0.1196	0.1002
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.3416)	(0.0000)
MV + EP nr	0.1513	0.1176	0.1176	0.1391	0.1230	0.1420	0.1238	0.0947
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.1940)	(0.0000)
MIN	0.0966	-0.0630	0.0921	-0.0716	0.1127	-0.0694	0.1555	0.0363
MIN2	0.1147	-0.0040	0.1039	-0.0091	0.1213	-0.0124	0.1561	0.0719
	(0.0411)	(0.0000)	(0.3122)	(0.0000)	(0.5259)	(0.0000)	(0.9836)	(0.0248)
MIN + EP	0.1376	0.1499	0.1188	0.1479	0.1423	0.1638	0.1558	0.1338
	(0.0002)	(0.0000)	(0.0409)	(0.0000)	(0.0320)	(0.0000)	(0.9884)	(0.0000)
MIN + EP nn	0.1414	0.1499	0.1222	0.1479	0.1456	0.1638	0.1544	0.1340
	(0.0001)	(0.0000)	(0.0250)	(0.0000)	(0.0230)	(0.0000)	(0.9470)	(0.0000)
MIN + EP nr	0.1452	0.1455	0.1334	0.1452	0.1489	0.1623	0.1574	0.1352
	(0.0000)	(0.0000)	(0.0028)	(0.0000)	(0.0167)	(0.0000)	(0.9306)	(0.0000)
OC	0.1464	0.0244	0.1131	-0.0079	0.1124	-0.0331	0.1200	0.1070
OC2	0.1486	0.0495	0.1235	0.0291	0.1161	0.0138	0.1263	0.1176
	(0.6897)	(0.0000)	(0.0193)	(0.0000)	(0.5797)	(0.0000)	(0.5712)	(0.3695)
OC + EP	0.1463	0.0994	0.1090	0.1257	0.1159	0.1292	0.1226	0.0830
	(0.9984)	(0.0000)	(0.5680)	(0.0000)	(0.7407)	(0.0000)	(0.7813)	(0.0252)
OC + EP nn	0.1443	0.1021	0.1088	0.1247	0.1121	0.1296	0.1229	0.1048
	(0.8303)	(0.0000)	(0.5620)	(0.0000)	(0.9880)	(0.0000)	(0.7598)	(0.8432)
OC + EP nr	0.1437	0.1028	0.1121	0.1202	0.1157	0.1237	0.1285	0.0987
	(0.7866)	(0.0000)	(0.8911)	(0.0000)	(0.7542)	(0.0000)	(0.3888)	(0.4761)
VT	0.1039	-0.0476	0.0951	-0.0564	0.1056	-0.0522	0.1519	0.0436
VT2	0.1179	0.0046	0.1143	0.0051	0.1231	0.0051	0.1611	0.0876
	(0.0106)	(0.0000)	(0.0001)	(0.0000)	(0.0008)	(0.0000)	(0.4164)	(0.0000)
VT + EP	0.1685	0.1510	0.1458	0.1543	0.1632	0.1621	0.1817	0.1492
	(0.0000)	(0.0000)	(0.0013)	(0.0000)	(0.0001)	(0.0000)	(0.0111)	(0.0000)
VT + EP nn	0.1685	0.1506	0.1458	0.1543	0.1632	0.1622	0.1817	0.1494
	(0.0000)	(0.0000)	(0.0013)	(0.0000)	(0.0001)	(0.0000)	(0.0111)	(0.0000)
VT + EP nr	0.1640	0.1439	0.1436	0.1498	0.1599	0.1560	0.1786	0.1435
	(0.0000)	(0.0000)	(0.0017)	(0.0000)	(0.0001)	(0.0000)	(0.0252)	(0.0000)
RRT	0.1056	-0.0465	0.0970	-0.0547	0.1063	-0.0496	0.1492	0.0421
RRT2	0.1188	0.0060	0.1149	0.0056	0.1242	0.0070	0.1575	0.0844
	(0.0082)	(0.0000)	(0.0000)	(0.0000)	(0.0002)	(0.0000)	(0.4342)	(0.0000)
RRT + EP	0.1675	0.1513	0.1447	0.1534	0.1639	0.1632	0.1811	0.1510
	(0.0000)	(0.0000)	(0.0026)	(0.0000)	(0.0000)	(0.0000)	(0.0062)	(0.0000)
RRT + EP nn	0.1675	0.1507	0.1447	0.1534	0.1639	0.1631	0.1811	0.1511
	(0.0000)	(0.0000)	(0.0026)	(0.0000)	(0.0000)	(0.0000)	(0.0062)	(0.0000)
RRT + EP nr	0.1628	0.1441	0.1425	0.1494	0.1602	0.1569	0.1786	0.1465
	(0.0001)	(0.0000)	(0.0033)	(0.0000)	(0.0001)	(0.0000)	(0.0115)	(0.0000)

