

Unfolding Structures in Financial Data using Copula and Topological Data Analysis for Portfolio Optimization

ANUBHA GOEL



DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY DELHI
October 2019

Unfolding Structures in Financial Data using Copula and Topological Data Analysis for Portfolio Optimization

by

Anubha Goel

Department of Mathematics

submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the



INDIAN INSTITUTE OF TECHNOLOGY DELHI
October 2019

Dedicated To My Family

Certificate

This is to certify that the thesis entitled “**Unfolding Structures in Financial Data using Copula and Topological Data Analysis for Portfolio Optimization**” submitted by **Ms. Anubha Goel**, to the Indian Institute of Technology Delhi, for the award of the Degree of **Doctor of Philosophy**, is a record of the original bonafide research work carried out by her under my supervision. The thesis has reached the standards fulfilling the requirements of the regulations relating to the degree. The results obtained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree or diploma.

Dr. Aparna Mehra

Professor

Department of Mathematics

Indian Institute of Technology Delhi

New Delhi 110 016

Date

Acknowledgements

This thesis is the culmination of my long voyage of Ph.D. that was much similar to mounting a high pinnacle gradually, accompanied by hardship, consolation, and frustration. This journey has genuinely been a transformative encounter for me, and it would not have been conceivable without the help of numerous individuals. I need to stretch out my genuine appreciation to each individual who fundamentally added to achieve this gigantic task.

I would like to express my sincere gratitude to my supervisor Prof. Aparna Mehra, for her constant support and encouragement through these significant years of my life. Her direction and encouragement was essential to achieve this thesis, which would not had been conceivable without her significant support and immense knowledge. I am especially thankful for her understanding and modesty in her dealings with me, particularly amid times in which I encountered difficulties in conveying everything that needs to be conveyed as fluently as one would have loved.

I wish to convey my gratitude to my SRC (Student Research Committee) members: Prof. S. Dharmaraja, Prof. Niladri Chatterjee and Prof. Sourabh B. Paul, for their valuable time and suggestions. I would also like to extend my sincere thanks to Prof. Karmeshu, Shiv Nadar University (SNU), for taking his time out for evaluating my research progress as and when required. I am particularly appreciative to all the staff individuals from the Department of Mathematics, IIT Delhi, who have dependably been agreeable, friendly and steady. I am grateful to the Mathematics department for the academic database "EIKON". I am thankful to Council of Scientific & Industrial Research (CSIR), India for providing me the financial grant to carry out my research work. An extraordinary affirmation to IIT Delhi authorities, security staff, and hostel authorities who all by and large make IIT Delhi a safe spot to live and work till late hours.

I want to express my sincere thanks to Dr. Amita Sharma (Indian Institute of Information and Technology, Guwahati, India) for the insightful discussions and collaborative research. I value her direction, dynamic help and her many lighting up ideas. My special words of thanks to her for her immense support and friendliness throughout this journey.

Besides, I need to thank my companions and especially my best friend and research companion Dr. Puneet Pasricha for the extraordinary time we had together all through our research and for supporting and dependably having an ear for me. He was always there close to me to spur me amid the up and downs of my voyage of Ph.D. A special thanks goes to all the friends Divya, Nitin, Rohit, Abhishek and Jagdeep for all the fun and parties we have had together. I would like to thank Jyoti and Ankita who have always been great pillars. I will always cherish their friendship and support. Last, but not the least, I would like to thank my fellow Ph.D students Nitu and Ruchiqa.

Finally, I owe all my most profound appreciation towards my parents for their endless support, unconditional love, and comprehension of my objectives and aspirations. I will be everlastingly obliged to them for their sacrificial care, tolerance, pains and penances they did to make me accomplish what I desired. My deepest gratitude to my brother Shubham and my sisters, Swati, Aakriti and Sanjoli for their infallible love, valuable prayers and constant support all throughout these years. Above all, I thank the Almighty for his bounty of blessings and sunshine in my life.

October 2019

Anubha Goel

Abstract

Financial portfolio optimization and investment decision making, which adhere to allocating weights to financial assets constituting a portfolio, are extensively researched areas. The banks, financial institutions, fund management firms, investors be it Government, corporate or individuals, all require to manage their assets and hedge their risk by selecting, balancing, and continuously evaluating their portfolios allocations. The breakthrough article by Markowitz in 1952, which introduced the World to the modern portfolio theory, has been the center of criticism of late due to the assumption of the multivariate normal distribution of returns, a fact far away from the truth in the real financial data.

