



UNIL | Université de Lausanne

Unicentre
CH-1015 Lausanne
<http://serval.unil.ch>

Year: 2024

Bears, Bulls and Monkeys

Bachelard Cyril

Bachelard Cyril, 2024, Bears, Bulls and Monkeys

Originally published at : Thesis, University of Lausanne
Posted at the University of Lausanne Open Archive <http://serval.unil.ch>
Document URN : urn:nbn:ch:serval-BIB_B27F2BC035E06

Droits d'auteur

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES
DÉPARTEMENT DES OPÉRATIONS

Bears, Bulls and Monkeys

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de

Doctorat en Business Analytics

par

Cyril BACHELARD

Directeur de thèse
Prof. Olivier Gallay

Co-directeur de thèse
Prof. Eric Jondeau

Jury
Prof. Paul André, président
Prof. Lorenzo Bretscher, expert interne
Prof. Erich Walter Farkas, expert externe

LAUSANNE
2024



UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES
DÉPARTEMENT DES OPÉRATIONS

Bears, Bulls and Monkeys

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de

Doctorat en Business Analytics

par

Cyril BACHELARD

Directeur de thèse
Prof. Olivier Gallay

Co-directeur de thèse
Prof. Eric Jondeau

Jury
Prof. Paul André, président
Prof. Lorenzo Bretscher, expert interne
Prof. Erich Walter Farkas, expert externe

LAUSANNE
2024

IMPRIMATUR

La Faculté des hautes études commerciales de l'Université de Lausanne autorise l'impression de la thèse de doctorat rédigée par

Cyril BACHELARD

intitulée

Bears, Bulls and Monkeys

sans se prononcer sur les opinions exprimées dans cette thèse.

Lausanne, le 03.07.2024



Professeure Marianne Schmid Mast, Doyenne



Thesis Committee

Olivier Gallay, Professor.

Faculty of Business and Economics, University of Lausanne.

Thesis supervisor.

Eric Jondeau, Professor.

Faculty of Business and Economics, University of Lausanne.

Thesis co-supervisor.

Lorenzo Bretscher, Assistant Professor.

Faculty of Business and Economics, University of Lausanne.

Internal expert.

Erich Walter Farkas, Professor.

Department of Finance, University of Zurich.

External expert.

Paul André, Professor and Vice Dean.

Faculty of Business and Economics, University of Lausanne.

President.

University of Lausanne
Faculty of Business and Economics

PhD in Business Analytics

I hereby certify that I have examined the doctoral thesis of

Cyril BACHELARD

and have found it to meet the requirements for a doctoral thesis.

All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: 10.06.2024

Prof. Olivier GALLAY
Thesis supervisor

University of Lausanne
Faculty of Business and Economics


PhD in Business Analytics

I hereby certify that I have examined the doctoral thesis of

Cyril BACHELARD

and have found it to meet the requirements for a doctoral thesis.

All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: _____  _____ Date: 10.06.2024

Prof. Eric JONDEAU
Thesis co-supervisor

University of Lausanne
Faculty of Business and Economics

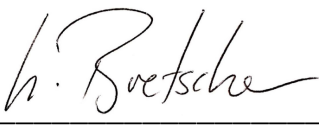
PhD in Business Analytics

I hereby certify that I have examined the doctoral thesis of

Cyril BACHELARD

and have found it to meet the requirements for a doctoral thesis.

All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: 10.06.2024

Prof. Lorenzo BRETSCHER
Internal expert

University of Lausanne
Faculty of Business and Economics

PhD in Business Analytics

I hereby certify that I have examined the doctoral thesis of

Cyril BACHELARD

and have found it to meet the requirements for a doctoral thesis.

All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: E.W. Farkas Date: 10.06.2024

Prof. Erich Walter FARKAS
External expert

“Financial analysts in pin-striped suits do not like being compared with bare-assed apes. They retort that academics are so immersed in equations and Greek symbols (to say nothing of stuffy prose) that they couldn’t tell a bull from a bear, even in a China shop.”

Burton G. Malkiel

Preface

I must apologize for any disappointment, but this dissertation is not about animals, regrettably. It is about stock markets and statistics, mostly. It tells about chance and its dual role as both protagonist and antagonist. It is a testament to the interdisciplinary nature of modern finance, where insights from economics, mathematics, statistics, and computer science converge to illuminate the workings of global markets. To spice up what might be a dry topic for some, we will be sprinkling in a pinch of geometry for good measure.

I hold hope that the discoveries uncovered within these pages will enrich the ongoing dialogue between academia and industry, fostering innovation and resilience. This, in light of statements such as those articulated by Mr. Malkiel, to whom I respectfully offer an alternative perspective.

Acknowledgments

I want to express my sincere appreciation to my supervisor, Olivier Gallay, and my co-supervisor, Eric Jondeau, for their invaluable guidance, unwavering support, and insightful feedback throughout this thesis journey.

I am deeply grateful to my co-authors, Apostolos Chalkis, Elias Tsigaridas, and Vissarion Fisikopoulos, for their exceptional collaboration. Their inspiring discussions, constructive criticism, and significant contribution to the initiation of GeomScale, the open-source software platform upon which our work is based, have been invaluable.

I also want to thank my partners at QuantArea AG, as well as my former employer, OLZ AG, for their encouragement and support. Their recognition of the importance of ongoing education and professional development has allowed me to pursue this research while managing my work responsibilities.

Further, my heartfelt thanks go to my parents and to my wife for their boundless love and unwavering support.

Lastly, I extend my gratitude to all individuals who have directly or indirectly contributed to this thesis. Special mentions include Orif Ibrogimov, Anna Timonina-Farkas, and Soros Chitsiripanich, among others. If I have inadvertently omitted anyone, I offer my apologies and extend my thanks to them as well.

Cyril Bachelard

Zurich, April 2024

Contents

1	Introduction	1
	References	10
2	Randomized Geometric Tools for Anomaly Detection in Stock Markets	11
2.1	Introduction	12
2.1.1	Financial background	13
2.1.2	Connection to geometry, methodology, and computational challenges	16
2.1.3	Contributions	20
2.2	Geometric modeling and algorithms	21
2.2.1	Geometric modeling	21
2.2.2	Sampling from a connected component of \mathcal{K}	23
	Great Cycle Walk (GCW)	24
	Reflective Great Cycle Walk (ReGCW)	25
2.2.3	Practical volume approximation	27
2.3	Implementation and experiments	30
2.3.1	Implementation	31
	Parameter tuning for practical performance	31
2.3.2	Construction of volatility-constrained random portfolios and back- testing framework	32
2.3.3	Data	33
2.4	Results	34
2.5	Discussion	44
	References	46
	Appendix	53
2A	Membership oracle for a connected component	53

2B	Sampling uniformly with Great Cycle Walk	53
2C	Proofs	54
	Theorem 1	54
	Theorem 3	55
2D	Tables	57
2E	Figures	60

3	Randomized Control in Performance Analysis and Empirical Asset Pricing	63
3.1	Introduction	64
	3.1.1 Contributions	68
	3.1.2 Previous work	68
3.2	Use and Abuse of Random Portfolios	71
	3.2.1 A (brief) review of performance analysis	71
	3.2.2 RP-based performance evaluation - An example	74
3.3	Random Portfolio Generation	82
	3.3.1 Naive Random Portfolios	82
	3.3.2 Basic Random Portfolios	85
	3.3.3 Regularized Random Portfolios	85
	3.3.4 Sampling	87
3.4	Geometric Random Walks	89
	3.4.1 Ball walk (BaW)	101
	3.4.2 Hit-and-Run (HaR)	104
	3.4.3 Coordinate Directions Hit-and-Run (CDHR)	106
	3.4.4 Billiard walk (BiW)	108
	3.4.5 Dikin walk	110
	3.4.6 Vaidya walk	112
	3.4.7 John walk	114
	3.4.8 Hamiltonian Monte Carlo (HMC)	115
	3.4.9 Reflective NUTS HMC (ReNHMC)	117
	3.4.10 Constrained Riemannian HMC (CoRHMC)	117
3.5	Empirical Study	119

3.5.1	Data	121
3.5.2	Implementation	123
3.5.3	Results	124
3.6	Conclusion	128
	References	130
	Appendix	142
3A	Factor mapping	142
4	On The Geometry of the Bootstrap	143
4.1	Introduction	144
4.2	Geometry of the bootstrap and an exact algorithm	146
4.2.1	Computing the exact Bayesian bootstrap distribution	149
4.2.2	Numerical example	151
4.2.3	Concentration phenomena	152
	Partitions and node activation probabilities	152
4.3	Exact moments	156
4.3.1	Exact standard error	157
4.3.2	Exact higher central moments	161
4.3.3	The moment problem	163
4.3.4	Numerical example revisited	164
4.4	Informative priors	164
4.4.1	Asymmetry	165
4.4.2	Concentration	167
4.4.3	Constrained domain	169
4.4.4	Mixtures	171
4.5	Applications	172
4.5.1	Regression	173
4.5.2	Eigenvalue Estimation	175
4.5.3	Nonparametric Density Estimation	176
4.6	Conclusion and further discussion	177
	References	180
	Appendix	184

4A	Algorithm of Varsi (1973)	184
4B	Algorithm of Cales et al. (2021)	184
4C	Proof of proposition 14 under $\omega \sim D_\alpha$	185
4D	Closed-form expressions	186
5	Good and Bad Turbulence	189
5.1	Introduction	190
5.2	Financial turbulence: Measures, properties and hedging rules	193
5.2.1	Properties	196
5.3	Discriminating good from bad turbulence	199
5.3.1	Turbulence and (quadratic) discriminant analysis	200
5.3.2	Bears and bulls	204
5.3.3	Smoothing	206
5.3.4	Refinements	209
5.4	Empirical study	210
5.4.1	Data and backtesting framework	211
5.4.2	Descriptive analysis	214
5.4.3	Backtest results	216
5.4.4	Simulation study	219
5.5	Conclusion	222
	References	224
	Appendix	227
5A	Drawdown Statistics	227
6	Conclusion	229

2 my 3 girls

Chapter 1

Introduction

The advent of computer simulation methods, such as sampling-based techniques like Markov Chain Monte Carlo (MCMC), has marked a milestone in scientific exploration. These methods have substantially enhanced our ability to investigate complex systems, make accurate predictions, and conduct rigorous statistical analyses. Their applications span various fields, from physics and climate science to biology and economics, empowering researchers to confront once-insurmountable challenges.

As scientific challenges grow in complexity, demanding precise modelling of real-world intricacies characterized by non-linear behaviour, high-dimensionality, temporal variations, and constraints, it becomes imperative for our computational tools to evolve accordingly. This dissertation seeks to address this need. It comprises four self-contained chapters focused on computational tools, particularly sampling-based methods. These methods are examined both as tools for addressing inquiries in finance, economics, or statistics, and as subjects of study in their own right.

Specifically, the work examines geometric random walk algorithms, a powerful class of MCMC procedures designed for high-dimensional constrained scenarios. Originating from computational geometry and previously untapped in finance, these methods are thoroughly surveyed, encompassing novel complexity results, and expanded to address specific inquiries in empirical asset pricing and performance evaluation.

The overarching theme across all chapters is a geometric perspective, which is relevant to both the modeling of economic questions and the analysis of algorithmic sampling procedures. This perspective yields analytical insights into one of the most renowned sampling-based methods for statistical inference - the bootstrap.

The core of this work lies in the practical applications, particularly in the area of financial economics, to better understand the characteristics of certain investment strategies and ultimately to refine them. Despite this practical emphasis, a thorough grasp of the underlying theoretical concepts forms an integral part of this study. This involves delving into complexity results of sampling algorithms, rooted in profound mathematical principles, as well as exploring economic relationships that demand a sound interpretation of the underlying economic theory. In the realm of implementing quantitative trading strategies, empirical observations are backed by theoretical arguments to enhance confidence in the proposed system's ability to generalize. However, practical considerations sometimes take precedence, guided by real-world experience. This fusion of theoretical and applied interdisciplinary dimensions leads to a comprehensive exploration of computational methodologies within the ambit of empirical finance.

Considerable efforts have been invested to implement the computational tools discussed in this work, in particular the sampling routines, within the open-source platform `GeomScale`¹, to which the author of this thesis contributes. `GeomScale` is associated with NumFOCUS, a nonprofit organization dedicated to promoting open source and reproducible scientific projects, alongside notable initiatives like NumPy, Julia, Jupyter, and Stan. Additionally, `GeomScale` is recognized as an open-source organization on the Google Summer of Code (GSoC)² project site.

`GeomScale` is a research and development project that provides open source code for state-of-the-art algorithms for geometric statistics, i.e., the intersection of data science, optimization, geometric and statistical computing. Its current focus is on scalable algorithms for sampling from high-dimensional distributions, integration, convex optimization and their applications. The core library, `volesti`, is written in C++, supported by interfaces for R and Python. Notably, `GeomScale`'s current implementation excels in handling problems spanning thousands of dimensions, making it a leading software package. Its applications span diverse fields including systems biology, finance, and artificial intelligence, where it has already made significant contributions.

The motivation for the thesis, which now encompasses various themes, arose from a practical challenge encountered while working in the financial industry as a quantitative

¹<https://geomscale.github.io/about/>

²GSoC is an online program aimed at engaging new contributors in open source software development, where the author of this thesis has served as a mentor for the past four years.

research analyst. The task was to demonstrate that the success of a particular investment fund could be attributed to the manager’s skill rather than mere chance (without using classical regression-based factor models). Initially, it seemed logical to formalize this inquiry as a statistical hypothesis test. To do so, one needed to define the null distribution representing the absence of skill, i.e., the set of all possible performance results that can be obtained without ability along with the corresponding probability measure. This could be achieved through a randomized control, i.e., a group of portfolios that by construction incorporate no strategy, bias or skill.

Indeed, the suggested experimental design outlined in the financial literature involves defining a random portfolio and deriving its performance distribution, either analytically or through sampling and simulation (see the literature review in Section 3.1.2 of Chapter 3). This distribution then serves as the null distribution against which to evaluate the performance of the investment strategy under scrutiny. Importantly, for the randomized control to constitute a fair representation, it should adhere to the investment guidelines of the investigated strategy. These guidelines, often stemming from regulatory requirements, are typically designed to restrict (risk) exposure to individual security issuers or groups of issuers. Moreover, the literature suggests that the random portfolio weights should follow a uniform distribution to ensure that any one portfolio configuration is just as likely to appear as any other.

In technical terms, let ω denote a random portfolio, characterized by a real-valued random vector with a probability density function (pdf) π and bounded support $\mathcal{K} \subset \mathbb{R}^n$. \mathcal{K} is determined by investor-imposed constraints, forming a convex body geometrically. As the weights are random, performance measures of the random portfolio are also (scalar) random variables. For instance, let $\theta = (\mu, \sigma)$ denote the vector of performance parameters, with mean $\mu = \langle \omega, r \rangle$, r representing the mean vector of asset returns, and variance $\sigma = \omega^\top \Sigma \omega$, where Σ is the covariance matrix of securities’ returns.

The challenge arises in finding the distribution of θ from the random weights with bounded support (assuming that r and Σ are known). While analytically intractable in general, the problem lends itself well to approximative numerical procedures such as MCMC. However, despite the common practice of using Monte Carlo methods in performance evaluation and the prevalence of MCMC in finance, there is no literature that uses MCMC to sample from a random portfolio. One of the main objectives of the current

work was originally to adapt geometric random walk-based sampling to close the existing gap.

The highly efficient implementation of geometrically inspired sampling methods within the GeomScale project has provided a solution to sampling from a constrained random portfolio, thereby addressing the performance evaluation problem. However, while the sampling issue has been resolved, a fundamental flaw has emerged regarding the very concept of formulating performance analysis as a statistical test using random portfolios. It has become apparent that such a formulation is ill-posed due to i) the lack of an unambiguous solution to the problem and ii) the high sensitivity of the solution to the initial conditions of the experimental design. This flaw is associated with the non-intuitive properties of high-dimensional spaces and is discussed in Chapter 3.

Coming to terms with this realization was a journey, not due to ego but because random sampling for performance evaluation, including the widely used bootstrap method, is prevalent in both the financial literature and practice (see the literature review in Section 3.1.2 of Chapter 3). Peer-reviewed articles such as Lisi (2011), which advocate for “A statistically correct procedure for evaluating management skill in the mutual fund industry” based on random portfolios, must be challenged and refuted.

In essence, the inappropriate use of a random portfolio’s performance distribution as the null distribution of no skill stems from a misconception that “The uniformly distributed random portfolio is analogous to an enumeration of all feasible portfolios without any prior on the market.” (Kim and Lee, 2016). The crux of the matter lies in the fact that in high dimensional spaces, the uniform measure heavily concentrates in the high-volume region of the distribution support. While the uniform measure guarantees that each portfolio has an equal probability of being sampled, the concentration phenomenon means that sets of portfolios with specific characteristics appear much more likely than others. For example, with overwhelming probability, a sampled portfolio would have weights whose entropy and L_p norm are close to constants in a given dimension. Consequently, a uniformly random portfolio resembles an enumeration of typical³ portfolios, all closely resembling the portfolio represented by the centroid of the sampling space defined by investment constraints. In the basic scenario where random portfolio weights must be non-negative

³Chapter 4 briefly touches upon a connection to the formal notion of typicality from information theory.

and sum to one, the centroid corresponds to the equally-weighted portfolio (also known as the naive or $1/N$ portfolio). Hence, the performance distribution of a random portfolio acts as a sampling distribution of the sample mean of the naive portfolio's performance measure (or as the posterior distribution of the mean of the naive portfolio's performance measure in Bayesian terms). Allocations such as those of capitalization-weighted indices or the elaborate investment strategy under evaluation, typically resulting from an optimization and located at the boundary of \mathcal{K} , are highly improbable to be sampled. As a result, statistical testing tends to exhibit a strong bias towards rejecting the null hypothesis of no skill.