Table C.11: CEQ (Risk-aversion=3) for empirical data (Zero Transaction Cost, In-Sample = 120 (JLW))

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.0036	0.0036	0.0032	0.0035	0.0037	0.0038	0.0042	0.0033
EWP2	0.0041	0.0039	0.0039	0.0043	0.0044	0.0044	0.0048	0.0045
	(0.1090)	(0.2854)	(0.0028)	(0.0006)	(0.0019)	(0.0087)	(0.2078)	(0.0011)
EWP + EP	0.0060	0.0058	0.0052	0.0056	0.0059	0.0062	0.0056	0.0046
	(0.0004)	(0.0010)	(0.0023)	(0.0041)	(0.0008)	(0.0024)	(0.0029)	(0.0106)
EWP + EP nn	0.0058	0.0057	0.0052	0.0056	0.0058	0.0061	0.0056	0.0047
	(0.0007)	(0.0013)	(0.0023)	(0.0041)	(0.0009)	(0.0032)	(0.0029)	(0.0062)
EWP + EP nr	0.0054	0.0052	0.0050	0.0052	0.0054	0.0055	0.0054	0.0043
	(0.0368)	(0.1394)	(0.0550)	(0.1769)	(0.0754)	(0.2620)	(0.0088)	(0.2853)
MV	0.0049	0.0045	0.0034	0.0040	0.0037	0.0040	0.0037	0.0039
MV2	0.0052	0.0047	0.0039	0.0044	0.0041	0.0041	0.0045	0.0040
	(0.1233)	(0.2355)	(0.0146)	(0.0913)	(0.0961)	(0.4145)	(0.0390)	(0.5580)
MV + EP	0.0058	0.0049	0.0038	0.0053	0.0045	0.0055	0.0038	0.0041
	(0.0227)	(0.3089)	(0.1962)	(0.0230)	(0.0605)	(0.0204)	(0.9308)	(0.7459)
MV + EP nn	0.0054	0.0048	0.0037	0.0052	0.0042	0.0055	0.0038	0.0043
	(0.1464)	(0.3230)	(0.3023)	(0.0241)	(0.1855)	(0.0222)	(0.9814)	(0.4211)
MV + EP nr	0.0056	0.0045	0.0039	0.0050	0.0043	0.0051	0.0040	0.0040
	(0.1172)	(0.7411)	(0.1136)	(0.2445)	(0.1948)	(0.3597)	(0.3931)	(0.9997)
MIN	0.0043	0.0045	0.0041	0.0044	0.0049	0.0050	0.0048	0.0045
MIN2	0.0049	0.0050	0.0044	0.0050	0.0052	0.0051	0.0052	0.0049
	(0.0365)	(0.0816)	(0.2395)	(0.1366)	(0.4341)	(0.4404)	(0.3961)	(0.2720)
MIN + EP	0.0045	0.0052	0.0038	0.0052	0.0047	0.0056	0.0046	0.0051
	(0.7664)	(0.0728)	(0.3871)	(0.1030)	(0.6522)	(0.1286)	(0.5463)	(0.1508)
MIN + EP nn	0.0046	0.0052	0.0038	0.0052	0.0047	0.0056	0.0046	0.0049
	(0.5343)	(0.0724)	(0.4799)	(0.1030)	(0.7884)	(0.1286)	(0.5274)	(0.2928)
MIN + EP nr	0.0048	0.0050	0.0043	0.0049	0.0049	0.0054	0.0049	0.0050