The complex dependence structure in high dimension due to the applications of big data has made it essential and challenging to model the multivariate dependence between financial variables. The dependency in the financial markets has increased manifold over the years due to advancement in technology, open economy, and global trades. Owing to the expansion of borders of the financial markets, with investors investing in different markets of the world, add on more to the complexity of the system. The random variables representing returns of the assets are typically asymmetric, possess anonymous cyclic behavior and stylized features, high kurtosis, tail dependence, and downside-upside momentum trends. Moreover, the economies worldwide have witnessed several significant crises, economic imbalances, and extremely volatile fluctuations in the financial data as a result of inappropriate risk management. As a repercussion, the methods of financial risks management were put to in-depth scrutiny majorly after the sub-prime global financial crisis of 2008. Coming up with renewing models has become an essential and uphill task for investors who are concerned with minimizing risk in their investments. Several models have been put forward in investment decision making and risk management, which goes hand in hand, to minimize the risk earning profits at the same time modifying the original Markowitz portfolio optimization model by bringing in more realistic features to capture the complex dependency structures in the underline of the financial assets.

This thesis aims to establish a new approach in financial investments and risk management which can proficiently encapsulate the dependency structure and shapes in the financial data. The aim is carried out using two approaches: Copula Theory and Topological Data Analysis.

Copulas have been used to model the underline dependence structure between the financial variables. According to the theory put forward by Sklar, copulas are the functions which couples the multivariate distribution to its one-dimensional marginals and thus enabling to investigate the relationship between marginals and joint distribution of the random variables. More specifically, in this thesis, we work with dynamic vine copulas, with bivariate copulas as their building blocks. A vine copula provides a flexible mechanism for modeling various distributions prevailing among the real data. They have been masterly designed to invade all possible and complex nonlinear dependency structure. We apply the regular vines copulas with rolling window scheme to make our study dynamic and robust towards data tweaks.

We design a dynamic copula based scenario generation method using the time series modeling. The residuals in the time series model, fitted in the return data series of

the assets, are pseudo-randomly generated using inverse probability transforms and vine copula. The simulated residues are in turn used to generate scenarios for returns of the assets. The proposed scenario (data) generation process is shown to ameliorate the investment process by boosting the risk-reward ratios in the optimal portfolios. This novel methodology is applied throughout the thesis for generating data from different scenarios that could be prevalent in the markets and captured by vine copula.

Besides modeling the dependency structure, this thesis investigates two crucial problems of index tracking and enhanced indexing in the domain of portfolio optimization. One of the highlighting features of our research is the development of the deviation based risk measures which were introduced long back but remained hidden since then. We propose the deviation based second order stochastic dominance criterion in portfolio optimization and apply it to the problem of enhanced indexing. The other salient part is the prelude of deviation measure with mixed conditional value at risk, inheriting all the useful properties of the conditional value at risk (CVaR). We also formulate the two-tailed mixed CVaR with deviation measure to develop a new model for index tracking problem.

Furthermore, in the same work, we design a two step procedure for enhanced indexing problem. Firstly, we introduce a Markov chain based filtration criterion to build-in the asset-picking ability in the model. The approach is quite beneficial yet straightforward and reduces the computational efforts in solving the proposed deviation based mixed CVaR enhanced indexing model which is solved on the filtered assets at the second step. The mixed CVaR with deviation measure holds more information on the distributions of returns than the conventional CVaR thereby resulting in portfolios with superior performance.

In another research work, included in the thesis, a new robust optimization model concerning change in the dependency structure is formulated for the risk-reward ratios to protect against the worst possible scenarios. We introduce a mixture copula within the framework of worst case robust optimization. The resulting uncertainty set comprising of the mixture copulas dispense away from pre-selecting the probability distributions of returns.

The cynosure of this thesis is the applications of Topological Data Analysis (TDA) in the financial decision-making process. TDA combines concepts from algebraic topological to extract geometry and shapes in the data. TDA has set its footprints in several fields primarily in the medical and fraud detection domains, but it is not yet fully utilized to its potential in investment decisions including portfolio optimization. In this thesis, we endeavor to decipher an application of TDA in asset allocation viz. enhanced index tracking. The superior empirical performance of the proposed model strengthens our belief that TDA holds promise in the domains of investment science, risk management and econometric.