What initially began as an effort to advance sampling-based performance evaluation morphed into a critique of such practices, as detailed in Chapter 3. Nonetheless, amidst empirical analyses, a somewhat related application for random portfolios emerged: factor analysis and anomaly detection.

Chapters 2 and 3 are devoted to these topics. Chapter 2 addresses the low-volatility anomaly and proposes novel randomized geometric tools to detect and analyze the empirical phenomenon. The low-volatility anomaly, which goes back to at least ?, refers to the empirical observation that stocks or portfolios with lower volatility (or more generally, risk) tend to provide higher returns than would be expected according to traditional economic theory like the Capital Asset Pricing Model (CAPM). This phenomenon challenges the basic tenets of modern finance, which suggest that higher risk should be compensated with higher returns. The anomaly suggests that investors can achieve superior risk-adjusted returns by investing in low-volatility assets.

The usual procedure to investigate the (existence of the) low-volatility anomaly is to form a small number of portfolios by first sorting stocks according to their historical volatility and grouping them into percentile portfolios with increasing volatility levels. Stocks within these portfolios are weighted equally or in proportion to their market capitalization. In the empirical asset pricing literature, a factor series is commonly generated from the percentile portfolios as a dollar-neutral portfolio which takes a long position in the first percentile portfolio and a short position in the last percentile portfolio. If this factor exhibits a positive return, meaning that the low-volatility portfolio outperforms the high-volatility portfolio, it is termed an anomaly.

The issue with this approach is its disregard for correlations and its reliance on the

chosen weighting scheme within the percentile portfolios. Chapter 2 proposes to replace the point estimates for the percentile portfolios with sets comprising all possible portfolios having a predefined volatility. This is done by defining multiple random portfolios, each having a pre-specified volatility, and then taking samples from those random portfolios and measuring their risk-return characteristics over time. Empirical results confirm the presence of the low-volatility effect in two regional universes and sub-markets discriminating for sector belongingness and company size. However, it is also shown that the effect is overestimated by sorting-based methods. In particular, it is demonstrated that the spread in realized volatility levels increases with higher volatility targets, indicating larger estimation errors for such portfolios. Overall, it is found that the classical sorting-based approach badly represents the set of possible portfolios having a certain volatility. Therefore, one should exercise caution when basing inference on them.

The article marks the first geometric formulation and MCMC-based analysis of the volatility puzzle in stock markets. The modeling of the (detection) problem results in sampling and estimating the (relative) volume of geodesically non-convex and nonconnected spherical patches that arise by intersecting a non-standard simplex with a sphere. To accomplish the task, two novel MCMC algorithms are introduced that exploit the geometry of the problem and employ state-of-the-art continuous geometric random walks adapted on spherical patches. Chapter 3 takes a step back and discusses the use of a random portfolio for performance evaluation. It explains in more detail that commonly employed random portfolios do not offer a statistically acceptable experimental design for performance comparisons and cautions the reader against such a use case with the aid of an illustrative example.

The article then moves on to suggest an alternative use case involving the investigation of the relationship between portfolio factor scores and performance in a constrained setup, which allows to empirically test for the presence of factor anomalies under stringent investment guidelines. Experiments on real-world datasets using the GeomScale software are conducted, replicating the investment guidelines of the MSCI Diversified Multi-Factor index⁴. The chosen index, representing a framework commonly encountered by institutional investors, aims to optimize factor exposures towards momentum, value, size, and quality, while adhering to risk, weight, and factor exposure constraints relative to the parent index.

⁴<https://www.msci.com/msci-diversified-multi-factor-indexes>

The study uncovers diverse impacts of the constraints imposed by the index on factor-exposure-performance relations compared to an otherwise unconstrained long-only framework. In summary, tilts towards value and momentum are found to offer higher returns with increased volatility, while quality tilts reduce risk without affecting returns, and size tilts lead to lower returns and higher risk. The findings suggest that factor patterns persist even within stringent constraints, justifying an active investment approach to capitalize on the positive ones.

Chapter 3 is rather demanding since it covers various facets. On the one hand, economic concepts of performance measurement and factor analysis are presented, while on the other hand, all existing geometric random walks are surveyed to sample random portfolio weights adhering to linear and quadratic constraints under various models, explaining their strengths, limitations, and computational complexities. For publication in an academic journal, the survey on geometric random walks and related novel results on computational complexity have been separated from the rest of the article. This has resulted in two papers targeting distinct audiences: one in finance and one in computational geometry.

In the process of categorising random portfolios on the basis of geometric properties, the well-known bootstrap method (Efron, 1979) is discussed. The adopted geometric perspective leads to insights into analytical solutions, to which a separate chapter is devoted. Chapter 4 examines nonparametric bootstrap techniques from a geometric perspective, based on a unified representation of classical and Bayesian frameworks using random weights. It establishes the connection between resampling plans and the weight distribution, and presents a direct computational method for the exact Bayesian bootstrap distribution of Rubin (1981) without the need for sampling. The study links exact moments of bootstrap distributions to the L_p norm of random weights and proposes ways to incorporate prior information into bootstrap procedures by adjusting the parametrization of the weight distribution. Eventually, it discusses applications where it is common to use bootstrap standard errors and shows how these can be given without resampling by leveraging the geometric perspective.

The last chapter of this thesis shifts focus from high-dimensional sampling methods to the development of a dynamic hedging strategy. While the previous chapters deal with sampling points, representing portfolios or probability vectors, Chapter 5 addresses the processing of a risk signal indicating moments for hedging investments in a stock index

over time.

Despite this shift, a thematic connection remains with the preceding chapters through the geometric approach. Risk management decisions are tied to a distance metric, namely the Mahalanobis distance, building upon the work of [Kritzman and Li \(2010\)](#). The authors introduced the squared Mahalanobis distance as a measure of financial turbulence and advocated for its use as a robust indicator for market instability, thereby assisting in portfolio decision-making and risk management.

Chapter 5 empirically and theoretically evaluates the predictive capability of the Mahalanobis-based indicator proposed by [Kritzman and Li \(2010\)](#). A set of jointly sufficient conditions regarding the data generating process of asset returns are defined to validate its suitability for dynamic hedging. The Mahalanobis-based indicator is framed as a tool for multivariate outlier detection, and methods for determining a threshold to distinguish periods of market turbulence from calmer times are discussed. The empirically observed property of the turbulence measure, which exhibits temporal clusters and generally coincides with negative market developments when experiencing spikes, is examined. These properties are traced back to two well-known stylized facts of asset returns, volatility clustering, and asymmetric volatility-return dependence, which are expected to persist in the future, providing a theoretical justification for a Mahalanobis-based hedging approach. However, through the help of a simulation study, it is found that these two properties alone are insufficient to support a Mahalanobis-based hedging strategy when the objective is not only to reduce risk (in terms of volatility and drawdowns) but also to maintain competitive returns. A third criterion, termed synchronous volatility, is identified as crucial.

As the main contribution of Chapter 5, a method is proposed to distinguish turbulence into positive and negative market developments, labeled as good and bad, utilizing a basic generative model, quadratic discriminant analysis. The approach entails computing the difference between two Mahalanobis distances under distinct market regimes, marking the second notable statistical application of the Mahalanobis distance beyond outlier detection in cluster analysis. The adaptation of a pattern recognition technique, initially designed for identifying economic cycles, facilitates the labeling of bear and bull markets, a prerequisite for learning class-specific parameters.

Various turbulence-based hedging rules are formulated and subjected to empirical examination. The assessment includes both a statistical framework, utilizing classification

metrics, and an economic framework, conducted through backtesting and the analysis of multiple performance measures. It becomes apparent that statistical significance does not seamlessly translate into commensurate economic reward, a phenomenon attributed to transaction costs. In this context, a data-driven smoothing approach, based on an orthogonal stochastic volatility model, is proposed to reduce trading activity while achieving a balanced trade-off between errors of type I (missed signals) and type II (false alarms).

References

- Efron, B. (1979). The bootstrap: Another look at the jackknife. *The Annals of Statistics*, 7:1 – 26.
- Kim, W. C. and Lee, Y. (2016). A uniformly distributed random portfolio. *Quantitative Finance*, 16:297–307.
- Kritzman, M. and Li, Y. (2010). Skulls, financial turbulence, and risk management. *Financial Analysts Journal*, 66:30 – 41.
- Lisi, F. (2011). Dicing with the market: Randomized procedures for evaluation of mutual funds. *Quantitative Finance*, 11:163–172.
- Rubin, R. (1981). The bayesian bootstrap. *The Annals of Statistics*, 9:130 – 134.

Chapter 2

Randomized Geometric Tools for Anomaly Detection in Stock Markets

Cyril Bachelard, Apostolos Chalkis¹, Vissarion Fisikopoulos¹ and Elias Tsigaridas²

¹ *National & Kapodistrian University of Athens, Greece*

² *Inria Paris and IMJ-PRG, Sorbonne Université and Paris Université*

A version of the chapter has been published in *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics, PMLR 206:9400-9416, 2023.*

Abstract

We propose novel randomized geometric tools to detect low-volatility anomalies in stock markets; a principal problem in financial economics. Our modeling of the (detection) problem results in sampling and estimating the (relative) volume of geodesically non-convex and non-connected spherical patches that arise by intersecting a non-standard simplex with a sphere. To sample, we introduce two novel Markov Chain Monte Carlo (MCMC) algorithms that exploit the geometry of the problem and employ state-of-the-art continuous geometric random walks (such as Billiard walk and Hit-and-Run) adapted on spherical patches. To our knowledge, this is the first geometric formulation and MCMC-based analysis of the volatility puzzle in stock markets. We have implemented our algorithms in C++ (along with an R interface) and we illustrate the power of our approach by performing extensive experiments on real data. Our analyses provide accurate detection and new insights into the distribution of portfolios' performance characteristics. Moreover, we use our tools to show that classical methods for low-volatility anomaly detection in finance form bad proxies that could lead to misleading or inaccurate results.

JEL classification: C15, C63, C88, C58, G12.

Keywords: Low-volatility anomaly, MCMC, geometric random walks.

2.1 Introduction

We consider two fundamental problems from two arguably distant disciplines: Computational Geometry and Financial Economics. The geometrical problem involves the computation of volume of and sampling from non-convex and possibly disconnected spherical patches arising by the intersection of a non-standard simplex with a sphere. The absence of convexity and the presence of many connected components make the problem very challenging from an algorithmic and implementation point of view. The geometric representation has a financial interpretation, which is the set of portfolios, i.e., investments in a collection of stocks, having a certain risk level. Thus, our motivation to solve this geometrically hard problem, aside from having an interest in its own right, stems from a concrete financial question about the feasible space of investable portfolios obeying certain regulatory and risk related constraints. The fundamental economic question that we eventually address is that of the relation between risk and return: do assets with higher risk levels provide higher returns, as suggested by economic theory, or is there an empirically observable anomaly? The question is a long-standing one and we give a glimpse into its history in the next section.

Our analysis and contribution to the topic consist of a differentiated perspective from the existing literature. The usual procedure is to form a small number of portfolios by first sorting stocks according to their historical volatility and grouping them into portfolios (typically equally weighted) with increasing riskiness in order to then assess their relative empirical performance. Instead of the simple heuristic approach to group stocks into risk buckets, we consider all possible combinations of stocks from a given investable universe, i.e., the entire space of feasible portfolios having a certain risk level (defined in terms of volatility). To achieve this goal we use sampling based on geometric random walks, i.e., Markov Chain Monte Carlo (MCMC) procedures, and compare their resulting empirical risk-return characteristic to those of portfolios sampled from other risk level sets. Hence, instead of comparing the descriptive performance statistics of a single portfolio to represent a risk level, we investigate the joint distribution of risk and return of the parameters of the average portfolio having a certain risk level. Analyzing two different stock markets (the USA and Europe) and subsets of the two markets distinguishing between company size (small versus large) and sector belongingness (cyclical versus defensive) we find that

the risk-return profiles of commonly employed equally weighted quintile portfolios formed by sorting stocks according to their historical volatility range anywhere in the bivariate distribution of realized risk and return. As such, we argue that one should be careful to draw conclusions drawn from the sorting-based point estimates. However, the analysis of sets of portfolios with given risk level, thus abstracting from the problem of a specific weighting scheme and the missing consideration of correlation structures among stocks, shows that the resulting risk-return cluster does indeed support the hypothesis of an anomaly, albeit less pronounced than when using the quintile approach. The sampling-based approach allows us to visualize how the distribution of risk and return widens with increasing ex-ante volatility. Having knowledge about the distribution of performance statistics is then valuable for statistical inference and significance testing, particularly in the context of financial data which display time-series structures and are non-Gaussian.

2.1.1 Financial background

It has long been recognized that the capital asset pricing model (CAPM), a cornerstone of financial economic theory and the workhorse model of classical capital market theory, independently developed by [Sharpe \(1964\)](#) and [Lintner \(1965\)](#) and [Mossin \(1966\)](#), does not do justice in explaining the complexity of real world market dynamics. Contrary to the equilibrium model's predicted simple positive linear relation between risk and expected return, higher risk is not generally rewarded with higher return in global stock markets. According to the CAPM, the return one should expect from an investment depends solely on the riskiness of the investment relative to a single factor which is the overall market. Investments which bear higher risk than the market portfolio should pay out a higher return in expectation, i.e., a risk premia. However, [Haugen and Heins \(1975\)](#) were the first to recognize that risk does not generate a special reward following the early warning signs coming from [Black et al. \(1972\)](#), [Miller and Scholes \(1972\)](#) and [Fama and MacBeth \(1973\)](#). Their finding has subsequently been confirmed by [Fama and French \(1992\)](#) and [Black \(1973\)](#).

Further studies have found a wealth of anomalies, i.e., systematic and persistent deviations of empirical observations from model prediction. Prominent examples include firm size ([Banz \(1981\)](#)), stocks with lower market capitalization tend to outperform stocks

with a higher market capitalization in the future), value (Rosenberg et al. (1985), stocks that have a low price relative to their fundamental value, commonly tracked via accounting ratios like price to book or price to earnings outperform high-value stocks), momentum (Jegadeesh and Titman (1993), stocks that have outperformed in the past tend to exhibit strong returns going forward) or quality (Asness et al. (2019), stocks which have low debt, stable earnings, consistent asset growth, and strong corporate governance, commonly identified using metrics like return to equity, debt to equity, and earnings variability).

Building upon the anomalous findings, the original single-factor CAPM has then been augmented with other factors besides the market, namely size and value Fama and French (1992), size value and momentum Carhart (1997) and size, value, and two quality factors Fama and French (2015). Nevertheless, despite the wealth of documented anomalies and cited extensions, the CAPM has shown great resilience vis-à-vis a transition to an alternative paradigm. This is particularly surprising in the light of a large body of literature subsumed under the term low-risk anomaly or low-volatility anomaly which directly attacks the very core of the CAPM by showing that, even after controlling for other factors as done by the latest CAPM extensions, (i) low-risk companies outperform high-risk companies and that (ii) low-risk portfolios produce higher risk-adjusted returns than capitalization weighted benchmarks.

In particular, Falkenstein (1994) found that, when controlling for size, the relation between risk and return gets reversed. Further, the outperformance of low-volatility stocks compared to high-volatility stocks has been shown to be robust among different markets, industries and sub-periods (see e.g., Blitz and van Vliet (2007), Blitz et al. (2013), Baker and Haugen (2012), van Vliet and de Koning (2017), Blitz et al. (2019)). Walkshäusl (2014)).

Note that in the CAPM, the risk is defined in terms of market beta, i.e., the ratio between an asset's volatility (standard deviation of returns) and the volatility of a market portfolio times their correlation (or the ratio of the asset's co-variation with the market portfolio over the variance of the market portfolio). The low-risk anomaly was found to be present irrespective of the employed measure of risk, be it beta or volatility (see e.g., Blitz and van Vliet (2007), Baker et al. (2011), Frazzini and Pedersen (2014), Liu et al. (2018), Asness et al. (2020)).

The typical approach pursued by the studies on the low-volatility anomaly is to sort



Figure 2.1: Risk-return profile of volatility quintile portfolios and the overall market portfolio for the U.S. stock market over nearly a century.

stocks according to their historical volatility and to form portfolios, either weighted equally or proportional to market capitalization, within quintiles of volatility levels. The process is then repeated on a monthly or quarterly basis, thus giving rise to, say, five (quintile) or ten (decile) backtested portfolios of increasing ex-ante volatility. Figure 2.1 illustrates the advantageous risk-return profile of the low-volatility portfolio compared to the other volatility quintiles and the overall market computed on nearly a century of U.S. stock market data¹.

The performance difference between the lowest and highest volatility stocks is called the low-volatility premia. It can be exploited, in principle, by forming long-short portfolios having positive weights (i.e., a long position) in the low-volatility stocks and negative weights (i.e., a short position meaning that one sells stocks) in the high-volatility stocks. In practice, however, many investors and investment vehicles like mutual funds are prohibited by regulation to short sell assets. Thus, their potential benefit from the anomalous risk-return relation is not to bet on the underperformance of high-volatility stocks, but to overweight low-volatility stocks while avoiding exposure to high-volatility titles.