	(0.1456)	(0.3852)	(0.5432)	(0.4159)	(0.6578)	(0.4529)	(0.6579)	(0.2545)
OC	0.0052	0.0044	0.0037	0.0036	0.0036	0.0035	0.0036	0.0036
OC2	0.0054	0.0046	0.0041	0.0042	0.0040	0.0038	0.0040	0.0041
	(0.3032)	(0.4266)	(0.0116)	(0.0321)	(0.0901)	(0.3108)	(0.2008)	(0.2665)
OC + EP	0.0056	0.0051	0.0037	0.0054	0.0043	0.0053	0.0039	0.0038
	(0.2619)	(0.1202)	(0.9023)	(0.0030)	(0.0746)	(0.0198)	(0.4974)	(0.7665)
OC + EP nn	0.0054	0.0045	0.0036	0.0052	0.0040	0.0052	0.0039	0.0042
	(0.4369)	(0.2994)	(0.9552)	(0.0062)	(0.1893)	(0.0224)	(0.5530)	(0.1767)
OC + EP nr	0.0054	0.0047	0.0038	0.0050	0.0042	0.0049	0.0042	0.0040
	(0.6708)	(0.7065)	(0.9294)	(0.0970)	(0.1815)	(0.2756)	(0.1543)	(0.6870)
VT	0.0042	0.0042	0.0038	0.0040	0.0043	0.0044	0.0048	0.0040
VT2	0.0045	0.0044	0.0043	0.0046	0.0047	0.0047	0.0052	0.0049
	(0.1670)	(0.3935)	(0.0037)	(0.0009)	(0.0174)	(0.0597)	(0.1621)	(0.0010)
VT + EP	0.0058	0.0055	0.0048	0.0056	0.0056	0.0059	0.0059	0.0056
	(0.0048)	(0.0250)	(0.0526)	(0.0120)	(0.0128)	(0.0224)	(0.0055)	(0.0008)
VT + EP nn	0.0058	0.0055	0.0048	0.0056	0.0056	0.0059	0.0059	0.0056
	(0.0048)	(0.0279)	(0.0526)	(0.0120)	(0.0128)	(0.0225)	(0.0055)	(0.0008)
VT + EP nr	0.0054	0.0050	0.0046	0.0052	0.0052	0.0053	0.0058	0.0052
	(0.1086)	(0.5144)	(0.3216)	(0.2644)	(0.2886)	(0.5406)	(0.0066)	(0.0478)
RRT	0.0043	0.0043	0.0038	0.0041	0.0043	0.0045	0.0046	0.0039
RRT2	0.0045	0.0044	0.0043	0.0046	0.0048	0.0047	0.0051	0.0047
	(0.1851)	(0.3495)	(0.0046)	(0.0011)	(0.0094)	(0.0462)	(0.1380)	(0.0013)
RRT + EP	0.0057	0.0055	0.0048	0.0056	0.0057	0.0060	0.0058	0.0056
	(0.0082)	(0.0247)	(0.0960)	(0.0158)	(0.0097)	(0.0207)	(0.0021)	(0.0002)
RRT + EP nn	0.0057	0.0055	0.0048	0.0056	0.0057	0.0060	0.0058	0.0056
	(0.0082)	(0.0288)	(0.0960)	(0.0158)	(0.0097)	(0.0212)	(0.0021)	(0.0002)
RRT + EP nr	0.0053	0.0050	0.0046	0.0052	0.0053	0.0054	0.0058	0.0053
	(0.1533)	(0.5075)	(0.4417)	(0.2803)	(0.2576)	(0.5191)	(0.0018)	(0.0144)