The core of this thesis lies in improvising reward-risk ratios of the portfolio's out-of-sample performance by unwrapping the underline dependency structure and the topological geometry shrouded in the financial data, thus bestowing the optimization models with the more realistic arrangements.

Besides, adding to the current knowledge of the copula theory and its applications in the

investment problems, we expect this thesis to open the doors for more research on TDA in the field of risk management and portfolio optimization.

सार

वित्तीय पोर्टफोलियो अनुकूलन और निवेश निर्णय लेना, जो एक पोर्टफोलियो बनाने वाली वित्तीय परिसंपत्तियों को भार आवंटित करने का पालन करते हैं, बड़े पैमाने पर शोध वाले क्षेत्र हैं। बैंकों, वित्तीय संस्थानों, फंड प्रबंधन फर्मों, निवेशकों को यह सरकार, कॉर्पोरेट या व्यक्ति होना चाहिए, सभी को अपनी संपत्ति का प्रबंधन करने और अपने पोर्टफोलियो आवंटन का चयन, संतुलन, और लगातार मूल्यांकन करके अपने जोखिम को हेज करने की आवश्यकता होती है। 1952 में मार्कोविट द्वारा सफलता लेख, जिसने विश्व को आधुनिक पोर्टफोलियो सिद्धांत से परिचित कराया, रिटर्न के बहुभिन्नरूपी सामान्य वितरण की धारणा के कारण देर से आलोचना का केंद्र रहा, जो वास्तविक डेटा डेटा में सच्चाई से बहुत दूर है। ।

बड़े डेटा के अनुप्रयोगों के कारण उच्च आयाम में जटिल निर्भरता संरचना ने वित्तीय चर के बीच बहुभिन्नरूपी निर्भरता को मॉडल करना आवश्यक और चुनौतीपूर्ण बना दिया है। प्रौद्योगिकी, खुली अर्थव्यवस्था और वैश्विक व्यापारों में प्रगति के कारण वित्तीय बाजारों में निर्भरता पिछले कई वर्षों में बढ़ी है। दुनिया के विभिन्न बाजारों में निवेश करने वाले निवेशकों के साथ वित्तीय बाजारों की सीमाओं के विस्तार के कारण, सिस्टम की जटिलता में और अधिक वृद्धि होती है। संपत्तियों के रिटर्न का प्रतिनिधित्व करने वाले यादृच्छिक चर आमतौर पर असममित होते हैं, जिनमें अनाम चक्रीय व्यवहार और शैलीगत विशेषताएं, उच्च कर्टोसिस, पूंछ पर निर्भरता, और उल्टा-उल्टा गति के रुझान होते हैं। इसके अलावा, दुनिया भर की अर्थव्यवस्थाओं ने अनुचित जोखिम प्रबंधन के परिणामस्वरूप वित्तीय आंकड़ों में कई महत्वपूर्ण संकटों, आर्थिक असंतुलन और अत्यधिक अस्थिर उतार-चढ़ाव को देखा है। एक नतीजे के रूप में, वित्तीय जोखिम प्रबंधन के तरीकों को 2008 के उप-प्रधान वैश्विक वित्तीय संकट के बाद गहराई से जांच के लिए रखा गया था। नए मॉडल के साथ आना उन निवेशकों के लिए एक आवश्यक और कठिन कार्य बन गया है जो जोखिम को कम करने से संबंधित हैं। उनके निवेश। निवेश के निर्णय लेने और जोखिम प्रबंधन में कई मॉडल सामने रखे गए हैं, जो एक ही समय में मूल मार्कोविटज़ पोर्टफोलियो ऑप्टिमाइज़ेशन मॉडल को संशोधित करने के जोखिम को कम करने के लिए अधिक यथार्थवादी सुविधाओं में लाकर जटिल निर्भरता संरचनाओं को पकड़ने के लिए आगे बढ़ाते हैं। वित्तीय परिसंपत्तियों के आधार पर।

इस थीसिस का उद्देश्य वित्तीय निवेशों और जोखिम प्रबंधन में एक नया दृष्टिकोण स्थापित करना है जो वित्तीय आंकड़ों में निर्भरता संरचना और आकृतियों को गहरा कर सकता है। उद्देश्य दो दृष्टिकोणों का उपयोग करके किया जाता है: कोपुला सिद्धांत और सामयिक डेटा विश्लेषण।