Alternatively to sorting, the minimum variance portfolio provides another solution for a low-risk portfolio by means of optimization. A parallel stream of literature investigating the empirical performance of minimum variance portfolios aligns with the findings on

¹The investment universe consists of the 1000 largest stocks listed on the three major U.S. exchanges (NYSE, AMEX, and NASDAQ) from 1929 to 2020. Five quintile portfolios, each consisting of 200 stocks, were formed by sorting the investment universe by 36-month historical volatility. The portfolios are rebalanced quarterly and the positions within each quintile are equally weighted (weighting of 0.5% per share). The data set is provided by Pim van Vliet and Jan de Koning on their private website <https://www.paradoxinvesting.com/data/>.

low-risk portfolios based on sorts of single-stock risk measures thus further strengthening the existence of a low-risk anomaly (see e.g., [Haugen and Baker \(1991\)](#), [Clarke et al. \(2006\)](#), [Clarke et al. \(2011\)](#)).

Contrary to the pure volatility sorting-based portfolio construction method, the minimum variance procedure includes information about correlations in the portfolio selection process. As a result, the minimum variance portfolio is, although related, different from the sorting-based low-volatility portfolio for it might very well include medium- or even high-volatility stocks as long as they contribute to an overall decrease in portfolio volatility through low correlations.

Vice-versa, the ex-ante variances of portfolios formed from volatility-sorted subsets of the entire investable universe do not have to be monotonically increasing even though the average ex-ante volatilities of the assets within the groups are. This is, again, because of correlations.

Hence, the ignorance of correlations in the formation of volatility-ranked portfolios poses a drawback in the existing literature analyzing the volatility puzzle. Instead of clustering stocks according to volatility and representing the subgroups by a single portfolio, an alternative would be to pre-define certain volatility targets and to sample portfolios with exactly those ex-ante volatilities from the entire collection of investable firms. Using sampling, one not only considers correlations but further overcomes the somewhat arbitrary choice of the weighting scheme for the quantile-portfolios (equally or relative to firm size), which is known to have a large impact on performance and inference ([Plyakha et al., 2014](#)).

2.1.2 Connection to geometry, methodology, and computational challenges

Finance and economic textbooks usually represent the set of feasible portfolios in a risk-return space as in the left plot of [Figure 2.2](#). The blue parabolic curve, whose concave part is termed the efficient frontier, highlights the boundary of the feasible set (grey area), and the vertical lines correspond to iso-variance portfolios (color-coded by increasing volatility from green to red). Despite the fact that we adopt this economic representation to analyze the results of our empirical study of the low-volatility anomaly, our approach consists of representing the set of volatility constrained portfolios with geometrical objects. In

particular, we represent the set of long-only portfolios, i.e., portfolios with positive or zero weights, with the canonical simplex. Then, given the covariance matrix of assets' returns, the set of portfolios with a fixed level of volatility is the intersection between the canonical simplex and the boundary of an ellipsoid centered at the origin and whose shape and orientation are determined by the eigenvectors and eigenvalues of that matrix. The right plot of Figure 2.2 illustrates the situation for three hypothetical assets and five variance levels (corresponding to the y -coordinates of the colored vertical lines in the left plot of Figure 2.2), giving rise to five iso-variance ellipsoids intersecting the simplex (grey dotted surface). The parabolic efficient frontier in the left plot of Figure 2.2 transforms to a piecewise-linear line in blue in the right plot of Figure 2.2. In either representations, it highlights the set of mean-variance Pareto-efficient portfolios. Harry Markowitz, the father of modern portfolio theory, called it the critical line² in his seminal paper [Markowitz \(1992\)](#).

In the forthcoming empirical analysis, we estimate the covariance matrix from historical data and set several volatility levels that define a sequence of concentric ellipsoids intersecting the simplex. By sampling independently and uniformly from each intersection, we obtain sets of volatility-constrained portfolios. By computing the future returns for each sample we then capture the dependency between portfolio volatility and future portfolio return using several statistical tools from quantitative analysis. Note that one could directly extend this approach to capture the dependency between volatility and other portfolio scores.

From a geometric point of view, the intersection between the canonical simplex and the boundary of an ellipsoid in \mathbb{R}^d is a $(d - 1)$ -dimensional (geodesically) non-convex and nonconnected body (see Figure 2.3), thus, forming a very challenging problem. An additional challenge, in our case, comparing to existing work on manifold sampling, comes from non-connected-ness i.e. we have to estimate the volume of each connected part to achieve uniform sampling. Our approach relies on MCMC sampling and efficient practical Multiphase Monte Carlo (MMC) schemes for volume approximation.

Considering previous approaches, quantile or minimum variance portfolios correspond to points that belong to the boundary of the set of constrained volatility portfolios. In particular, they lie on the intersection between several facets of the simplex and the

²Note that Markowitz chose a geometric exposition of his ideas.

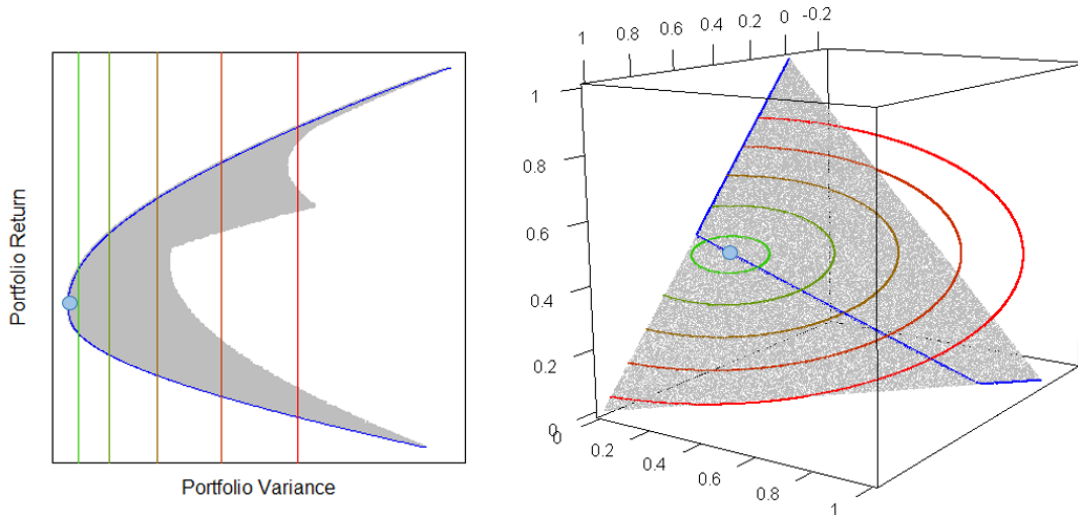


Figure 2.2: Left plot: Feasible set and iso-volatility lines in risk-return space. The grey area represents the set of feasible portfolios. Iso-volatility level sets form vertical lines and are color-coded by increasing variance from green to red. The blue curve highlights the set of Pareto-efficient portfolios in the trade-off between risk and return. The part above the minimum variance portfolio (light blue dot) is called the efficient frontier. Right plot: Feasible set and iso-volatility curves in asset weights space. The grey area depicts the simplex of feasible portfolios. Iso-volatility level sets form ellipsoidal curves, again color-coded by increasing variance from green to red. The blue line, called the critical line, highlights the set of Pareto-efficient portfolios in the trade-off between risk and return. Note that the minimum variance portfolio, again illustrated with a light blue dot, forms the centroid of the 2-dimensional iso-volatility ellipses and indicates the point where the 3-dimensional ellipsoid centred at the origin touches the simplex.

boundary of the ellipsoid. Consequently, the region around those portfolios is very unlikely to be sampled, i.e., be visited by the random walk. Thus, it is likely that the analysis of the volatility puzzle with those portfolios results in outliers compared to the results from a sampling-based approach.

High-dimensional sampling from multivariate distributions with MCMC algorithms is a fundamental problem with many applications in science and engineering (see e.g., [Iyengar \(1988\)](#), [Somerville \(1998\)](#), [Genz and Bretz \(2009\)](#), [Schellenberger and Palsson \(2009\)](#), [Venzke et al. \(2021\)](#)). In particular, multivariate integration over a convex set and volume approximation of such sets – a special case of integration – have accumulated a broad amount of effort over the last decades. Nevertheless, these problems are computationally hard if we want to solve them exactly in general dimension ([Dyer and Frieze, 1988](#)).

MCMC sampling algorithms have made remarkable progress in solving efficiently the problems of sampling and volume approximation of full-dimensional convex bodies in \mathbb{R}^d

while enjoying great theoretical guarantees (Chen et al., 2018; Lee and Vempala, 2018; Mangoubi and Vishnoi, 2019). Sampling from the boundary of a convex body has also been studied (Dieker and Vempala, 2015). However, these algorithms could not be used in our setting since they are focused either on full-dimensional convex bodies or on sampling from the entire boundary of a convex body instead of a part of it (as in our case).

Manifold sampling is a well-studied problem (Narayanan and Niyogi, 2006; Diaconis et al., 2013) with various applications, e.g., in machine learning (Byrne and Girolami, 2013). Of special interest is the case where the manifold is a hypersphere (Davidson et al., 2018; Grattarola et al., 2019; Reisinger et al., 2010). Moreover, sampling efficiently on constraint manifolds is a core problem in robotics (Ortiz-Haro et al., 2021). Finally, in Cong et al. (2017) they propose a method for sampling multivariate normal distributions truncated on the intersection of a set of hyperplanes.

From a practical perspective, theoretical sampling algorithms cannot be applied efficiently for real-life computations. For example, the asymptotic analysis by Lovász and Vempala (2006) hides some large constants in the complexity, and in Lee and Vempala (2018) the step of the random walk used for sampling is too small to be an efficient choice in practice.

Recently, practical volume algorithms have been designed by relaxing the theoretical guarantees and applying new algorithmic and statistical techniques; they are very efficient in practice and they also guarantee high accuracy results (Emiris and Fisikopoulos, 2014; Cousins and Vempala, 2016; Chalkis et al., 2019). Volume computation and uniform sampling have been shown to have useful applications in finance for crises detection (Calès et al., 2018) and efficient portfolio allocation and analysis (Pouchkarev et al., 2004; Hallerbach et al., 2002). In Chandrasekaran et al. (2010) they propose a polynomial-time algorithm for the more general problem of sampling and volume computation of star-shaped bodies, an important non-convex generalization of convex bodies.

In Abbasi-Yadkori et al. (2017) they prove that Hit-and-Run mixes fast in a more general setting that includes star-shaped bodies and spiral bodies appearing in motion planning. For the more general problem of approximating the volume of basic semi-algebraic sets (that is a set defined by a disjunction of polynomial equalities and inequalities) based on the so-called Moment-SOS hierarchy we refer the reader to Tacchi et al. (2020, 2022).

2.1.3 Contributions

We contribute to the literature on the low-volatility anomaly via a novel methodology to generate risk-sorted portfolios in high dimensions. Our approach directly samples uniformly distributed long-only portfolios having a certain level of volatility, which considers correlations and provides insights into the distribution of portfolios' performance statistics. Our empirical application of the geometric approach shows how the distributions of portfolios' performance statistics vary with the ex-ante volatility level and where the performance statistics of the standard sorting-based approach reside: are they close to the modes or outlying? Given that we find sorting-based results which are strong outliers with respect to (w.r.t.) the distribution of sampling-based statistics, we conclude that the classical sorting-based portfolios form bad proxies and one should be careful to base inference on them.

On the technical side, our contributions consist of the geometric modeling of the financial problem and the construction of efficient randomized geometric tools. First, we apply proper linear transformations to end up operating on the intersection of the unit ball with an arbitrary full-dimensional simplex. Then, we sample at the intersection and apply the inverse transformation to obtain volatility-constrained portfolios. We develop two new geometric random walks to sample from such spherical, non-convex, and nonconnected patches according to any given probability distribution. We also design a new MMC scheme to estimate the volume of a spherical patch. That is a practical randomized method based on simulated annealing and sampling from the Von-Mises Fischer distribution. Our MMC scheme generalizes and extends existing randomized volume approximation schemes (Cousins and Vempala (2015)). Last but not least, we offer an efficient open-source implementation in C++ with interface in R (see Section 2.3 for details).

Structure of the paper. The next section describes the geometric modeling of the financial problem in detail as well as the randomized geometric methods that we build, such as MCMC sampling and volume approximation. Section 2.3 contains details of our implementation and details about the algorithmic pipeline we introduce to analyze the volatility puzzle. Section 2.4 presents the results and the analysis of the volatility anomaly

in European and U.S. stock markets using our methods. Last, in Section 2.5 we discuss possible future work to improve the efficiency of our methods as well as other financial problems that our framework could be used to address.

2.2 Geometric modeling and algorithms

Our approach consists of the geometric modeling of long-only volatility-constrained portfolios. Within this framework, we develop two sampling algorithms and an MMC scheme for volume approximation. We first present the geometric framework and certain mathematical tools before presenting our algorithms.

Notations We denote a full dimensional (convex) body with a capital letter and if applicable with an index we denote its dimension; for example, the unit ball in \mathbb{R}^d is B^d . We denote the lower dimensional, usually (geodesically) non-convex, bodies with a calligraphic letter; for example \mathcal{S}_{d-1} is the $(d-1)$ dimensional sphere in \mathbb{R}^d . They are usually parts of the surface of a full-dimensional convex body.

2.2.1 Geometric modeling

In finance, a portfolio is a collection of assets. Each portfolio allocates a percentage of a given budget to every asset. The set of long-only portfolios is geometrically defined by the canonical d -dimensional simplex

$$\Delta_d := \{x \in \mathbb{R}^{d+1} \mid x_i \geq 0, \sum_{i=1}^{d+1} x_i = 1\} \subset \mathbb{R}^{d+1}, \quad (2.1)$$

where each point represents a portfolio and $d+1$ is the number of assets. The vertices represent portfolios composed entirely of a single asset. The portfolio weights, i.e., the fractions of investment for each asset, are non-negative and sum up to 1. Notice that Δ_d is a d -dimensional body that lies in \mathbb{R}^{d+1} ; that is, it is a *lower dimensional* body.

Given a vector of assets' returns $R \in \mathbb{R}^{d+1}$, and a portfolio $x \in \Delta_d$, we say that the return of x is $f_{ret}(x, R) = R^T x$. Similarly, given the positive definite covariance matrix $\Sigma \in \mathbb{R}^{(d+1) \times (d+1)}$ of the distribution of the assets' returns, we define the portfolio volatility as $f_{vol}(x, \Sigma) = x^T \Sigma x$. Thus, to model portfolios' volatility we employ ellipsoids intersecting

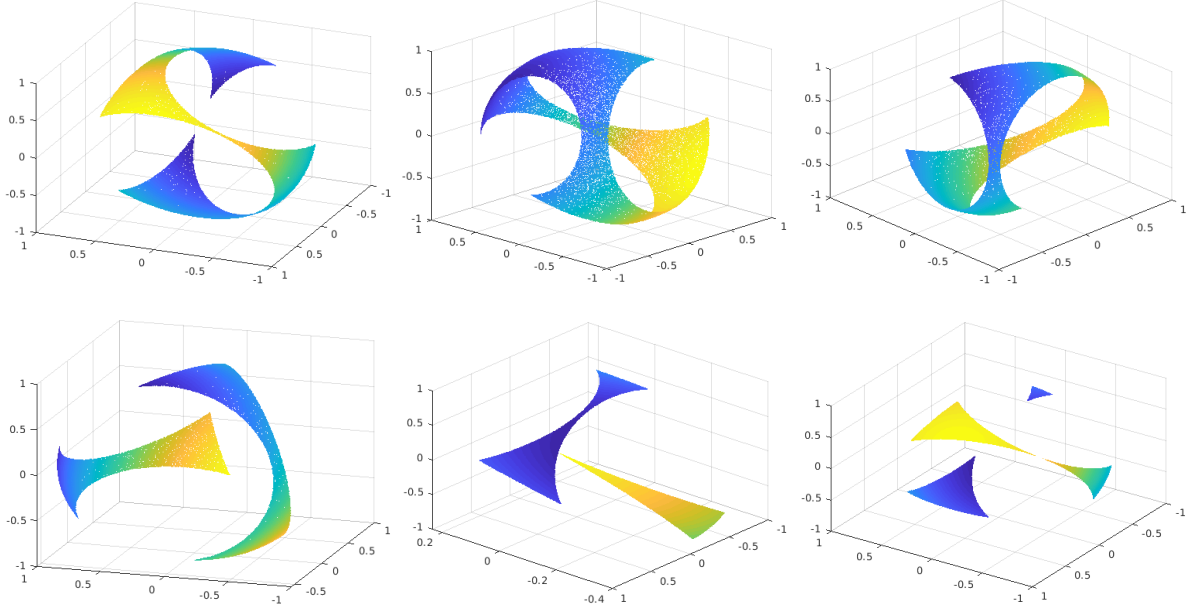


Figure 2.3: Six examples in \mathbb{R}^3 of the body $\mathcal{K} = \mathcal{S}_2 \cap \Delta$ (the unit sphere intersected by the interior of a simplex) that we sample from. In general, \mathcal{K} is a non-connected and geodesically non-convex body in \mathbb{R}^d .

Δ_d . The set,

$$\mathcal{E}_d^c \cap \Delta_d = \{x \in \Delta_d \mid x^T \Sigma x = c, c \in \mathbb{R}_+\} \subseteq \mathbb{R}^{d+1}, \quad (2.2)$$

corresponds to the portfolios with volatility c . Notice that portfolios (i.e., points) that belong to the ellipsoid (centered at the origin), $\mathcal{E}_d^c := \{x \in \mathbb{R}^{d+1} \mid x^T \Sigma x = c, c \in \mathbb{R}_+\} \subseteq \mathbb{R}^{d+1}$, achieve volatility equal to c . The portfolios in the interior of \mathcal{E}_d^c achieve lower volatility, while those in the complement of \mathcal{E}_d^c achieve higher volatility than c .