Table C.12: CEQ (Risk-aversion=3) for empirical data (50 bps Transaction Cost + 1 bps fixed Transaction Cost on each trading stock, In-Sample = 120 (JLW))

Strategy	BM-25	BM-100	OP-25	OP-100	INV-25	INV-100	IND-10	IND-49
EWP	0.0010	-0.0061	0.0006	-0.0064	0.0011	-0.0062	0.0031	-0.0016
EWP2	0.0018	-0.0036	0.0017	-0.0033	0.0022	-0.0033	0.0036	0.0006
	(0.0038)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2600)	(0.0000)
EWP + EP	0.0047	0.0038	0.0039	0.0038	0.0045	0.0044	0.0044	0.0022
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0047)	(0.0000)
EWP + EP nn	0.0045	0.0038	0.0039	0.0038	0.0045	0.0043	0.0044	0.0024
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0047)	(0.0000)
EWP + EP nr	0.0042	0.0035	0.0037	0.0035	0.0042	0.0038	0.0042	0.0021
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0002)	(0.0000)	(0.0132)	(0.0000)
MV	0.0013	-0.0058	-0.0001	-0.0064	0.0002	-0.0066	0.0020	-0.0019
MV2	0.0021	-0.0034	0.0010	-0.0039	0.0010	-0.0043	0.0028	-0.0007
	(0.0003)	(0.0000)	(0.0000)	(0.0000)	(0.0008)	(0.0000)	(0.0313)	(0.0023)
MV + EP	0.0039	0.0024	0.0020	0.0033	0.0026	0.0036	0.0023	0.0008
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.4560)	(0.0000)
MV + EP nn	0.0036	0.0024	0.0019	0.0033	0.0023	0.0036	0.0024	0.0014
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.3689)	(0.0000)
MV + EP nr	0.0038	0.0023	0.0022	0.0032	0.0025	0.0033	0.0026	0.0012
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.1087)	(0.0000)
MIN	0.0014	-0.0055	0.0013	-0.0058	0.0021	-0.0054	0.0036	-0.0006
MIN2	0.0022	-0.0030	0.0017	-0.0031	0.0025	-0.0031	0.0037	0.0006
	(0.0142)	(0.0000)	(0.1849)	(0.0000)	(0.3522)	(0.0000)	(0.6171)	(0.0091)
MIN + EP	0.0031	0.0037	0.0024	0.0036	0.0033	0.0043	0.0035	0.0029
	(0.0001)	(0.0000)	(0.0334)	(0.0000)	(0.0158)	(0.0000)	(0.7422)	(0.0000)
MIN + EP nn	0.0033	0.0037	0.0025	0.0036	0.0034	0.0043	0.0035	0.0029
	(0.0000)	(0.0000)	(0.0131)	(0.0000)	(0.0070)	(0.0000)	(0.7759)	(0.0000)
MIN + EP nr	0.0035	0.0035	0.0030	0.0035	0.0036	0.0041	0.0038	0.0030
	(0.0000)	(0.0000)	(0.0003)	(0.0000)	(0.0017)	(0.0000)	(0.5472)	(0.0000)
OC	0.0036	-0.0024	0.0020	-0.0039	0.0020	-0.0051	0.0023	0.0015
OC2	0.0037	-0.0012	0.0025	-0.0021	0.0022	-0.0028	0.0026	0.0021
	(0.5235)	(0.0000)	(0.0050)	(0.0000)	(0.4478)	(0.0000)	(0.4103)	(0.2570)
OC + EP	0.0036	0.0011	0.0018	0.0025	0.0022	0.0027	0.0025	0.0004
	(0.8981)	(0.0000)	(0.5041)	(0.0000)	(0.6620)	(0.0000)	(0.9358)	(0.0080)
OC + EP nn	0.0035	0.0011	0.0017	0.0025	0.0020	0.0027	0.0025	0.0014
	(0.9672)	(0.0000)	(0.5590)	(0.0000)	(0.8470)	(0.0000)	(0.8818)	(0.8962)
OC + EP nr	0.0034	0.0015	0.0019	0.0024	0.0022	0.0025	0.0027	0.0012
	(0.8053)	(0.0000)	(0.7595)	(0.0000)	(0.7370)	(0.0000)	(0.3671)	(0.3373)
VT	0.0016	-0.0055	0.0012	-0.0059	0.0017	-0.0056	0.0036	-0.0009
VT2	0.0022	-0.0031	0.0020	-0.0030	0.0025	-0.0030	0.0040	0.0010
	(0.0031)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.2406)	(0.0000)
VT + EP	0.0046	0.0038	0.0035	0.0040	0.0044	0.0043	0.0047	0.0036
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0050)	(0.0000)
VT + EP nn	0.0046	0.0038	0.0035	0.0040	0.0044	0.0044	0.0047	0.0037
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0050)	(0.0000)
VT + EP nr	0.0043	0.0034	0.0034	0.0037	0.0041	0.0039	0.0047	0.0033
	(0.0000)	(0.0000)	(0.0012)	(0.0000)	(0.0002)	(0.0000)	(0.0065)	(0.0000)
RRT	0.0017	-0.0055	0.0012	-0.0059	0.0017	-0.0055	0.0035	-0.0009
RRT2	0.0023	-0.0031	0.0021	-0.0030	0.0025	-0.0030	0.0039	0.0009
	(0.0020)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2122)	(0.0000)
RRT + EP	0.0046	0.0038	0.0035	0.0039	0.0044	0.0044	0.0047	0.0037
	(0.0000)	(0.0000)	(0.0003)	(0.0000)	(0.0000)	(0.0000)	(0.0017)	(0.0000)
RRT + EP nn	0.0046	0.0038	0.0035	0.0039	0.0044	0.0044	0.0047	0.0037
	(0.0000)	(0.0000)	(0.0003)	(0.0000)	(0.0000)	(0.0000)	(0.0017)	(0.0000)
RRT + EP nr	0.0043	0.0034	0.0034	0.0037	0.0041	0.0039	0.0047	0.0035
	(0.0000)	(0.0000)	(0.0028)	(0.0000)	(0.0002)	(0.0000)	(0.0016)	(0.0000)