वित्तीय चर के बीच रेखांकन संरचना को मॉडल करने के लिए कॉपल्स का उपयोग किया गया है। स्केलर द्वारा सामने रखी गई थ्योरी के अनुसार, कोपल्स ऐसे फ़ंक्शंस हैं जो कपल्स मल्टीवेरिएट डिस्ट्रीब्यूशन को इसके वन-डायमेंशनल मार्जिन्स में जोड़ते हैं और इस तरह मार्जिन और रैंडम वेरिएबल्स के जॉइंट डिस्ट्रीब्यूशन के बीच रिलेशनशिप की जांच करते हैं। अधिक विशेष रूप से, इस थीसिस में, हम गतिशील बेल कॉपल्स के साथ काम करते हैं, उनके बिल्डिंग ब्लॉक्स के रूप में बाइवेरिएट कॉपुलस के साथ। एक बेल

कॉपुला वास्तविक डेटा के बीच प्रचलित विभिन्न वितरणों के मॉडलिंग के लिए एक लचीला तंत्र प्रदान करता है। उन्हें सभी संभावित और जटिल गैर-निर्भरता संरचना पर आक्रमण करने के लिए मास्टर बनाया गया है। हम अपने अध्ययन को गतिशील और डेटा ट्वीक के प्रति मजबूत बनाने के लिए रोलिंग विंडो योजना के साथ नियमित दाखलताओं को लागू करते हैं।

हम समय श्रृंखला मॉडलिंग का उपयोग करते हुए एक गतिशील कॉपुला आधारित परिदृश्य पीढ़ी पद्धति को डिजाइन करते हैं। समय श्रृंखला मॉडल में अवशिष्ट, परिसंपत्तियों के रिटर्न डेटा श्रृंखला में फिट किए गए, उलटे प्रायिकता परिवर्तन और बेल कॉपुला का उपयोग करके छद्म यादृच्छिक रूप से उत्पन्न होते हैं। नकली अवशेषों का उपयोग परिसंपत्तियों के रिटर्न के लिए परिदृश्य उत्पन्न करने के लिए किया जाता है। प्रस्तावित परिदृश्य (डेटा) पीढ़ी प्रक्रिया को इष्टतम पोर्टफोलियो में जोखिम-इनाम अनुपात को बढ़ाकर निवेश प्रक्रिया को संशोधित करने के लिए दिखाया गया है। यह उपन्यास पद्धति विभिन्न परिदृश्यों से डेटा उत्पन्न करने के लिए थीसिस के लिए लागू की जाती है जो बाजारों में प्रचलित हो सकती है और बेल कोपला द्वारा कब्जा कर लिया जा सकता है।

निर्भरता संरचना के मॉडलिंग के अलावा, यह थीसिस पोर्टफोलियो ऑप्टिमाइज़ेशन के डोमेन में इंडेक्स ट्रेकिंग और एन्हांस्ड इंडेक्सिंग की दो महत्वपूर्ण समस्याओं की जांच करता है। हमारे शोध की मुख्य विशेषताओं में से एक विचलन आधारित जोखिम उपायों का विकास है जो लंबे समय से शुरू किए गए थे, लेकिन तब से छिपे हुए थे। हम पोर्टफोलियो अनुकूलन में विचलन आधारित दूसरे क्रम स्टोचैस्टिक प्रभुत्व मानदंड का प्रस्ताव करते हैं और इसे बढ़ाया अनुक्रमण की समस्या पर लागू करते हैं। अन्य मुख्य भाग जोखिम में मिश्रित सशर्त मूल्य के साथ विचलन उपाय का प्रस्ताव है, जो सशर्त मूल्य के सभी उपयोगी गुणों को जोखिम में डालते हैं (CVaR)। हम सूचकांक ट्रेकिंग समस्या के लिए एक नया मॉडल विकसित करने के लिए विचलन उपाय के साथ दो-पूंछ मिश्रित सीवीआरआर भी तैयार करते हैं।

इसके अलावा, एक ही काम में, हम बढ़ाया अनुक्रमण समस्या के लिए एक दो कदम प्रक्रिया डिजाइन करते हैं। सबसे पहले, हम एक मार्कोव श्रृंखला आधारित निस्पंदन क्रिट पेश करते हैं