Geodesics on ellipsoids (as distinct from spheres) are not, in general, closed. That is, while we can compute geodesics on a sphere by exploiting spherical trigonometry, this is not the case for ellipsoids where one has instead to solve differential equations (Karney, 2012). Thus, for efficiency, we map the ellipsoid \mathcal{E}_d^c onto the unit hypersphere $\mathcal{S}_{d-1} \subset \mathbb{R}^d$ and we apply the same transformation to the simplex Δ_d to obtain a full-dimensional simplex $\Delta \in \mathbb{R}^d$. In particular, first, we use an orthonormal basis that spans the linear subspace that Δ_d lies on to obtain both an ellipsoid and a simplex in \mathbb{R}^d . Then, we transform the latter ellipsoid into \mathcal{S}_{d-1} and use this transformation to obtain a simplex $\Delta \in \mathbb{R}^d$. Therefore, instead of sampling from $\mathcal{E}_d^c \cap \Delta_d$, we sample from

$$\mathcal{K} := \mathcal{S}_{d-1} \cap \Delta. \quad (2.3)$$

We can use the inverse transformations to obtain uniformly distributed portfolios with volatility c since the transformation is isometric (Chalkis et al., 2021).

In general, \mathcal{K} is a set of geodesically non-convex³ spherical surface patches (see Figure 2.3 for a few examples). We call these patches the (connected) components of \mathcal{K} ; we denote them by \mathcal{K}_i , $i \in [M]$, where M is their cardinality. To sample uniformly from \mathcal{K} , we sample uniformly from each component according to its relative volume. In particular, first we sample $u \sim \mathcal{U}(0, 1)$. Afterwards, if $u \in [\sum_{i=1}^m w_i, \sum_{i=1}^{m+1} w_i]$, for some $m < M$, then we sample a uniformly distributed point from \mathcal{K}_m , where $w_i = \text{vol}(\mathcal{K}_i)/\text{vol}(\mathcal{K})$, $i \in [M]$ is the relative volume of the i -th component.

To identify and represent the components of \mathcal{K} we use the vertices and the edges of Δ . In particular, we consider the 1-skeleton of Δ i.e. the graph whose vertices are the vertices of Δ , with two vertices adjacent if they form the endpoints of an edge of Δ . Note that in the case of Δ the 1-skeleton is a clique.

We identify the edges of Δ that \mathcal{S}_{d-1} intersects and remove them from the 1-skeleton. We also identify the vertices of Δ that lie in the interior of \mathcal{S}_{d-1} and remove them as well as their adjacent edges. We denote the resulting graph by G . There is a bijection between the connected components of G and the connected components of \mathcal{K} . Thus, we represent each connected component of \mathcal{K} using the set of vertices of the corresponding connected component of G . To decide if a given point $p \in \mathcal{S}_{d-1}$ belongs to a certain component of \mathcal{K} we develop a membership oracle. One call costs $O(d^2)$ operations (see Appendix 2A).

2.2.2 Sampling from a connected component of \mathcal{K}

We introduce two geometric random walks, namely the Great Cycle Walk (GCW) and the Reflective Great Cycle Walk (ReGCW), to sample from a (connected) component of \mathcal{K} . To design these algorithms we employ the geodesics of \mathcal{S}_{d-1} . Note that the great circles on \mathcal{S}_{d-1} are the intersection of \mathcal{S}_{d-1} with 2-dimensional hyperplanes that pass through the origin in \mathbb{R}^d . These great circles are the geodesics of \mathcal{S}_{d-1} (Byrne and Girolami, 2013).

³A subset C of a surface is said to be a geodesically convex set if, given any two points in C , there is a unique minimizing geodesic contained within C that joins those two points. A geodesic is a curve representing in some sense the shortest path or arc between two points on a surface.

ALGORITHM 1: GCW(Δ, p, π)

Input : Simplex Δ ; point p ; probability density function π .

Require : $\mathcal{S}_{d-1} \cap \Delta \neq \emptyset$; point $p \in \mathcal{K} = \mathcal{S}_{d-1} \cap \Delta$

Output : Next Markov point in \mathcal{K}

Pick a uniform vector v from $\mathcal{S}_{d-1} \cap \{x \in \mathbb{R}^d \mid p^T x = 0\}$;

Let the great cycle $\ell(\theta) := \{p \cos \theta + v \sin \theta, \theta \in [0, 2\pi]\}$;

Let (θ^-, θ^+) be the values such that $\ell_p(\theta) := \{p \cos \theta + v \sin \theta, \theta \in [\theta^-, \theta^+]\}$ is the part of $\ell(\theta) \cap \Delta$ that contains p ;

Pick $\tilde{\theta}$ from $\ell_p(\theta)$ according to π_ℓ ;

return $p \cos \tilde{\theta} + v \sin \tilde{\theta}$;

Great Cycle Walk (GCW)

Great Cycle Walk (GCW) is a random walk to sample from any probability density function $\pi(x)$ supported on a connected component of \mathcal{K} . GCW generalizes the Hit-and-Run sampler (Bélisle et al., 1993) on a spherical patch. At each step, GCW starts from a point $p \in \mathcal{K}$ and picks uniformly a great cycle ℓ of S_{d-1} passing through p . Then, it computes the part of the great cycle that lies in \mathcal{K} and contains p . It samples a point from that part of ℓ according to π_ℓ to set the next Markov point, where π_ℓ is the restriction of π on ℓ . We prove (Theorem 1) that the (unique) stationary distribution of this algorithm is π and moreover, that GCW converges to π from any starting point in \mathcal{K} .

Some details of the computations are in order. GCW (at each step) chooses uniformly a great cycle passing from a point p by sampling uniformly at random a unit vector v from the hypersphere S_{d-1} restricted to the hyperplane $\mathcal{H}_p := \{x \in \mathbb{R}^d \mid p^T x = 0\}$. The parametric equation of the great cycle is

$$\ell(\theta) := \{p \cos \theta + v \sin \theta, \theta \in [-\pi, \pi]\}. \quad (2.4)$$

We compute v by sampling uniformly at random a point u in $\mathcal{S}_{d-1} \cap \mathcal{H}_p$ as

$$v = \frac{(I_d - pp^T)u}{\|(I_d - pp^T)u\|_2}, \quad (2.5)$$

where I_d is the $d \times d$ identity matrix. In this way, v is the normalized projection of u on the hyperplane \mathcal{H}_p and it is uniformly distributed in \mathcal{S}_{d-1} .

GCW computes the connected part of $\ell(\theta)$ that lies in Δ and contains p by computing

the intersection of $\ell(\theta)$ with each facet of Δ . For this, it computes the smallest positive and the largest negative solution of the following equations,

$$a_j^T \ell(\theta) = b_j \iff a_j^T x_i \cos \theta + a_j^T v_i \sin \theta = b_i, \theta \in [-\pi, \pi], j \in [d+1], \quad (2.6)$$

where $a_j \in \mathbb{R}^d$ are normal vectors of the facets of Δ . Let $z_j = b_j / \sqrt{(a_j^T x_i)^2 + (a_j^T v_i)^2}$. If $z_j \in [-1, 1]$, then the values of θ that correspond to the intersections are

$$\theta_j^-, \theta_j^+ = \pm \cos^{-1}(z_j) + \tan^{-1} \left(\frac{a_j^T v_i}{a_j^T p} \right). \quad (2.7)$$

Otherwise, the great cycle does not intersect with the j -th facet. Thus, GCW keeps $\theta^+ = \min_{j \leq m} \{\theta_j^+\}$ and $\theta^- = \max_{j \leq m} \{\theta_j^-\}$ and the intersection points are $\ell(\theta^+)$ and $\ell(\theta^-)$, respectively.

Theorem 1. *The stationary distribution of Alg. 1, $CGW(\Delta, p, \pi)$, where the starting point p belongs to \mathcal{K} , is π for any starting point p in \mathcal{K} .*

Finally, Δ has d facets while the computations of the intersection of $\ell(\theta)$ with a facet of Δ takes $O(d)$ arithmetic operations, which lead to the following remark.

Remark 2. *The cost per step of GCW is $O(d^2)$ arithmetic operations.*

GCW can also be used to sample uniformly from a component of \mathcal{K} (see Appendix 2B). However, the second geometric random that we introduce, namely Reflective Great Cycle Walk, has superior practical performance compared to GCW for uniform sampling.

Reflective Great Cycle Walk (ReGCW)

We introduce the Reflective Great Cycle Walk (ReGCW), a geometric random walk that operates on a connected component of \mathcal{K} and converges to the uniform distribution. ReGCW is a generalization of Billiard Walk (Gryazina and Polyak, 2014) on a spherical patch. Similar to GCW, it starts from a point p in a connected component \mathcal{K}_i . At each step, it generates uniformly a great cycle passing through p and computes a trajectory length $L = -\tau \ln \eta$, where η is a uniform number in $[0, 1]$, i.e., $\eta \sim \mathcal{U}[0, 1]$, and τ is a predefined constant. When the generated great cycle hits $\partial\Delta$, it reflects, and the reflected curve also forms part of a great cycle of \mathcal{S}_{d-1} . ReGCW returns the next Markov point

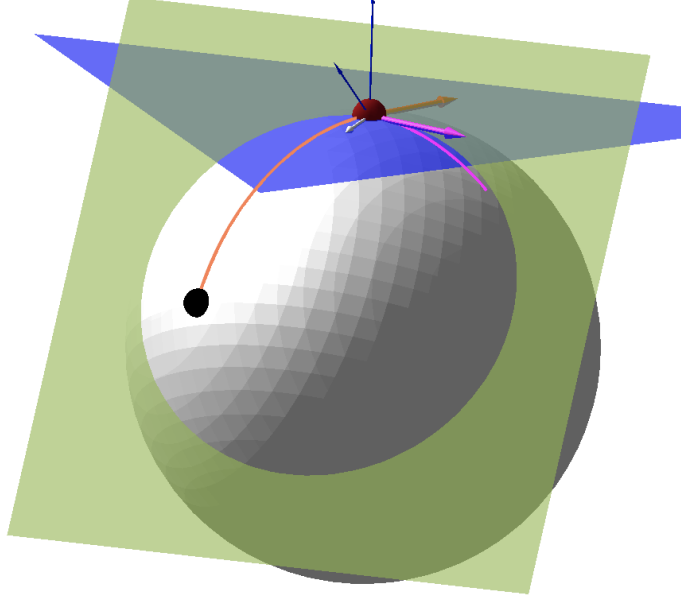


Figure 2.4: An illustration of the reflection of the ReGCW. The orange trajectory, starting from the black point, hits the boundary of the component defined by the intersection of the sphere with the green hyperplane (red point). The blue hyperplane is the tangent space at the intersection point. The resulted trajectory is in magenta.

after traveling a distance L . If the number of reflections exceeds a given upper bound $\rho \in \mathcal{N}_+$, then the next point is p itself. It is useful to set a bound on the number of reflections to avoid computationally hard cases where the trajectory may stick in corners. We explain our choices for τ and ρ in Section 2.3.

Some details of the computations are in order. At each step of ReGCW, when we are at a point p , we denote by $\ell(\theta) := \{p \cos \theta + v \sin \theta, \theta \in [0, L]\}$ the part of the great cycle emanating from p and having length L . Since we are on the unit sphere, geodesic length is numerically equal to the radian measure of the angles that the great circle arcs subtend at the center. We compute the smallest angle $\tilde{\theta}$ as in Equation (2.6), hence, $q := \ell(\tilde{\theta}) \in \partial\Delta$ is the point on the components boundary hit by the ReGCW trajectory. If $\tilde{\theta} < L$, then we compute the reflection of $\ell(\theta)$ at $\tilde{\theta}$ as follows: Let $a \in \mathbb{R}^d$ be the normal vector of the facet of P that $\ell(\theta)$ hits. We compute the normalized projection of a onto the hyperplane $\mathcal{H}_q := \{x \in \mathbb{R}^d \mid q^T x = 0\}$, say a' , using Equation (2.5). Then, the reflection of $\ell(\theta)$ at q is,

$$\ell_r(\theta) = \{q \cos \theta + v_r \sin \theta, \theta \in [0, L - \tilde{\theta}]\}, \quad (2.8)$$

where $v_r = v - 2(v^T a')a'$. We also update the length travelled so far, i.e., $L \leftarrow L - \tilde{\theta}$. When

ALGORITHM 2: Reflective Great Cycle Walk(Δ, p, ρ, τ)

Input : Simplex $\Delta \in \mathbb{R}^d$; current Markov point p ; upper bound on the number of reflections ρ ; length of trajectory parameter τ ;

Require : $\mathcal{S}_{d-1} \cap \Delta \neq \emptyset$; point $p \in \mathcal{S}_{d-1} \cap \Delta$

Output : Next Markov point in the same component of $\mathcal{S}_{d-1} \cap \Delta$ as p

$L \leftarrow -\tau \ln \eta, \eta \sim \mathcal{U}(0, 1)$ // length of the trajectory

$i \leftarrow 0$ // current number of reflections

$p_0 \leftarrow p$ // initial point of the step

Let the hyperplane $\mathcal{H}_p := \{x \in \mathbb{R}^d \mid p^T x = 0\}$;

Pick a uniform vector v from $\mathcal{S}_{d-1} \cap \mathcal{H}_p$;

do

 Let the curve $\ell(\theta) := \{p \cos \theta + v \sin \theta, \theta \in [0, L]\}$;

$\tilde{\theta} \leftarrow \arg \min_{\theta \in [0, L]} \{\ell(\theta) \in \partial \Delta\}$; // intersection angle

if $L < \tilde{\theta}$ **or** $\partial \Delta \cap \ell(\theta) = \emptyset$ **then return** $p \cos L + v \sin L$;

$p \leftarrow \ell(\tilde{\theta})$ and $v \leftarrow \ell'(\tilde{\theta})$; // point and direction update

 Let s the normalized projection of inner vector of the facet of Δ at p on \mathcal{H}_p ;

$v \leftarrow v - 2(v^T s)s$ // reflected direction

$L \leftarrow L - \tilde{\theta}$;

$i \leftarrow i + 1$;

while $i \leq \rho$;

if $i = \rho$ **then return** p_0 **else return** p ;

$L < \tilde{\theta}$ we set $\ell(L)$ as the next Markov point. The defined reflection operator guarantees that $v_r \in \mathcal{H}_q$ which implies that $\ell_r(\theta)$ is a part of a great cycle. A single reflection of the ReGCW is depicted in Figure 2.4.

Theorem 3. *The Reflective Great Cycle Walk of Alg. (2) has a unique stationary distribution, which is the uniform distribution. It converges from any starting point.*

Finally, ReGCW performs at most ρ reflections per step, while the computation of the intersection of $\ell(\theta)$ with $\partial \Delta$ costs $O(d^2)$ arithmetic operations, which leads to the following remark.

Remark 4. *The cost per step of ReGCW is $O(\rho d^2)$ arithmetic operations.*

2.2.3 Practical volume approximation

At a high level, our method is based on that of Lovász and Vempala (2006) and Cousins and Vempala (2015), as well as the practical variant presented in Cousins and Vempala (2016). However, those algorithms are designed for full dimensional bodies in \mathbb{R}^d . Thus, we

have to make the necessary practical adjustments to develop a practical volume estimation method. Since in our case we estimate the volume of geodesically non-convex spherical patches we do not have most of the theoretical guarantees found in previous work.

In particular, given a connected component \mathcal{K}_i , $i \in [M]$, for any sequence of k functions $f_j : \mathcal{S}_{d-1} \rightarrow \mathbb{R}_+$, $j \in [k]$ we consider the following representation of $\text{vol}(\mathcal{K}_i)$,

$$\text{vol}(\mathcal{K}_i) = \int_{\mathcal{K}_i} f_k dx \frac{\int_{\mathcal{K}_i} f_{k-1} dx}{\int_{\mathcal{K}_i} f_k dx} \frac{\int_{\mathcal{K}_i} f_{k-2} dx}{\int_{\mathcal{K}_i} f_{k-1} dx} \dots \frac{\int_{\mathcal{K}_i} dx}{\int_{\mathcal{K}_i} f_1 dx} \quad (2.9)$$

$$= \left(\frac{1}{\int_{\mathcal{K}_i} f_k dx} \frac{\int_{\mathcal{K}_i} f_k dx}{\int_{\mathcal{K}_i} f_{k-1} dx} \dots \frac{\int_{\mathcal{K}_i} f_1 dx}{\int_{\mathcal{K}_i} dx} \right)^{-1}, \quad \text{for } i \in [M], \quad (2.10)$$

while we prefer the right hand of the Equation (2.9) for reasons we explain in the sequel. We set each f_j to be proportional to the von-Mises Fischer (vMF) distribution, i.e.,

$$f_j(x) = e^{a_j(\mu^T x)}, \quad x, \mu \in \mathcal{S}_{d-1}, \quad a_j > 0, \quad (2.11)$$

where μ is the mean and a_j is the inverse of the variance. The vMF distribution is the restriction of the spherical Gaussian distribution on the hypersphere \mathcal{S}_{d-1} (Mardia, 1975).

Lovász and Vempala (2006) and Cousins and Vempala (2015) used exponential or Gaussian functions in Equation (2.9). They first compute the smallest variance α_k such that (s.t.) the mass of f_K is almost inside the convex body. Then, they compute the sequence of functions by cooling the variance until they reach the uniform distribution. In our case, we can not have a guarantee for α_k , and thus, we start from the uniform distribution and then we decrease the variance until we reach a distribution with sufficiently small variance. To stop decreasing the variance, in each step of our schedule, we sample from f_j and we probabilistically bound the proportion of the mass outside \mathcal{K}_i exploiting Bernoulli trials.

By standard error analysis (Jeter, 2005), to estimate $\text{vol}(\mathcal{K}_i)$ within relative error ϵ it suffices to estimate each integral ratio in Equation (2.9) within error $\epsilon_k = O(\epsilon/\sqrt{k})$ while the $\int_{\mathcal{K}_i} f_k dx$ is computed within an error $\epsilon_0 < \epsilon$.