Appendix D

Proof of Theorem 1.1.4 according to Wang and Tan (2012)

Here's the proof of Theorem 1.1.4 according to Wang and Tan (2012).

Proof: From (1.13), we have

$$\boldsymbol{q}^T \boldsymbol{A_0} = D\boldsymbol{U_1}^T.$$

Define the columns of $\boldsymbol{U}$ as $\boldsymbol{U_1}, ..., \boldsymbol{U_d}$. Under the transformation $\boldsymbol{x} = \boldsymbol{A_0}\boldsymbol{U}\boldsymbol{z}$, we have

$$\boldsymbol{q}^T\boldsymbol{x} = \boldsymbol{q}^T\boldsymbol{A_0}\boldsymbol{U}\boldsymbol{z} = D\boldsymbol{U_1}^T(\boldsymbol{U_1}z_1 + \boldsymbol{U_2}z_2 + \cdots + \boldsymbol{U_d}z_d) = Dz_1,$$

by the orthogonality of the columns $\boldsymbol{U_1}, ..., \boldsymbol{U_d}$ of $\boldsymbol{U}$. Therefore,

$$h(\boldsymbol{q}^T\boldsymbol{x}) = h(Dz_1).$$

This proves the first part of the theorem.

The remaining results follow form the following equivalences (under the transformations

$\boldsymbol{x} = \boldsymbol{A_0 U z}$ and $\boldsymbol{z} = \Phi^{-1}(\boldsymbol{u})$) that

$$\{h(\boldsymbol{q}^T \boldsymbol{x} < H\} \Longleftrightarrow \{h(Dz_1) < H\} \Longleftrightarrow \{u_1 < c\}.\square$$

Appendix E

Special Case of Effective Portfolio

Theorem E.0.1. *Suppose* $\tilde{\mu}_1 = \cdots = \tilde{\mu}_N = \tilde{\mu}$, *and*

$$\mathbf{\Sigma}_{aa} = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_1 \\ \cdots & \cdots & \cdots & \cdots \\ \rho_1 & \rho_1 & \cdots & 1 \end{bmatrix},$$

$$\mathbf{\Sigma}_{ba} = \begin{bmatrix} \rho_2 & \rho_2 & \cdots & \rho_2 \\ \rho_2 & \rho_2 & \cdots & \rho_2 \\ \cdots & \cdots & \cdots & \cdots \\ \rho_2 & \rho_2 & \cdots & \rho_2 \end{bmatrix}.$$

We have $\beta_{EP} < 1$ *and* $E(\tilde{R}_{EP}) > E(\tilde{R}_M)$ *if and only if* $\frac{a\rho_2}{1+(a-1)\rho_1} > 1$. *i.e.* $\rho_2 > \rho_1$ *and* $a \gg 0$.

Proof: If $\boldsymbol{w}$ are equal weight, then by inspection, $E(\tilde{R}_{EP}) > E(\tilde{R}_M)$ iff

$$\frac{a}{N} + \frac{b}{N}\frac{a\rho_2}{1+(a-1)\rho_1} > 1.$$

i.e. $\frac{a\rho_2}{1+(a-1)\rho_1} > 1$.

Note: this could be generalized for any arbitrary weight $\boldsymbol{w}$, $E(\tilde{R}_{EP}) > E(\tilde{R}_{TP})$ if

$$\frac{a\rho_2}{1+(a-1)\rho_1} > 1.$$

i.e. $\rho_2 > \rho_1$ and $a \gg 0$.