Table of Contents

Certificate	i
Acknowledgements	iii
Abstract	v
List of Figures	xii
List of Tables	xv
List of Abbreviations	xix
1 Introduction and Literature Survey	1
1.1 Urge of copulas and Topological Data Analysis	3
1.2 Applications of Copulas in Finance	6
1.2.1 Risk Management	6
1.2.2 Portfolio Decision Problems	7
1.2.3 Tail Risk Measures	8
1.2.4 Deviation Risk Measures	10
1.3 Stochastic Dominance	11
1.4 Robust Optimization	13
1.5 Index Tracking and Enhanced Indexing	15
1.6 Contagion	20
1.7 Topological Data Analysis (TDA)	23
1.8 Performance Measures	25
1.9 Organization of the Thesis	27
2 The Trinity of Copula Theory, Time Series and Topological Data Analysis	31
2.1 Motivation	31
2.2 Copula Theory: Definitions and Basic Properties	32
2.2.1 Dependence Measures	34
2.2.2 Bivariate Copulas	35
2.3 Pair Copula Construction (PCC)-Vines	39
2.3.1 Estimation of Copula Parameters	43
2.4 Financial Time Series	45
2.4.1 General Representation of a Time Series	47
2.4.2 Statistical Tests	48
2.5 Topological Data Analysis	50
2.5.1 Simplicial Complex and Homology	51
2.5.2 Vietoris-Rips Complex and Persistence Diagrams	51
2.5.3 Persistence Landscape	53
2.5.4 Taken's Embedding	54
3 Scenario Generation Technique Assimilating Copula and Time Series	57

3.1	Empirical Modeling	58
4	Stochastic Autoregressive Canonical Vine Model for Analyzing Contagion Effects	65
4.1	Motivation and Objective	65
4.2	SCAR C-Vine Model	66
4.3	Methodology	68
4.4	Data and their Descriptive Statistics	69
4.5	Evidences of Contagion Effect	71
4.5.1	Parameter Estimates of Marginal Models	71
4.5.2	Parameters Estimate of the Copula Models	77
4.6	Quantitative Measurement of Contagion Effect	80
4.6.1	Analysis of Contagion Effect using Kendall's Tau	80
4.6.2	Time Varying Kendall's Tau	86
4.7	Conclusions	89
5	Deviation Measure in Second Order Stochastic Dominance with an Application to Enhanced Indexing	91
5.1	Motivation and Objective	91
5.2	Second Order Stochastic Dominance	92
5.2.1	LSSD and TSSD Optimization Models	93
5.2.2	The Proposed Deviation SSD Optimization Model	95
5.3	Empirical Analysis on Historical Data	98
5.3.1	Historical Data	98
5.3.2	Methodology	99
5.4	Empirical Analysis using Simulation	104
5.5	Conclusions	108
6	Index Tracking and Enhanced Indexing using Mixed Conditional Value-at-Risk and STARR Ratio	109
6.1	Motivation and Objective	109
6.2	Mixed CVaR and Two Tail CVaR	110
6.2.1	CVaR Deviation Measure	112
6.3	Proposed Models for IT and EI	113
6.3.1	Two-Tail Mixed CVaR Model for IT	113
6.3.2	A Two-Step Procedure for EI	114
6.4	Empirical Analysis and Performance Evaluation	119
6.4.1	Sample Data	119
6.4.2	Methodology	120
6.5	Results	121
6.5.1	IT model	121
6.5.2	Results for the EI models	124
6.6	Conclusions	125
7	Robust Optimization of STARR Ratios using Copulas	129
7.1	Motivation and Objective	129
7.2	Robust STARR Ratio Model	130

7.2.1	General structure of Robust optimization model	130
7.2.2	Robust STARR ratio model with MCVaR and Deviation MCVaR as Risk Measure	131
7.2.3	The Data	137
7.3	Empirical Analysis	137
7.3.1	Results: First dataset	138
7.3.2	Results: Second dataset	141
7.3.3	Third dataset: Rolling Window Analysis	141
7.4	Conclusions	146
8	Topological Data Analysis in Investment Decisions	147
8.1	Motivation and Objective	147
8.2	Data and Methodology	148
8.2.1	Sample Data	148
8.2.2	Methodology	149
8.3	Significance of TDA Norm Value	152
8.4	Application of TDA to Enhanced Indexing	153
8.4.1	A Two-Step Procedure	153
8.4.2	Empirical Analysis and Results	155
8.4.3	Empirical Analysis and Results	157
8.5	Conclusions	159
9	Contributions, Limitations and Future Outlook	169
9.1	Contributions of the Thesis	169
9.2	Limitations and Future Outlook	171
	Bibliography	173
	Curriculum Vitae	195