To estimate each integral ratio within ϵ_k , we rewrite it as

$$Y_j = \frac{\int_{\mathcal{K}_i} f_j dx}{\int_{\mathcal{K}_i} f_{j-1} dx} = \int_{\mathcal{K}_i} \frac{f_j}{f_{j-1}} \frac{f_{j-1}}{\int_{\mathcal{K}_i} f_{j-1} dx}. \quad (2.12)$$

Then, we use GCW to generate N random samples from a distribution proportional to f_{j-1} and restricted to \mathcal{K}_i and we estimate each integral ratio using the following estimator

$$R_j = \frac{1}{N} \sum_{l=1}^N \frac{f_j(x_l)}{f_{j-1}(x_l)}, \quad (2.13)$$

while $E[Y_j] = \lim_{N \rightarrow \infty} R_j$. Using Chebyshev's inequality, when $Var[Y_j]/E[Y_j]^2 \leq 1$ we guarantee that $N = \tilde{O}(1)$ points suffice to approximate (2.12) within relative error ϵ_k (Cousins and Vempala, 2015).

Fixing the sequence. In Cousins and Vempala (2015) they prove that when f_j are Gaussian functions in \mathbb{R}^d , if $\alpha_j = \alpha_{j-1} \left(1 + \frac{1}{d}\right)$, then $Var[Y_j]/E[Y_j]^2 \leq 1$. To the best of our knowledge it is unclear whether Lemma 3.2 in Cousins and Vempala (2015) can be extended in our framework. To fix the sequence of variances we define the following practical annealing schedule based on the practical techniques in Cousins and Vempala (2016). We set

$$\alpha_j = \alpha_{j-1} \left(1 + \frac{1}{d}\right)^r, \quad r \in \mathbb{R}_+. \quad (2.14)$$

Then, we sample N' points with GCW from f_{j-1} and we search for the maximum r s.t. the ratio of the variance over the square of the average value of f_j/f_{j-1} evaluated on this sample lie in an interval $[1 - \delta, 1]$ for a predefined small value of δ . To estimate the desired value of r we binary search in an interval r_{\min}, r_{\max} , where $r_{\min} = 0$ and r_{\max} is found by setting $r_{\max} = 2^n$ and n the smallest integer s.t. the average value of the ratio f_j/f_{j-1} with $\alpha_j = \alpha_{j-1} \left(1 + \frac{1}{d}\right)^{r_{\max}}$ is larger than 1. We follow Cousins and Vempala (2016) to set N' for practical computations, and thus, we choose the value $1200 + d^2$.

First variance. To compute α_1 we set $\alpha_0 = 0$ and we use N' uniformly distributed points in \mathcal{K}_i generated by ReGCW. Then, we binary search for α_1 from a proper interval s.t. the average ratio between f_1 and the uniform distribution is smaller than 1.

Last variance. To stop we compute a large enough α_k s.t. almost the entire mass of the distribution proportional to f_k and restricted to \mathcal{K}_i is inside \mathcal{K}_i , i.e., for a predefined ϵ_0 we have,

$$\int_{\mathcal{K}_i} f_k dx = (1 - \epsilon_0) \int_{\mathcal{S}_{d-1}} f_k dx \quad (2.15)$$

with high probability. To achieve this objective, when we compute a new f_j , we sample

ν points from the corresponding exact vMF distribution supported on \mathcal{S}_{d-1} using the algorithm in [Kurz and Hanebeck \(2015\)](#) and we stop when less than $\epsilon_0\nu$ points are outside \mathcal{K}_i . Clearly, from Hoeffding’s inequality $\nu = O\left(\log\left(\frac{1}{1-\zeta}\right)/\epsilon_0^2\right)$ points suffices to guarantee that $\int_{\mathcal{K}_i} f_k dx \geq (1 - \epsilon_0) \int_{\mathcal{S}_{d-1}} f_k dx$ with probability $1 - \zeta$. For practical computations we set $\epsilon_0 = 0.05$.

Ratio convergence. If the points generated by GCW were independent, then we would use the theoretical bound on N , derived from Chebyshev’s inequality in [Cousins and Vempala \(2015\)](#) to estimate each integral ratio. However, these samples are correlated and thus, we use the same convergence criterion as in [Cousins and Vempala \(2016\)](#). In particular, for each point, we update the value of the integral ratio and we store the last values on a sliding window W . We declare convergence when

$$(\max(W) - \min(W))/\min(W) \leq \epsilon_k/2, \quad (2.16)$$

where $\max(W)$ and $\min(W)$ correspond to the maximum and minimum values of the sliding window respectively. As in [Cousins and Vempala \(2016\)](#) it is unclear how to obtain a good bound on the probability of failure with relation to the window size. To set the length of the sliding window we follow [Cousins and Vempala \(2016\)](#) and we set it equal to $4d^2 + 500$.

Sampling from a segment. To sample from the mVF distribution $\pi(x) \propto e^{\alpha(\mu^T x)}$ restricted to a connected component \mathcal{K}_i we use the GCW algorithm. In each step of the random walk we have to sample from mVF restricted to a curve $\ell(\theta) = \{p \cos \theta + v \sin \theta, \theta \in [\theta_1, \theta_2]\}$, i.e.,

$$\pi_\ell(\theta) \propto e^{\alpha(\mu^T p \cos \theta + \mu^T v \sin \theta)}, \quad \theta \in [\theta_1, \theta_2]. \quad (2.17)$$

To sample from the latter univariate distribution we use the Metropolis-Hastings algorithm ([Chib and Greenberg, 1995](#)) (see Section [2.3](#)).

2.3 Implementation and experiments

We perform an empirical study using previously introduced randomized geometric tools for the detection and analysis of the low-volatility anomaly on real data from global stock markets and comment on the performance of our implementation.

2.3.1 Implementation

We present the implementation of our algorithms and the tuning of various parameters. We provide a complete open-source software framework to address low-volatility detection in stock markets with hundreds of assets. Our code lies on a public domain⁴. The core of our implementation is in C++ to optimize performance while the user interface is implemented in R. The package employs `eigen` (Guennebaud et al., 2010) for linear algebra, `boost` (Maurer and Watanabe, 2017) for random number generation, and expands `volesti` (Chalkis and Fisikopoulos, 2021), an open-source library for high dimensional MCMC sampling and volume approximation. In our software we also use the package `LogConcDEAD` (Cule et al., 2009) to fit logconcave distributions for the analysis of the samples generated by our random walks (see Section 2.4).

Parameter tuning for practical performance

To achieve an efficient implementation for our methods, we introduce the following parameterizations.

Computing a starting point on K_i . We first compute the maximum inscribed ball in the intersection $\Delta \cap B_d$, where B_d is the unit ball. Since this is the intersection of a convex polytope with a ball, the maximum inscribed ball with center x_c and radius r is given by the optimal solution of the following conic program,

$$\max r, \quad \text{subject to : } a_i^T x_c + r \|a_i\| \leq b_i, \quad \|x_c\| \leq 1 - r, \quad (2.18)$$

where Δ is defined as the intersection of the half-spaces $a_i^T x \leq b_i$, $i = 1, \dots, d + 1$. Then, any vertex v of Δ that also belongs to the connected component of graph G (defined in Section 2.2.1) corresponds to a component \mathcal{K}_i and thus we can use it to obtain a point in \mathcal{K}_i . For this, we consider the segment defined by v and x_c , then, the intersection of the segment with \mathcal{S}_{d-1} lies in \mathcal{K}_i .

Convergence to the target distribution. We assess the quality of our results by employing a widely used MCMC diagnostic, namely the potential scale reduction factor (PSRF) (Gelman and Rubin, 1992). In particular, we compute the PSRF for each univariate

⁴https://github.com/TolisChal/volume_approximation/tree/low_volatility

marginal of the sample that both GCW and ReGCW output. Following e.g., [Gelman and Rubin \(1992\)](#), a convergence is satisfying according to PSRF when all the marginals have PSRF smaller than 1.1.

mVF restricted on a segment. To sample from mVF using GCW, in each step of the random walk, we have to sample from mVF constrained on a part of a great cycle. The goal is to sample from the univariate distribution $\pi_\ell(\theta) \propto e^{\alpha(\mu^T p \cos \theta + \mu^T v \sin \theta)}$, for $\theta \in [\theta_1, \theta_2]$. To perform this operation, we use Metropolis-Hastings ([Chib and Greenberg, 1995](#)). The proposal probability density we use is the uniform distribution on a segment of length $(\theta_2 - \theta_1)/3$, with the median being the current Markov point of GCW. We set the walk length (the number of Markov points to burn until storing a point) equal to a fixed value, i.e., 10, which we found to be an efficient choice empirically.

Parameters of ReGCW. To employ ReGCW (see Section [2.2.2](#)), we have to select efficient values for the parameter τ that controls the length of the trajectory in each step, and for ρ , the maximum number of reflections per step. To set τ for a component \mathcal{K}_i we sample $20d$ points with GCW with uniform target distribution. Then, we set τ equal to the length of the maximum geodesic chord in Equation [2.7](#) computing from those $20d$ steps of GCW. For the maximum number of reflections, we experimentally found that $\rho = 100d$ is a reasonable choice, leading to less than 0.1% of violations among all ReGCW steps.

2.3.2 Construction of volatility-constrained random portfolios and backtesting framework

We apply ReGCW sampling to construct sets of portfolios having a predefined variance and investigate whether the out-of-sample performance of so constructed portfolios varies in a systematic way with the variance level (do portfolios with higher variance deliver higher, lower, or equal returns?). Out-of-sample means that we analyze the future (ex-post) performance of portfolios formed with volatility targets derived from past (ex-ante) stock price information and that, at every point of the back-testing procedure, we only use information that was effectively available at that point in time. Starting in March 2002, the earliest possible date for our data set described below, the implementation consists of a three-step process that is applied initially and repeated every three months in order to account for new information about stock risks and index composition. The three steps

are data cleaning, covariance estimation, and sampling. In total, the quarterly reviews amount to 80 time points where portfolios are rebuilt by first, cutting out a historical data sample of five years of cleaned weekly returns to estimate the covariance matrix which defines the ellipsoidal portfolio-variance level sets. Our choice of covariance estimator is the non-linear shrinkage estimator of [Ledoit and Wolf \(2020\)](#) which has been shown to possess desirable properties in large-dimensional setups and is guaranteed to produce non-singular matrices and thus non-degenerate ellipsoids. Given the covariance matrix, five variance targets are computed from volatility-sorted quintile portfolios with equal weighting of within-quintile assets. From each of the five volatility level sets, 1000 portfolios are sampled.⁵ The investments are held over the following three months until the process is repeated. Ultimately, we arrive at a total of 5000 backtested portfolio price paths capturing the profits and losses endured over a period spanning from March 2002 to Dezember 2021 by randomly concatenating time series within each volatility cluster at the 80 rebalancing dates.

The entire procedure is repeated several times in order to control for size and sector effects by distinguishing between the 50% smallest and largest companies and by further labeling companies as either defensive or cyclical according to the sector classification methodology employed by MSCI⁶ building on the Global Industry Classification Standards (GICS)⁷.

2.3.3 Data

The data basis for our empirical study consists of two large universes of stock price series of companies covered by the MSCI USA and the MSCI Europe indices⁸ which encompass large and mid-cap equities traded in the U.S. and across 15 developed countries in Europe.

Our estimations are based on discrete weekly total⁹ returns using Wednesday closing

⁵Increasing the number of samples is not a bottleneck computationally (the complexity is linear). However, we found no additional economic value by adding more samples since qualitatively, our results did not change.

⁶Defensive sectors: Staples, Utilities, Energy and Health Care. Cyclical sectors: Financials, Real Estate, Information Technology, Discretionary, Industrials, Materials, Communication. For more detailed information see: https://www.msci.com/eqb/methodology/meth_docs/MSCI_Cyclical_and_Defensive_Sectors_Indexes_Methodology_Nov18.pdf

⁷<https://www.msci.com/our-solutions/indexes/gics>

⁸<https://www.msci.com/our-solutions/indexes/developed-markets>

⁹Returns, i.e., the percentage changes in prices from time $t - 1$ to t , are termed *total* when adjusted for dividends (i.e., dividends are re-invested).

prices denoted in local currencies. We use local currencies since we do not want to model any foreign exchange rates which would add another layer of volatility to the return series when expressed in a particular base currency (except for stocks that are already denominated in that currency). The currency effect is only relevant for the European market. Out-of-sample simulations are based on discrete daily total returns denoted in U.S. Dollars. Finally, we calculate all descriptive statistics and significance tests on discrete monthly returns as is customary in the financial industry.

The data cleaning process starts with adjustments for past corporate actions such as dividends, mergers and acquisitions, name changes, and other corporate actions. In addition, stocks that do not have enough history are excluded from the sample. To be included in the study, stocks need a consistent price history of five years, the equivalent of 260 weekly returns, without any gaps larger than two weeks. Further, otherwise, illiquid stocks are removed from the investable universe. As a threshold, we require a median trading volume over the previous 365 days to be above USD 1.5 million. We do this because, on the one hand, such illiquid stocks are not easily tradable and therefore would lead to a large implementation shortfall (i.e., the difference between a simulated performance and one obtained from real investments) and on the other hand, such companies display artificially low volatility due to a lack of trading and not because they are not risky. The cleaning process is necessary to ensure that, at every point in time, the investable universe only contains information that was effectively available at that point in time and to avoid any positive survivorship bias. Reference index membership over the full sample period is therefor not a requirement.

2.4 Results

Our experimental investigation of the low-volatility effect over the past nearly twenty years using introduced geometric tools allows for a series of interesting conclusions. In particular, the sampling-based method provides insights into the distribution of risk and return statistics as visualized in Figure 2.5. The charts show the risk-return profiles of the five clusters of backtested sampling-based portfolios for the U.S. (left plot) and European (right plot) markets where clusters are color-coded by increasing variance from green to red. Each point indicates the annualized performance statistics of a backtested

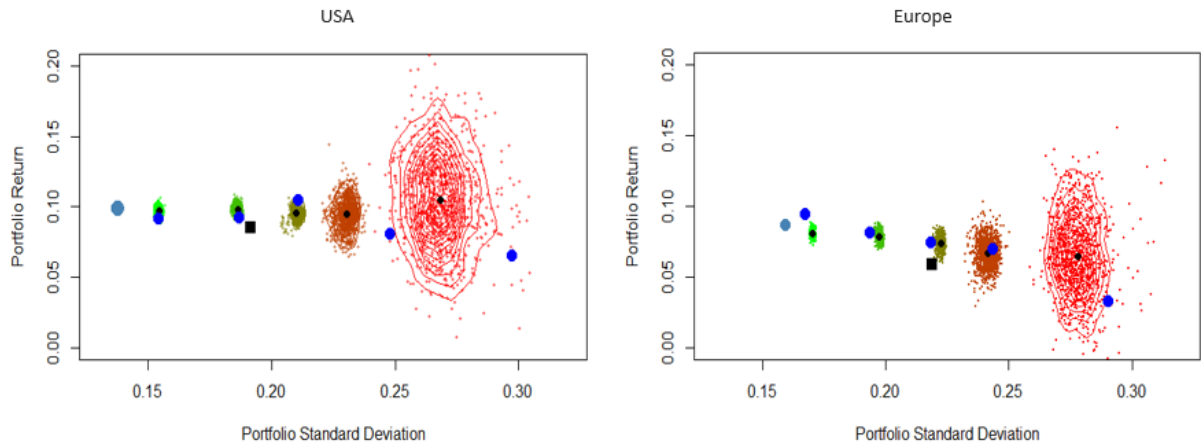


Figure 2.5: Risk-return profiles of the five clusters of backtested sampling-based portfolios for the U.S. (left) and European (right) market where clusters are color-coded by increasing variance from green to red. The black dots represent the cluster averages whereas blue dots depict the performance of the classical sorting-based quintile portfolios. The light blue dot further shows the performance of a minimum-variance portfolio backtest and the black square tells the performance of the capitalization-weighted market index.

strategy using the procedure described in section 2.3.2. The points per volatility group are overlaid by bivariate (nonparametric kernel) density contours lines. The black dots represent the cluster averages whereas blue dots depict the performance of the classical sorting-based quintile portfolios. The light blue dot further shows the performance of a minimum-variance portfolio backtest and the black square tells the performance of the capitalization-weighted market index. Inspection of Figure 2.5, as well as Figure 2.6 and 2.7 which display the same insights for the four sub-markets filtering for large defensive, small defensive, large cyclical, and small cyclical companies, reveals the following results.

First, we can confirm the presence of the low-volatility effect in both universes and corresponding sub-markets. Low-volatility portfolios historically delivered higher risk-adjusted returns than high-volatility portfolios, on average. However, the anomalous pattern is much more pronounced in Europe where we find a monotone decrease in return with increasing variance, i.e., very much the opposite of what would be predicted by the CAPM. In the U.S. market, the negative relation only holds for risk-adjusted performance measures like the Sharpe ratio¹⁰. The risk-return pattern is essentially flat with a slight increase in returns for the highest volatility level, an effect that could not be seen in the introductory example (see Figure 2.1) using the long-term U.S. dataset. In fact, the

¹⁰It is calculated by dividing the excess return of the investment (the return over the risk-free rate) by its standard deviation. For simplicity, we omit the risk-free rate.

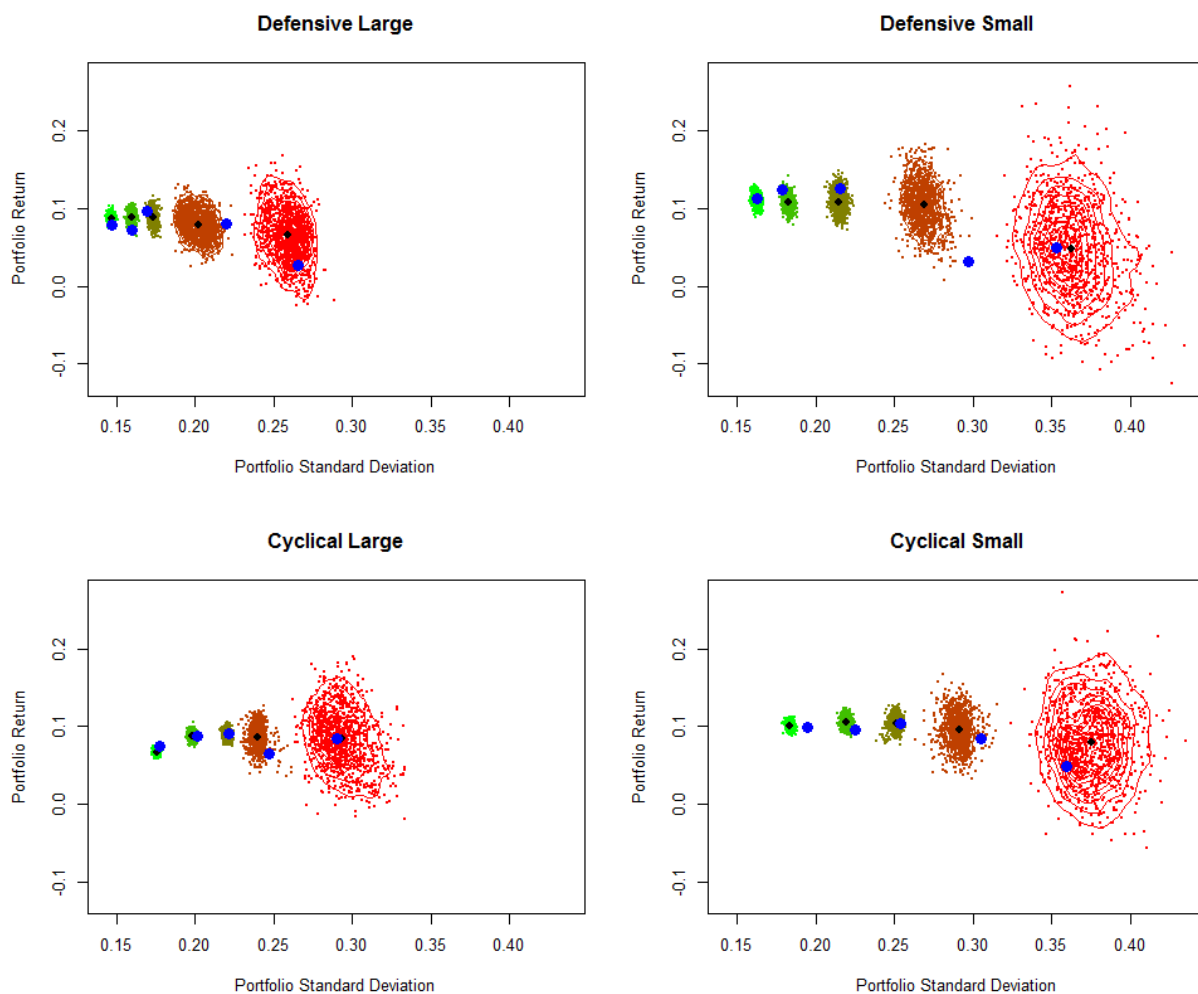


Figure 2.6: Risk-return profiles of the five clusters of backtested sampling-based portfolios for the U.S. market, grouped into four sub-markets according to company size and sector belongingness.

effect is to a large extent attributable to the stellar performance of mega-cap tech firms following the crash of March 2020 induced by the outbreak of the Covid-19 pandemic. If we omit the last two years in the analysis, i.e., stopping backtests at the end of 2019, the effect weakens and one finds a picture close to the one in the introductory example where the average of the highest volatility portfolios has the lowest cumulative returns while the largest cumulative returns are obtained by portfolios with moderately large variance (quintiles 2 and 3) and not by the low-volatility portfolio (quintile 1).

Detailed performance statistics are reported in Tables 2D.3 to 2D.8. They give the annualized returns, standard deviations, and Sharpe ratios per (sub-) universe for the five sorting-based quintile portfolio backtests and the average of the corresponding sampling-based portfolio backtests together with performance measures of the minimum variance

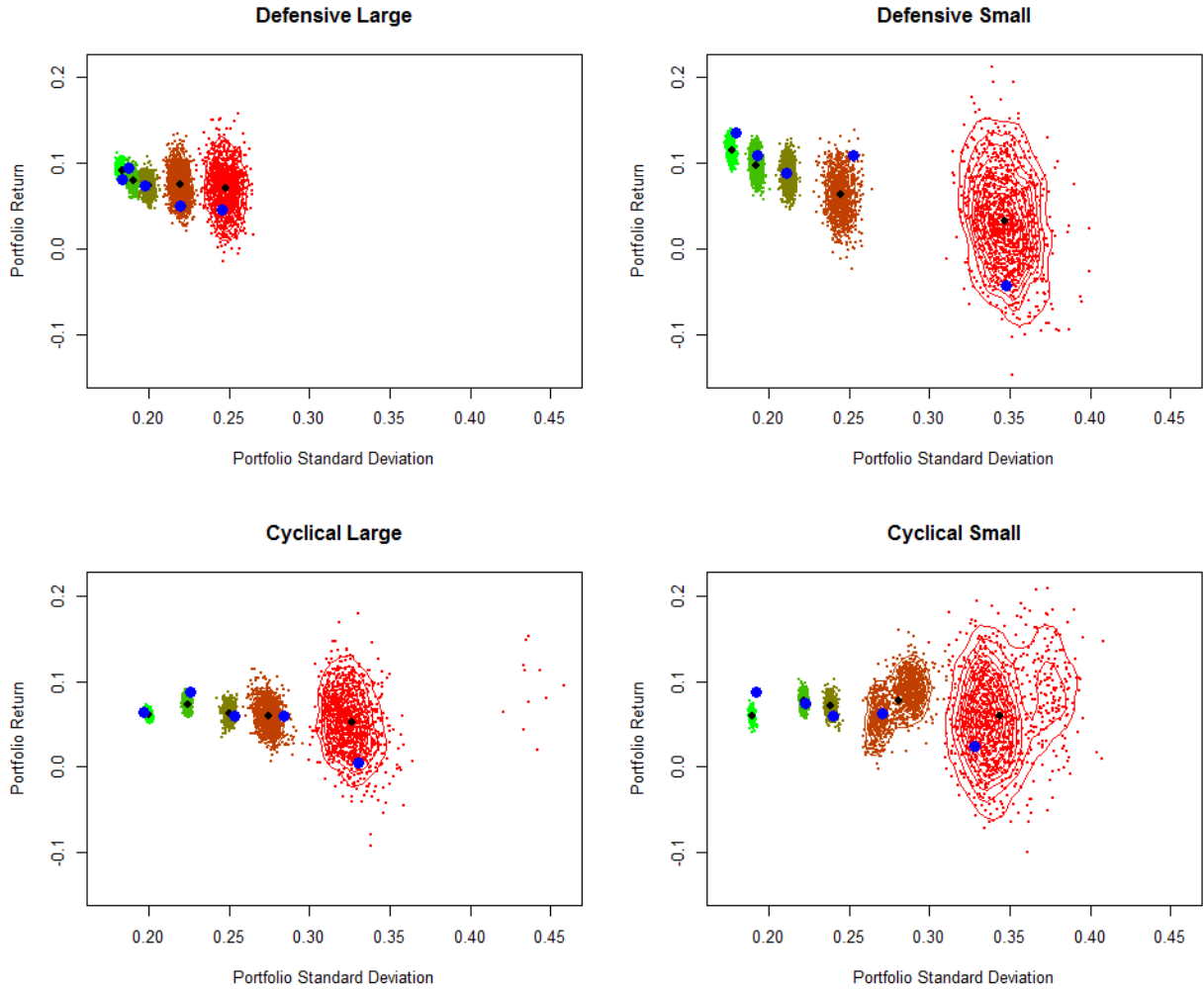


Figure 2.7: Risk-return profiles of the five clusters of backtested sampling-based portfolios for the European market, grouped into four sub-markets according to company size and sector belongingness.

portfolio and the capitalization-weighted index. Figure 2.8 further visualizes the decrease in Sharpe ratio with increasing variance, which is present in all (sub-) markets for both, sampling- and sorting-based backtests.

Second, Figures 2.5, 2.6 and 2.7 show that the variation in portfolio returns increases with the increasing ex-ante variance target and is largest for the highest variance group which contains both, the worst and the best performing portfolios. This shows that the construction method, i.e., the choice of weighting to form the volatility-targeting portfolios can heavily impact the final outcome. Further, also the spread in realized volatility levels increases with higher volatility targets, meaning that there is a larger estimation error for such portfolios.

Third, we observe that inference from the classical sorting-based method can be

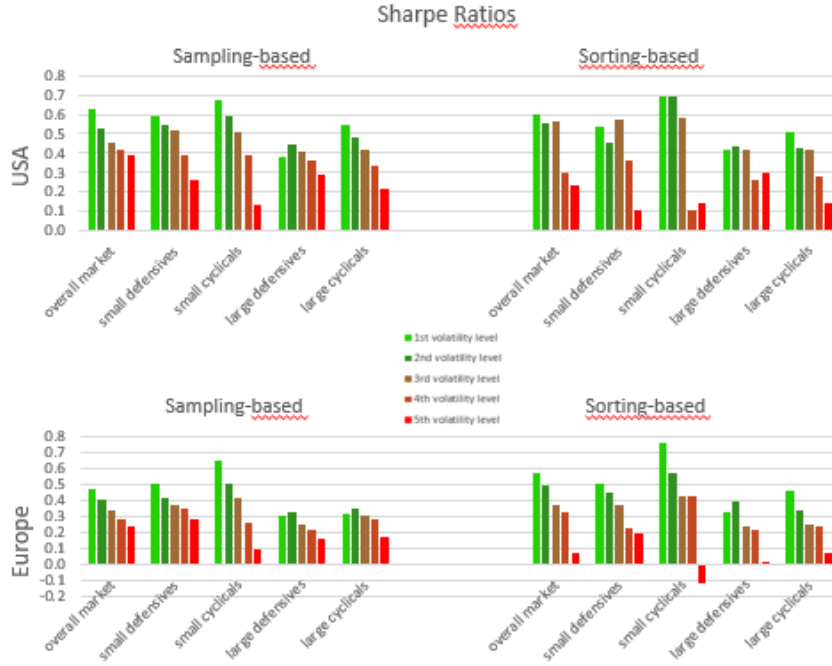


Figure 2.8: Sharpe ratios for the average sampling-based backtests and sorting-based backtests (quintile portfolios) in the U.S. and Europe, for the overall markets and the four sub-universes.

rather misleading in that resulting risk-return statistics can deviate substantially from the average sampling-based portfolio characteristics. Moreover, cross-checking the different sub-universes, the deviation seems to be non-systematic. In order to evaluate the likelihood of finding the sorting-based risk-return vector within the cloud of sampling-based performance vectors we parameterize the risk-return distribution by fitting a log-concave model. We use the non-parametric model in [Cule et al. \(2010\)](#) and its implementation in the R package [LogConcDEAD \(Cule et al., 2009\)](#). The function fits a log-concave density function to the data, where the logarithm of the density function is a tent function, and its support is the convex hull of the cluster. The fitted density gives the mode of the risk-return distribution (i.e. the most likely risk-return vector). Then, we compute the measures of a small rectangle centered on i) the mode, ii) the empirical cluster average and iii) the risk-return vectors of the sorting based quintile portfolios. In all ten (sub-) universes, we take the volume of the small rectangle to be 1% of the volume of the support of the distribution. The plot in [Figures 2E.14 and 2E.13](#) illustrate the probability density function (pdf) computed with the [LogConcDEAD](#) package and the quintile portfolios for each volatility level that corresponds to the plots in [Figure 2.5](#). Moreover, [Table 2D.9](#) reports the measures of the

rectangles. We notice that the measure of the area around the average is almost equal to that of the area around the mode, which is strong empirical evidence that the model in [Cule et al. \(2010\)](#) is a reasonable choice to evaluate the likelihood of finding a the performance vector of a particular backtest within each volatility cluster. Next, there is only one case – Europe, 4th volatility level – where the rectangular area around the sorting-based quintile portfolio statistics achieves almost the same probability as the one around the average or the mode. In four cases the performance vector of the quintile portfolio is even outside the convex hull of the cluster, i.e., the measure is zero. In another four cases, the probability of the area around the quintile portfolio statistics is more than ten times smaller than that of the average or the mode and in one case it is four times smaller. Consequently, according to the employed log-concave model, in most of the cases it is unlikely to find the performance results obtained from the sorting-based quintile portfolio backtests among the results of the sampling-based backtests within a volatility cluster. Moreover, there is not any pattern with respect to the volatility levels that would indicate when it is likely or not for the quintile portfolio performance to be found.

By design, the sorting-based portfolio is a particular solution to the sampling-based set. Hence, our finding that the sorting-based backtest statistics can vary strongly from the cluster averages may be surprising. However, the reason as to why the sorting-based quintile portfolio statistics are sometimes far outlying the distributions of sampling-based portfolio statistics has a geometric explanation. Sorting-based quintile portfolios with equal weighting form centroidal points on faces of the simplex, i.e., boundary points of both the simplex and the ellipsoid component of a volatility level (see [Section 2.2](#)). Typically, the areas around these points are highly unlikely to be sampled as they usually lie in a low-volume area of the sampling space, e.g., a corner. Hence, quintile portfolios should not be used to represent the group of portfolios with a specific volatility level.

Looking at the backtests on the different (sub-) universes we find that, overall, the low-volatility effect in terms of return gap between high-volatility and low-volatility is overestimated by the sorting-based approach, and particularly so in the European markets. Moreover, the variation in return among the five volatility clusters is more erratic for the sorting-based method than for the mode of the sampling-based portfolio return distributions. Not surprising is the observation that stocks from small and cyclical companies are more volatile than those of large and defensive companies (irrespective of whether we consider

sorting- or sampling-based portfolios).

A fourth insight, which however is not apparent visually, is that the risk-return relation of portfolios within volatility cluster are mostly negative. I.e., even for portfolios that, by construction, have the same ex-ante volatility, one observes that ex-post, those with lower volatility generally lead to a higher risk-adjusted return. This shows in terms of negative coefficients of the correlation between annualized returns and standard deviations of backtests within a volatility cluster, i.e., a sort of within-cluster low-volatility anomaly.

In the next step, we analyze the statistical significance in the difference of Sharpe ratios between the lowest and highest volatility portfolios. We employ a Sharpe ratio test [Ledoit and Wolf \(2008\)](#) which accounts for time series structures in the data by employing heteroscedasticity and autocorrelation consistent (HAC) estimates of standard error. In the U.S., the null of equal Sharpe ratios among the sorting-based portfolios can not be rejected at the 5% significance level. This also holds for any of the analyzed sub-markets except for the group of small defensives stocks. In Europe, the results are more mixed. The overall market as well as the sub-markets small defensives and small cyclicals show a significant Sharpe ratio difference, while large companies on the two-sector groups do not. Using the sampling-based simulations, we then run the test on all pairwise combinations of high minus low volatility simulations and count the number of significant t-statistics. We get less than 10% significant test results in the U.S. and about 15% in Europe, which, at first glance, does not exactly speak for the existence of an anomaly. However, if we discriminate the pairwise differences at a level of zero, i.e., we divide the sample between positive and negative Sharpe ratio differences, we observe that 98% and 99% of the Sharpe ratio differences are positive in the U.S. and European markets respectively and all significant test results come from the subset with positive Sharpe ratio differences. I.e., we only find significant positive differences and no significant negative differences. [Figure 2.9](#) further shows the distribution of p-values of the pairwise test statistics for series with positive Sharpe ratio difference in green and negative Sharpe ratio difference in red for the U.S. (left plot) and European (right plot) markets. Details of the test statistics for the two universes and sub-markets are contained in [Tables 2.1](#) and [2.2](#).

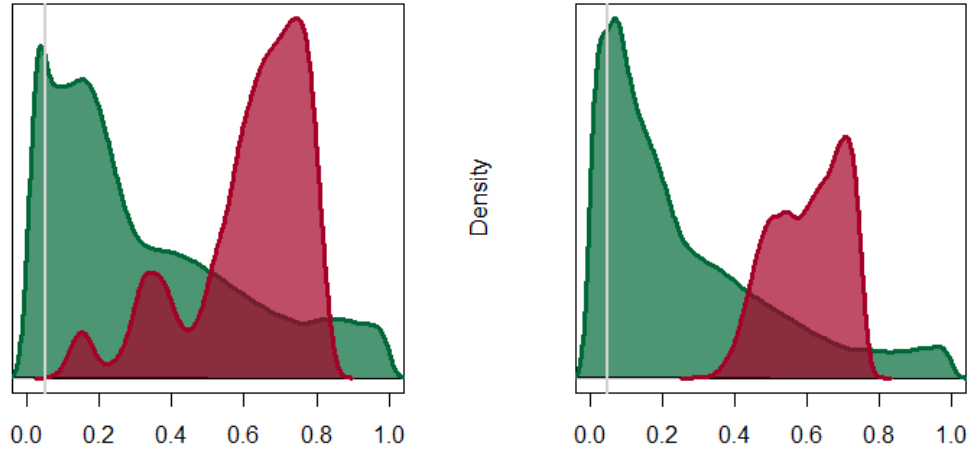


Figure 2.9: Distributions of p-values of the Ledoit-Wolf Sharpe ratio test between low-volatility and high-volatility portfolios, grouped by positive (green) and negative (red) Sharpe ratio differences of low minus high volatility portfolio backtest in the U.S. (left) and European (right) market. The grey vertical lines mark the 5% significance level.