List of Figures

2.1	Dependence modeling with Gaussian copula: (a) $\theta = -0.8$, (b) $\theta = 0.8$. . .	36
2.2	Dependence modeling with Student's- t copula: (a) $(\theta, \nu) = (-0.8, 8)$, (b) $(\theta, \nu) = (0.8, 8)$	36
2.3	Dependence modeling with Frank copula: (a) $\theta = -10$, (b) $\theta = 10$	37
2.4	Dependence modeling with Clayton copula: (a) $\theta = 5$, (b) $\theta = 10$	38
2.5	Dependence modeling with Gumbel copula: (a) $\theta = 5$, (b) $\theta = 10$	38
2.6	Dependence modeling with Joe copula: (a) $\theta = 5$, (b) $\theta = 10$	39
2.7	C-Vine trees	42
2.8	D-Vine trees	42
2.9	Rips Filtration	52
2.10	Persistence Diagram and Persistence Landscape	54
3.1	Rolling window scheme	63
4.1	Returns plots of Russian flu period	71
4.2	Returns plots of the subprime crisis period	72
4.3	Returns plots of China slowdown period	73
4.4	Kendall's tau values for the entire period of the Russian flu between Russia and (reading top to bottom) India, the USA, Hong Kong, Japan, China . .	81
4.5	Kendall's tau values for the entire period of the subprime crisis between the USA and (reading top to bottom) India, Russia, Hong Kong, Japan, China	82
4.6	Kendall's tau values for entire period of China slowdown between China and (reading top to bottom) India, Russia, the USA, Hong Kong, Japan .	82
4.7	Kendall's tau values (Tree 2) for the entire period of the Russian flu (Tree 2) with the USA as root node conditioned on Russia	84
4.8	Kendall's tau values (Tree 2) for the entire period of the subprime crisis with India as root node conditioned on the USA	85
4.9	Kendall's tau values (Tree 2) for the entire period of the China slowdown with India as root node conditioned on China	85
4.10	Rate of change in Kendall's tau post the Russian flu	86
4.11	Rate of change in Kendall's tau post subprime crisis	87
4.12	Rate of change in Kendall's tau post China slowdown	87
4.13	Rate of change of Kendall's tau (Tree 2) post Russian flu (conditioned on Russia)	88
4.14	Rate of change of Kendall's tau (Tree 2) post the subprime crisis (conditioned on USA)	88
4.15	Rate of change of Kendall's tau (Tree 2) post China slowdown (conditioned on China)	89
5.1	In-sample period daily return of each index in dataset A	104
5.2	In-sample period daily return of indices in dataset B	105
5.3	The scatter plot of residuals series of indices comprising data set A.	107

6.1	Out-of-sample cumulative returns of optimal portfolios from IT Models . .	123
6.2	Out-of-sample cumulative returns of optimal portfolios from EI Models . .	127
8.1	Daily returns, standard deviation and TDA norms of S&P Global Index .	152
8.2	Daily returns and TDA norms of a random constituent of S&P Global Index	152
8.3	Daily returns and TDA norms of Athens composite	153
8.4	Out-of-sample cumulative returns of optimal portfolios from EI Models . .	158