The picture on the more granular analysis of the sub-markets discriminating among sectors and company size shows a rather consistent (relative) pattern among the two universes. In both cases, the highest proportion of positive Sharpe ratios is found among small defensives with a proportion of significant p-values of almost 60% in the U.S. (while the p-value obtained on the sorting-based backtests is barely significant) and almost 70% in Europe. Further, both universes display the smallest proportion of positive Sharpe ratios among the group of large cyclicals.

USA		overall market	large defensives	small defensives	large cyclicals	small cyclicals
Sorting-based	p-values	0.073	0.084	0.049	0.626	0.069
Sampling-based	Proportion of significant p-values	0.114	0.254	0.567	0.004	0.119
	Proportion of positive SR	0.981	0.993	0.998	0.772	0.995
	Proportion of significant p-values on positive SR subset	0.116	0.256	0.569	0.006	0.120
	Proportion of significant p-values on negative SR subset	0	0	0	0	0

Table 2.1: Sharpe ratio test results for the U.S. market.

Europe		overall market	large defensives	small defensives	large cyclicals	small cyclicals
Sorting-based	p-values	0.010	0.084	0.000	0.060	0.044
Sampling-based	Proportion of significant p-values	0.152	0.140	0.678	0.019	0.028
	Proportion of positive SR	0.990	0.972	1	0.901	0.822
	Proportion of significant p-values on positive SR subset	0.153	0.144	0.678	0.021	0.033
	Proportion of significant p-values on negative SR subset	0	0	0	0	0.002

Table 2.2: Sharpe ratio test results for the European market.

We conclude that, while test statistics based on the sorting-based quintile portfolios do not provide an unambiguous picture, it is overwhelmingly clear from the descriptive analysis of sampling-based portfolios that low-volatility portfolios have delivered higher Sharpe ratios than high-volatility portfolios, irrespective of the weighting scheme used to form the portfolios.

As a final step, we repeat our analysis in a slightly modified version. Instead of randomly concatenating the sampling-based portfolio simulations at the rebalancing dates, we now order sampled portfolios within a volatility level by their last period return such that the currently best performing allocation matches with the previously best-performing portfolio, the currently second best with the previous second best, and so forth up to the current worst with the previous worst. The idea is that a concatenation based on a systematic ordering of the simulations according to a cross-sectional (in-sample) momentum criteria might produce a distinctive pattern out-of-sample. And indeed it does. Doing the concatenation over time as described, each of the 1000 backtests per volatility level describes a dynamic strategy in the sense that the allocation changes every three months and the backtests are ranked by their quarterly performances. The first backtest concatenates all portfolios which ranked first in terms of quarterly returns. The pattern that emerges is a positive relation between in-sample rank and out-of-sample volatility. The backtest composed of the best performing portfolios has one of the lowest volatility over the entire nearly twenty years period while the backtest composed of the worst quarterly performers shows one of the highest volatility (always within a volatility-cluster). The pattern is very strong and shows up in all (sub-) markets. Portfolios with the same ex-ante volatility, when sorted by ex-ante momentum and concatenated over time, display a strong negative relation

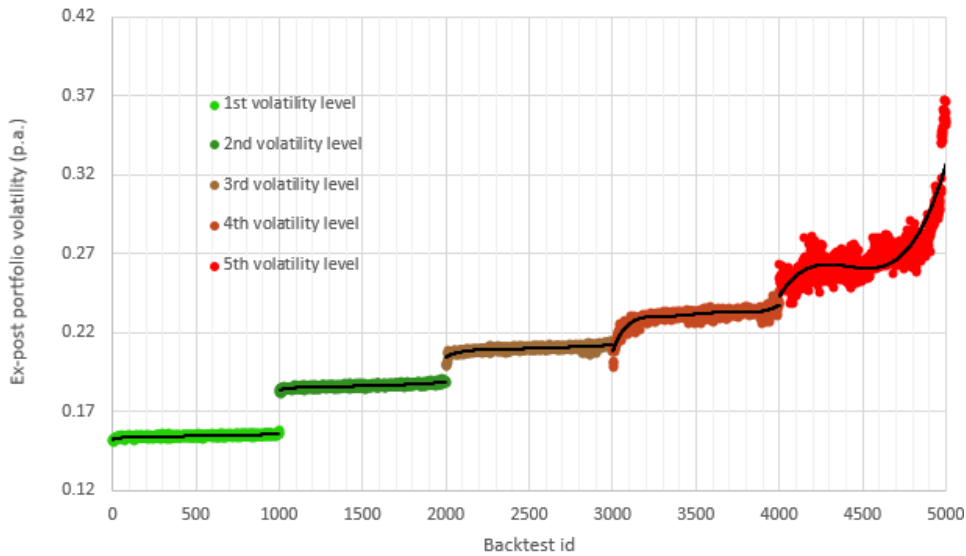


Figure 2.10: Ex-post volatility of sampling-based backtests on the U.S. market where concatenation of random portfolios at the rebalancing dates is made according to a momentum sort. Points with the same color have the same in-sample volatility. Points per volatility cluster are overlayed by a polynomial regression curve. The curves are increasingly upward sloping with increasing volatility level meaning that backtests combined of the best past performers display (much) lower volatility than backtests composed of past losers.

between the sorting-rank and ex-post volatility. The effect clearly shows in the increasingly upward sloping scatterplots in Figure 2.10 for the five color-coded volatility clusters in the U.S. market. In each volatility cluster, the backtests are ordered by rank (from best to worst) from left to right. Points are overlayed with a polynomial regression line to highlight the upward trend. Results for Europe are equivalent qualitatively, as can be seen in Figure 2.11 which shows the behavior for the four European sub-markets.

The altered perspective provides further evidence in favor of the low-volatility anomaly. However, the economic rationale and the causal nature of the very pronounced return-risk (rather than risk-return) relation is not clear to us at this stage. We leave the investigation to future research.

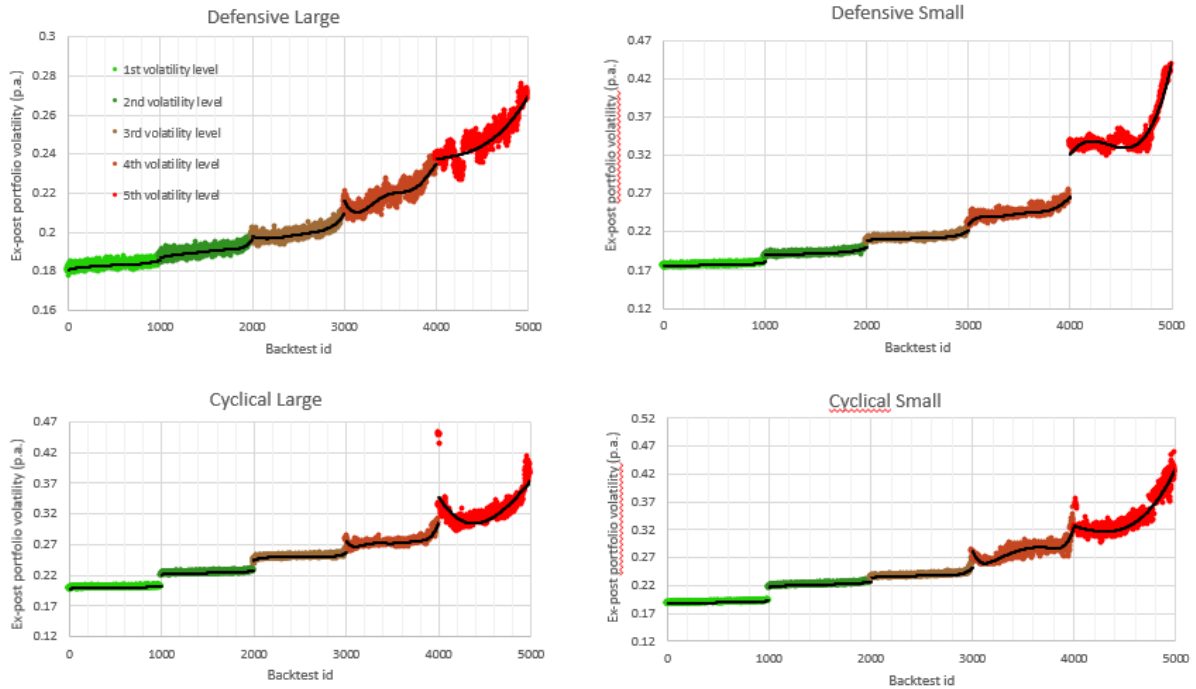


Figure 2.11: Ex-post volatility of sampling-based backtests on the four subsets of the European market where concatenation of random portfolios at the rebalancing dates is made according to a momentum sort. Points with the same color have the same in-sample volatility. Points per volatility cluster are overlaid by a polynomial regression curve. The curves are increasingly upward sloping with increasing volatility level meaning that backtests combined of the best past performers display (much) lower volatility than backtests composed of past loser portfolios.

2.5 Discussion

We proposed novel randomized geometric methods that can be used to analyze stock markets for anomalies. The tools have allowed us to sample portfolios with a fixed variance and to investigate their performance characteristic over time, leading us to conclude that the low-volatility anomaly is very much present in the U.S. and Europe as well as in sub-markets discriminating for sector belongingness and company size. Additionally, we discovered that the classical sorting-based approach, which constructs quantile portfolios, poorly represents the range of possible portfolios with a specific volatility. Therefore, it is essential to exercise caution when drawing inferences based on these portfolios.

The main limitation of our method lies in the "bottlenecks" that appear in the non-convex components (i.e. higher-dimensional analogs of the ones depicted in Figure 2.3) which prevents us to perform computations with more complex data.

Performance-wise, our method can be enhanced with more efficient samplers such as

Hamiltonian Monte-Carlo (HMC) ([Byrne and Girolami, 2013](#)) as well as by exploiting properties of the spherical patches. Furthermore, instead of using the simplex, general polytopes could be employed to incorporate additional constraints such as upper bounds on individual stocks or groups of stocks.

References

- Abbasi-Yadkori, Y., Bartlett, P., Gabillon, V., and Malek, A. (2017). Hit-and-run for sampling and planning in non-convex spaces. In *Proc. 20th Intern. Conf. Artificial Intelligence & Stat. (AISTATS)*, pages 888–895.
- Asness, C. S., A. Frazzini, N. J. G., and Pedersen, L. H. (2020). Betting against correlation. *Journal of Financial Economics*, 135:629 – 652.
- Asness, C. S., Frazzini, A., and Pedersen, L. H. (2019). Quality minus junk. *Review of Accounting Studies*, 24:34 – 112.
- Baker, M., Bradley, B., and Wurgler, J. (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analysts Journal*, 67:40 – 54.
- Baker, N. L. and Haugen, R. A. (2012). Low risk stocks outperform within all observable markets of the world. *Working paper no. 2055431*.
- Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9:3 – 18.
- Black, F. (1973). Beta and return: Announcements of the ‘death of beta’ seem premature. *The Journal of Portfolio Management*, 20:11 – 18.
- Black, F., Jensen, M. C., and Scholes, M. (1972). Asset pricing model: Some empirical tests. pages 79 – 121.
- Blitz, D., Baltussen, G., and van Vliet, P. (2019). When equity factors drop their shorts. *Working paper, 2019*.
- Blitz, D., Pang, J., and van Vliet, P. (2013). The volatility effect in emerging markets. *Emerging Markets Review*, 16:31 – 45.
- Blitz, D. and van Vliet, P. (2007). The volatility effect. *The Journal of Portfolio Management*, 34:102 – 113.
- Byrne, S. and Girolami, M. (2013). Geodesic monte carlo on embedded manifolds. *Scandinavian Journal of Statistics*, 40(4):825–845.

- Bélisle, C. J. P., Romeijn, H. E., and Smith, R. L. (1993). Hit-and-run algorithms for generating multivariate distributions. *Mathematics of Operations Research*, 18(2):255–266.
- Calès, L., Chalkis, A., Emiris, I. Z., and Fisikopoulos, V. (2018). Practical Volume Computation of Structured Convex Bodies, and an Application to Modeling Portfolio Dependencies and Financial Crises. In Speckmann, B. and Tóth, C. D., editors, *34th International Symposium on Computational Geometry (SoCG 2018)*, volume 99 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 19:1–19:15, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52:57 – 82.
- Chalkis, A., Emiris, I. Z., and Fisikopoulos, V. (2019). Practical volume estimation by a new annealing schedule for cooling convex bodies.
- Chalkis, A. and Fisikopoulos, V. (2021). volesti: Volume Approximation and Sampling for Convex Polytopes in R. *The R Journal*.
- Chalkis, A., Fisikopoulos, V., Tsigaridas, E., and Zafeiropoulos, H. (2021). Geometric Algorithms for Sampling the Flux Space of Metabolic Networks. In Buchin, K. and Colin de Verdière, E., editors, *37th International Symposium on Computational Geometry (SoCG 2021)*, volume 189 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 21:1–21:16, Dagstuhl, Germany. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- Chandrasekaran, K., Dadush, D., and Vempala, S. (2010). Thin partitions: Isoperimetric inequalities and a sampling algorithm for star shaped bodies. pages 1630–1645.
- Chen, Y., Dwivedi, R., Wainwright, M., and Yu, B. (2018). Fast MCMC sampling algorithms on polytopes. *Journal of Machine Learning Research*, 19(55):1–86.
- Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335.

- Clarke, R. H., de Silva, H., and Thorley, S. (2006). Minimum-variance portfolios in the us equity market. *The Journal of Portfolio Management*, 33:10 – 24.
- Clarke, R. H., de Silva, H., and Thorley, S. (2011). Minimum-variance portfolio composition. *The Journal of Portfolio Management*, 37:31 – 45.
- Cong, Y., Chen, B., and Zhou, M. (2017). Fast Simulation of Hyperplane-Truncated Multivariate Normal Distributions. *Bayesian Analysis*, 12(4):1017 – 1037.
- Cousins, B. and Vempala, S. (2015). Bypassing KLS: Gaussian cooling and an $O^*(n^3)$ volume algorithm. In *Proc. ACM STOC*, pages 539–548.
- Cousins, B. and Vempala, S. (2016). A practical volume algorithm. *Mathematical Programming Computation*, 8(2).
- Cule, M., Gramacy, R. B., and Samworth, R. (2009). Logconcded: An r package for maximum likelihood estimation of a multivariate log-concave density. *Journal of Statistical Software*, 29(2):1–20.
- Cule, M., Samworth, R., and Stewart, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(5):545–607.
- Davidson, T. R., Falorsi, L., Cao, N. D., Kipf, T., and Tomczak, J. M. (2018). Hyper-spherical variational auto-encoders. In *UAI*.
- Diaconis, P., Holmes, S., and Shahshahani, M. (2013). Sampling from a manifold. In *Advances in modern statistical theory and applications. A Festschrift in honor of Morris L. Eaton*, pages 102–125. Beachwood, OH: IMS, Institute of Mathematical Statistics.
- Dieker, A. B. and Vempala, S. S. (2015). Stochastic billiards for sampling from the boundary of a convex set. *Mathematics of Operations Research*, 40(4):888–901.
- Dyer, M. and Frieze, A. (1988). On the complexity of computing the volume of a polyhedron. *SIAM Journal on Computing*, 17(5):967–974.
- Emiris, I. and Fisikopoulos, V. (2014). Practical polytope volume approximation. *ACM Transactions of Mathematical Software*, 2018, 44(4):38:1–38:21.

- Falkenstein, E. G. (1994). Mutual funds, idiosyncratic variance, and asset returns. *PhD thesis, Northwestern University*.
- Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47:427 – 465.
- Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116:1 – 22.
- Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81:607 – 636.
- Frazzini, A. and Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, 111:1 – 25.
- Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472. Publisher: Institute of Mathematical Statistics.
- Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Springer Publishing Company, Incorporated, 1st edition.
- Grattarola, D., Livi, L., and Alippi, C. (2019). Adversarial autoencoders with constant-curvature latent manifolds. *Applied Soft Computing*, 81:105511.
- Gryazina, E. and Polyak, B. (2014). Random sampling: Billiard walk algorithm. *European Journal of Operational Research*, 238(2):497 – 504.
- Guennebaud, G., Jacob, B., et al. (2010). *Eigen v3*.
- Hallerbach, W., Hundack, C., Pouchkarev, I., and Spronk, J. (2002). A broadband vision of the development of the dax over time. Technical Report ERS-2002-87-F&A, Erasmus University Rotterdam.
- Haugen, R. A. and Baker, N. L. (1991). The efficient market inefficiency of capitalization-weighted stock portfolios. *The Journal of Portfolio Management*, 17:35 – 40.

- Haugen, R. A. and Heins, A. J. (1975). Risk and the rate of return on financial assets: Some old wine in new bottles. *Journal of Financial and Quantitative Analysis*, 10:775 – 784.
- Iyengar, S. (1988). Evaluation of normal probabilities of symmetric regions. *SIAM Journal on Scientific and Statistical Computing*, 9(3):418–423.
- Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48:65 – 91.
- Jeter, S. (2005). A handy tool for convenient error propagation analysis: A user form for error influence coefficients. In *2005 Annual Conference*, number 10.18260/1-2-14693, Portland, Oregon. ASEE Conferences. <https://peer.asee.org/14693>.
- Karney, C. F. F. (2012). Algorithms for geodesics. *Journal of Geodesy*, 87(1):43–55.
- Kurz, G. and Hanebeck, U. D. (2015). Stochastic sampling of the hyperspherical von mises–fisher distribution without rejection methods. In *2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pages 1–6.
- Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15:850 – 859.
- Ledoit, O. and Wolf, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics*, 48:3043 – 3065.
- Lee, Y. and Vempala, S. (2018). Convergence rate of Riemannian Hamiltonian Monte Carlo and faster polytope volume computation. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 1115–1121.
- Lintner, J. (1965). Security prices, risk and maximal gains from diversification. *The Journal of Finance*, 20:587 – 615.
- Liu, J., Stambaugh, R. F., and Yuan, Y. (2018). Absolving beta of volatility’s effects. *Journal of Financial Economics*, 128:1 – 15.
- Lovász, L. and Vempala, S. (2006). Hit-and-run from a corner. *SIAM Journal on Computing*, 35(4):985–1005.