List of Tables

4.1	Copulas with respective parameter range and corresponding Kendall's tau	68
4.2	Descriptive statistics of log returns (Russian flu): Sep 1, 1995 to Aug 15, 2000. Here AC, BC and WP refers to After crisis, Before crisis and Whole period respectively. Q(12) denotes the Ljung Box Q test of serial correlation up to 12 lags in the returns.	74
4.3	Descriptive statistics of log returns (the subprime crisis): Jan 1, 2005 to Dec 31, 2010	74
4.4	Descriptive statistics of log returns (China slowdown): Jan 1, 2015 to Dec 1, 2015	75
4.5	Parameters estimate for the marginal distribution model (before the Russian flu). The values in the parentheses denote the standard errors.	75
4.6	Parameters estimate for the marginal distribution model (after the Russian flu)	76
4.7	Parameters estimate for the marginal distribution model (before the subprime crisis)	76
4.8	Parameters estimate for the marginal distribution model (after the subprime crisis)	77
4.9	Parameters estimate for the marginal distribution model (before China slowdown)	77
4.10	Parameters estimate for the marginal distribution model (after China slowdown)	78
4.11	Parameters estimate for the copula models (Russian flu period)	78
4.12	Parameters estimate for the copula models (the subprime crisis)	78
4.13	Parameters estimate for the copula models (China slowdown period)	79
4.14	Parameters estimate for the copula models (Tree 2) with the USA as root node conditioned on Russia (Russian flu period)	79
4.15	Parameters estimate for the copula models (Tree 2) with India as root node conditioned on the USA (subprime crisis period)	79
4.16	Parameters estimate for the copula models (Tree 2) with India as root node conditioned on China (China crisis period)	80
4.17	Parameters estimate for the last stage of the C-Vine model	80
5.1	Total number of rolling windows in the in-sample periods in which allocation to a single stock in LSSDP, DSSDP, TSSDP falls in the intervals $I_1 = [70\%, 80\%)$, $I_2 = [80\%, 90\%)$, $I_3 = [90\%, 99\%)$, $I_4 = [99\%, 100\%]$. The analysis is reported on 5 indices, Track 1 (Hang Seng), Dow Jones, FF 49, BSE Sensex and CNX Nifty, comprising of 50 or fewer stocks. The total number of rolling windows is quoted in bracket in first column alongside each index.	100

5.2	Performance statistics of the out-of-sample returns from DSSDP and LSSDP on the historical datasets of 16 global indices, where the values are quoted by Mean = Mean $\times 10^{-3}$, Variance = Variance $\times 10^{-3}$, and Median = Median $\times 10^{-3}$. The ***, **, *, are used to mark the significant values in the statistical tests for mean and variance at the significance levels $0 < \alpha \leq 0.01$, $0.01 < \alpha \leq 0.05$, $0.05 < \alpha \leq 0.1$, respectively.	101
5.3	Out-of-sample performance of optimal portfolios DSSDP, LSSDP and TSSDP.	102
5.4	Summary statistics for indices in dataset A in the in-sample period along with results of hypothesis tests. Also, read values of Mean = Mean $\times 10^{-4}$. The value 1 indicate rejection of the null hypothesis for a given statistical test, otherwise 0 is used.	105
5.5	The BIC values for different ARMA models on in-sample returns data of 6 indices from the real dataset A.	106
5.6	Estimated values of the parameters for best fit ARMA(1, 1)-GJR-GARCH(1, 1) model in equations (3.1) - (3.2), where read the values of $\mu = \mu \times 10^{-4}$, $\omega = \omega \times 10^{-4}$ and $\alpha_1 = \alpha_1 \times 10^{-4}$	106
5.7	The Pearson's correlation matrix of 6 residuals series obtained from ARMA(0, 1)-GJR-GARCH(1, 1) model applied on real data of the in-sample period of each index comprising dataset A.	106
5.8	Out-of-sample performance analysis of three models where the models are solved using simulated data for two datasets A and B. Read values of Mean = Mean $\times 10^{-3}$, Std dev = Std dev $\times 10^{-2}$, Min = Min $\times 10^{-2}$, Max = Max $\times 10^{-2}$, EMR = EMR $\times 10^{-4}$, Sortino ratio = Sortino ratio $\times 10^{-2}$, DD = DD $\times 10^{-3}$, $CVaR_{0.95} = CVaR_{0.95} \times 10^{-2}$, $CVaR_{0.97} = CVaR_{0.97} \times 10^{-2}$, $SR_{0.95} = SR_{0.95} \times 10^{-2}$, $SR_{0.97} = SR_{0.97} \times 10^{-2}$	108
6.1	The out-of-sample performance of the portfolios obtained from three IT models on the correlation between portfolios and the indices, tracking error, maximum loss with respect to the index, information ratio and turnover ratio; the best values amongst the three models are highlighted in bold	122
6.2	The out-of-sample performance of optimal portfolios from six EI models, on seven performance measures including EMR, Sortino ratio, Omega excess ratio, Rachev ratio and VaR ratio. Each value v (except for the p -value) is to be read as $v \times 10^{-3}$. The best and the worst values amongst all models are highlighted in bold and <i>italic</i> , respectively. The significance level $\alpha \leq 0.01$, $0.01 < \alpha \leq 0.05$, $0.05 < \alpha \leq 0.1$, $0.1 < \alpha \leq 0.12$ are displayed by ****, ***, **, *, respectively.	126
7.1	Summary statistics of six indices along with the outcomes of three statistical tests over the in-sample period; the value 1 indicates rejection of the null hypothesis for a given statistical test.	138
7.2	Bayes information criterion (BIC) values of ARMA(p, q) model for different choices of parameter pair (p, q) on the return series of six indices over the in-sample period.	139

7.3	The estimated parameters for the best fit ARMA(0, 1)-GJR-GARCH(1, 1) model calculated over the in-sample data where the residuals follow skewed student's t distribution with parameters ζ and ν , the skewness and shape parameters, respectively.	139
7.4	The Pearson's correlation coefficients and the outcomes of KS and LB tests obtained from the in-sample data; the value 0 indicates rejection of the alternative hypothesis for a given statistical test, while 1 is used otherwise.	139
7.5	The out-of-sample values of the performance measures of four portfolios ($WS_C(X)$), ($WS_C^\Delta(X)$), ($S_G(X)$) and ($S_G^\Delta(X)$) in the out-of-sample period Sept 2008 - April 2009 of the first dataset.	140
7.6	The out-of-sample values of the performance measures of four portfolios ($WS_C(X)$), ($WS_C^\Delta(X)$), ($S_G(X)$) and ($S_G^\Delta(X)$) on the data of PIIGS countries in the out-of-sample period April 2003 - August 2017.	141
7.7	The in-sample and out-of-sample periods of analysis (in days) along with total number of out-of-sample rolling windows.	142
7.8	The out-of-sample performance of optimal portfolios from all the models with $d_1 = 1026$ days in-sample and $d_2 = 504$ days out-of-sample period in each window. The best values amongst all models are highlighted in bold	143
7.9	The out-of-sample performance of optimal portfolios from all the models with $d_1 = 1026$ days in-sample and $d_2 = 504$ days out-of-sample period in each window. The best values amongst all models are highlighted in bold	144
7.10	The average out-of-sample performance analysis of the returns data series obtained by serially concatenating the out-of-sample returns series from the out-of-sample windows of the optimal portfolios of the corresponding model for all the in-sample and out-of-sample periods considered. The best values among all models, for each in-sample and out-of-sample period, are highlighted in bold	145
8.1	Out-of-sample rolling windows.	149
8.2	Out-of-sample performance analysis of the filtered portfolios using the concatenated series with 6 months and 3 months in-sample and out-of-sample periods, respectively. In case the mean return is negative, the values of Sharpe ratio and Sortino ratios are taken to be zero.	160
8.3	Out-of-sample performance analysis of the filtered portfolios using the concatenated series with 6 months and 2 months In-sample and Out-of-sample periods respectively	161
8.4	Out-of-sample performance analysis of the filtered portfolios using the concatenated series with 6 months and 1 month In-sample and Out-of-sample periods respectively	162
8.5	Out-of-sample performance analysis of the filtered portfolios using the concatenated series with 3 months and 1 month In-sample and Out-of-sample periods respectively	163
8.6	Out-of-sample performance analysis of the proposed EI model using the concatenated series	164
8.7	Out-of-sample performance analysis of the proposed EI model using the concatenated series	165

8.8	Out-of-sample performance analysis of the proposed EI model using the concatenated series	166
8.9	Out-of-sample performance analysis of the proposed EI model using the concatenated series	167

List of Abbreviations

$CVaR_\alpha$	Conditional Value-at-Risk at α degree of confidence level
EI	Enhanced Indexing
EMR	Excess Mean Return
IR	Information ratio
IT	Index Tracking
LP	Linear Program
MAD	Mean Absolute Deviation
MCVaR	Mixed Conditional Value-at-Risk
PO	Portfolio Optimization
RO	Robust Optimization
RR	Rachev Ratio
SD	Stochastic Dominance
SHR	Sharpe Ratio
SSD	Second order Stochastic Dominance
STARR	The Stable Tail-adjusted Return Ratio
TDA	Topological Data Analysis
TYR	Treynor Ratio
VaR_α	Value-at-Risk at α degree of confidence level