- Lovász, L. and Vempala, S. (2006). Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithms. *J. Computer & System Sciences*, 72:392–417.
- Mangoubi, O. and Vishnoi, N. K. (2019). Faster polytope rounding, sampling, and volume computation via a sub-linear ball walk. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1338–1357.
- Mardia, K. V. (1975). Distribution theory for the von mises-fisher distribution and its application. In Patil, G. P., Kotz, S., and Ord, J. K., editors, *A Modern Course on Statistical Distributions in Scientific Work*, pages 113–130, Dordrecht. Springer Netherlands.
- Markowitz, H. (1992). Portfolio selection. *The Journal of Finance*, 7:77 – 91.
- Maurer, J. and Watanabe, S. (2017). Boost random number library. Software.
- Miller, M. H. and Scholes, M. (1972). Rates of return in relation to risk: A reexamination of some recent findings. pages 47 – 78.
- Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, 34:768 – 783.
- Narayanan, H. and Niyogi, P. (2006). Sampling hypersurfaces through diffusion. In *Neural Information Processing Systems (NIPS)*, page 7.
- Ortiz-Haro, J., Ha, J.-S., Driess, D., and Toussaint, M. (2021). Structured deep generative models for sampling on constraint manifolds in sequential manipulation. In *5th Annual Conference on Robot Learning (CoRL)*.
- Plyakha, Y., Uppal, R., and Vilkov, G. (2014). Equal or value weighting? implications for asset-pricing tests. *Working paper*.
- Pouchkarev, I., Spronk, J., and Trinidad, J. (2004). Dynamics of the spanish stock market through a broadband view of the IBEX 35 index. *Estudios Econom. Aplicada*, 22(1):7–21.
- Reisinger, J., Waters, A., Silverthorn, B., and Mooney, R. J. (2010). Spherical topic models. In *Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML’10*, page 903–910, Madison, WI, USA. Omnipress.

- Rosenberg, B., Reid, K., and Lanstein, R. (1985). ersuasive evidence of market inefficiency. *The Journal of Portfolio Management*, 11:9 – 16.
- Schellenberger, J. and Palsson, B. (2009). Use of randomized sampling for analysis of metabolic networks. *The Journal of biological Chemistry*, 284 9:5457–61.
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19:425 – 444.
- Smith, R. L. (1984). Efficient monte carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32(6):1296–1308.
- Somerville, P. (1998). Numerical computation of multivariate normal and multivariate-t probabilities over convex regions. *Journal of Computational and Graphical Statistics*, 7(4):529–544.
- Tacchi, M., Lasserre, J., and Henrion, D. (2020). Stokes, gibbs and volume computation of semi-algebraic sets. *arXiv preprint arXiv:2009.12139*.
- Tacchi, M., Weisser, T., Lasserre, J. B., and Henrion, D. (2022). Exploiting sparsity for semi-algebraic set volume computation. *Foundations of Computational Mathematics*, 22(1):161–209.
- van Vliet, P. and de Koning, J. (2017). High returns from low risk: A remarkable stock market paradox.
- Vempala, S. (2005). Geometric random walks: a survey. *Combinatorial and Computational Geometry*, pages 573–612.
- Venzke, A., Molzahn, D., and Chatzivasileiadis, S. (2021). Efficient creation of datasets for data-driven power system applications. *Electric Power Systems Research*, 190:106614.
- Walkshäusl, C. (2014). International low-risk investing. *The Journal of Portfolio Management*, 41:45 – 56.

Appendix

2A Membership oracle for a connected component

Lemma 5. *Given a body $\mathcal{K} = \mathcal{S}_{d-1} \cap \Delta$ and a point $p \in \mathcal{S}_{d-1}$ to decide if $p \in \mathcal{K}$ costs $O(d^2)$ operations.*

Proof. Figure 2A.12 describes the steps of the algorithm in 2D. This procedure can be generalized for any dimension d . In particular, we consider the half-line l_{op} which starts from the origin and passes through p . Let l_{op} to intersect $\partial\Delta$ at point q . Let also u one vertex of Δ that q has visual contact with, i.e., the segment defined by q and u does not intersect \mathcal{S}_{d-1} . Then, due to convexity, p belongs to the component of \mathcal{K} that corresponds to the component of graph G that contains u .

The computation of q takes $O(d^2)$ operations and the detection of the vertex u takes $O(d^2)$ operations too. □

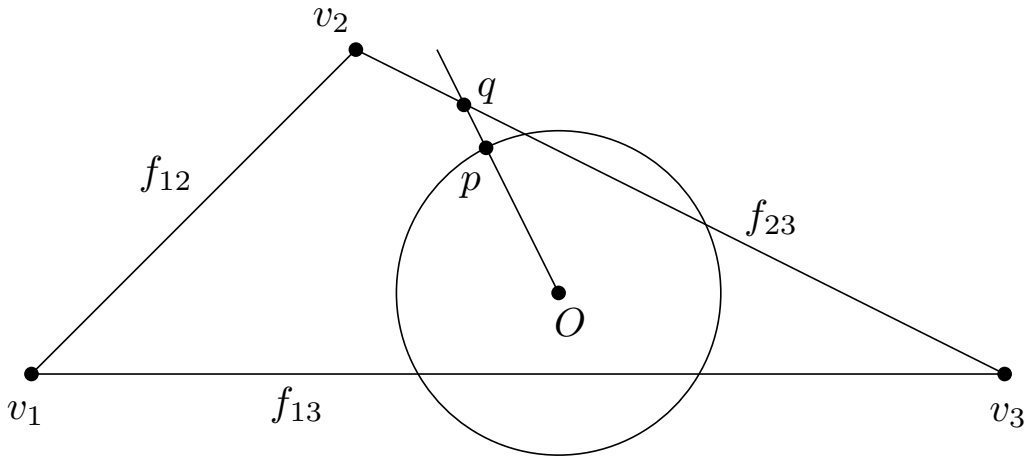


Figure 2A.12: Illustration of how the membership oracle works. In this example, there are two connected subsets of \mathcal{S}_1 . We would like to answer in which component the point p belongs. The half-line l_{op} intersects ∂P on q which lies in the facet f_{23} . The point q has visible contact with the vertex v_2 , and thus, belongs to the connected component of \mathcal{S}_1 that corresponds to the connected component $\{v_1, v_2\}$ of graph G .

2B Sampling uniformly with Great Cycle Walk

To sample uniformly a connected component of \mathcal{K} , GCW samples uniformly a point from $\ell(\theta)$ in each step. In general, to sample uniformly from a parametric curve, we have to

sample from the univariate probability density induced by the norm of the derivative of the curve. In our case, the curve is a (part of a) great circle and so GCW samples from

$$\phi(\theta) \propto \|\ell'(\theta)\|_2 = \|-p \sin \theta + v \cos \theta\|_2 = 1, \quad \theta \in [\theta^-, \theta^+], \quad (2.19)$$

which is the uniform distribution over the segment $[\theta^-, \theta^+]$. Consequently, starting from p the next Markov point is $p \cos \tilde{\theta} + v \sin \tilde{\theta}$, for $\tilde{\theta}$ sampled uniformly from $[\theta^-, \theta^+]$.

2C Proofs

Theorem 1

Proof. We build upon the results and the methodology in [Bélisle et al. \(1993\)](#). Let \mathcal{D} be a set of subsets of K and $P(p, A)$, with $p \in K$ and $A \in \mathcal{D}$, denotes the one step transition probabilities (Markov kernel) of CGW algorithm. We will prove that P is reversible with respect to π , that is

$$\int_A P(p, B) \pi(dp) = \int_B P(r, A) \pi(dr), \quad \text{for all } A, B \in \mathcal{D} \quad (2.20)$$

Then, the stationarity of π follows at once since we set $B = K$ in Equation (2.20) and we get,

$$\pi(A) = \int_K P(r, A) \pi(dr), \quad \text{for all } A \in \mathcal{D}, \quad (2.21)$$

which implies stationarity of π ([Vempala, 2005](#)).

For $p, q \in K$ with $p^T q = 0$ let,

$$\Phi(p, q) = \{\theta \in [-\pi, \pi] \mid p \cos \theta + q \sin \theta \in K\}. \quad (2.22)$$

Let f be the pdf of the constrained π on a great cycle, then, let $f_{(p,q)}$ be the pdf on $[\theta^-, \theta^+]$ defined by

$$f_{(p,q)}(\theta) = \begin{cases} \frac{f(p \cos \theta + q \sin \theta)}{\int_{\Phi(p,q)} f(p \cos t + q \sin t) dt} & \text{if } \theta \in \Phi(p, q), \\ 0, & \text{otherwise.} \end{cases} \quad (2.23)$$

Consider the random variables Q and U , where Q follows the uniform distribution over $\mathcal{S}_{d-1} \cap \mathcal{H}_p$, where $\mathcal{H}_p := \{x \in \mathbb{R}^d \mid p^T x = 0\}$, U follows the uniform distribution over $(0, 1)$, and $F_{(p,q)}$ is the cumulative distribution function (cdf) of $f_{(p,q)}$. Then, the Markov kernel

$P(p, A)$ is

$$\begin{aligned}
P(p, A) &= \Pr[F_{(p,Q)}^{-1}(U) \in A] \\
&= \int_{\mathcal{S}_{d-1} \cap \mathcal{H}_p} \frac{1}{\text{vol}(\mathcal{S}_{d-1} \cap \mathcal{H}_p)} \Pr[F_{(p,q)}^{-1}(U) \in A] dq \\
&= \int_{\mathcal{S}_{d-1} \cap \mathcal{H}_p} \frac{1}{\text{vol}(\mathcal{S}_{d-1} \cap \mathcal{H}_p)} \int_{\Phi(p,q)} \mathbf{1}_A(p \cos \theta + q \sin \theta) f_{(p,q)}(\theta) d\theta dq,
\end{aligned} \tag{2.24}$$

where $\mathbf{1}_A(\cdot)$ is the indicator function of the set A and $F_{(p,q)}^{-1}$ is the left-continuous inverse of cdf $F_{(p,q)}$. Let W the conic bundle on \mathcal{H}_p centered at p with small spatial angle $d\phi$ that defines the geodesic trajectories starting from p and have non-empty intersection with an infinitesimally small neighborhood dr in K . Then, the Markov kernel becomes $P(p, A) = \int_A g(p, r) \pi(dr)$, where

$$g(p, r) = \frac{\text{vol}(W \cap \mathcal{S}_{d-1} \cap \mathcal{H}_p)}{\text{vol}(\mathcal{S}_{d-1} \cap \mathcal{H}_p) \int_{\Phi_{pr}} p \cos t + q \sin t dt}, \tag{2.25}$$

where $\Phi_{pr} = \{\theta \in [-\pi, \pi] \mid p \cos \theta + q \sin \theta \in \mathcal{C}(pr)\}$, for well-chosen $q \in \mathcal{S}_{d-1} \cap \mathcal{H}_p$ and $\mathcal{C}(pr)$ being the part of the great cycle defined by $p, r \in \mathcal{S}_{d-1}$ inside K . Notice that $g(p, r)$ is symmetric.

Then, the left side in Equation (2.20) becomes $\int_A \int_B g(p, r) \pi(dr) \pi(dp)$ while the left side becomes $\int_B \int_A g(r, p) \pi(dp) \pi(dr)$. Reversibility then follows from Fubini's theorem and from the fact that $g(p, r)$ is symmetric. \square

Theorem 3

Proof. We build upon the results and the methodology in Gryazina and Polyak (2014). Let K_i a connected component of $K = \Delta \cap \mathcal{S}_{d-1}$. Theorem 2 in Smith (1984) proves that if the transition density $r(q|p)$ exists and is symmetric as well as it is positive for all $p, q \in K_i$ then, the uniform distribution over K_i is a unique stationary distribution, and it is achieved for any starting point $p \in K_i$. To prove convergence to the uniform distribution we consider two cases: when K_i is geodesically convex set and when K_i is geodesically non-convex set. In both cases, being at a Markov point p^j , the next Markov point p^{j+1} is obtained with positive probability with less than $\rho + 1$ reflections.

For the first case, the existence of the probability density $r(q, p)$ for any $q, p \in K_i$ is

implied when the transition probability from p to an infinitesimally small neighborhood dq is proportional to the volume of dq . Considering all possible piece-wise geodesic trajectories—defined as pieces of great cycles—that go from p to dq , take those that perform $0 \leq k \leq \rho$ reflections. With this set of trajectories there is a conic bundle on the plane H_p centered at p with with small spatial angle $d\theta$ that define these trajectories. The area of reflection can be approximated as plain region, and thus, a reflection does not change the geometry of the bundle. Then,

$$\Pr[\delta q|p] \propto \Pr[\delta\theta] \Pr[\delta L], \quad (2.26)$$

where $\Pr[d\theta]$ is the probability of choosing the spatial angle (proportional to the volume of the base of the cone) and $\Pr[dL]$ is the probability of choosing a certain trajectory length $L \in dL$. Thus, $\Pr[dq|p] \propto \text{vol}(dq)$. For a geodesically convex K_i the density $r(q|p)$ as all the points are reachable from any $p \in K_i$ with a trajectory with no reflections. The symmetry of the probability density function $r(q|p)$ follows from the uniformity of the distribution of the directions and reversibility of a billiard trajectory due to the reflection law: the angle of incidence equals the angle of reflection.

For the case of a geodesically non-convex K_i , the connectedness guarantees that starting from any point, we can reach a measurable neighborhood of any other point of K_i . Thus, there exists a piece-wise geodesic trajectory that connects any two points in K_i . Therefore, the transition probability density function $r(q|p) > 0$ for any $q, p \in K_i$. The symmetry of $r(q, p)$ holds using the same arguments as in the geodesically convex case. \square

2D Tables

USA		overall market	small defensives	small cyclicals	large defensives	large cyclicals	
Annualized Returns	Min Variance CapW-BM	0.099					
		0.085					
	Sampling-based	1st volatility level	0.097	0.087	0.109	0.067	0.101
		2nd volatility level	0.098	0.088	0.108	0.087	0.105
		3rd volatility level	0.095	0.089	0.108	0.089	0.104
		4th volatility level	0.095	0.079	0.105	0.085	0.096
		5th volatility level	0.105	0.066	0.048	0.084	0.080
	Sorting-based	1st volatility level	0.094	0.079	0.113	0.074	0.099
		2nd volatility level	0.098	0.072	0.124	0.088	0.095
		3rd volatility level	0.109	0.096	0.126	0.091	0.104
		4th volatility level	0.072	0.080	0.031	0.064	0.084
5th volatility level		0.064	0.027	0.049	0.085	0.049	

Table 2D.3

USA		overall market	small defensives	small cyclicals	large defensives	large cyclicals	
Annualized st.d.	Min Variance CapW-BM	0.138					
		0.191					
	Sampling-based	1st volatility level	0.154	0.147	0.163	0.176	0.184
		2nd volatility level	0.186	0.160	0.183	0.198	0.220
		3rd volatility level	0.210	0.173	0.215	0.221	0.251
		4th volatility level	0.230	0.202	0.269	0.240	0.291
		5th volatility level	0.268	0.259	0.363	0.294	0.375
	Sorting-based	1st volatility level	0.157	0.147	0.163	0.177	0.195
		2nd volatility level	0.176	0.160	0.179	0.201	0.225
		3rd volatility level	0.195	0.169	0.216	0.221	0.254
		4th volatility level	0.242	0.220	0.297	0.247	0.305
5th volatility level		0.282	0.266	0.353	0.290	0.360	

Table 2D.4

USA		overall market	small defensives	small cyclicals	large defensives	large cyclicals	
Annualized Sharpe Ratios	Min Variance CapW-BM	0.716					
		0.445					
	Sampling-based	1st volatility level	0.630	0.595	0.673	0.380	0.548
		2nd volatility level	0.525	0.549	0.590	0.441	0.479
		3rd volatility level	0.455	0.514	0.504	0.405	0.415
		4th volatility level	0.413	0.390	0.390	0.356	0.328
		5th volatility level	0.391	0.256	0.133	0.287	0.214
	Sorting-based	1st volatility level	0.598	0.540	0.690	0.418	0.507
		2nd volatility level	0.557	0.453	0.692	0.438	0.424
		3rd volatility level	0.559	0.569	0.583	0.412	0.411
		4th volatility level	0.299	0.364	0.106	0.260	0.275
5th volatility level		0.228	0.103	0.138	0.293	0.135	

Table 2D.5

Europe		overall market	small defensives	small cyclicals	large defensives	large cyclicals	
Annualized Returns	Min Variance	0.087					
	CapW-BM	0.059					
	Sampling-based	1st volatility level	0.081	0.092	0.115	0.062	0.060
		2nd volatility level	0.079	0.079	0.097	0.073	0.078
		3rd volatility level	0.074	0.073	0.087	0.063	0.072
		4th volatility level	0.067	0.075	0.063	0.059	0.078
		5th volatility level	0.065	0.070	0.033	0.053	0.060
	Sorting-based	1st volatility level	0.098	0.094	0.135	0.064	0.087
		2nd volatility level	0.092	0.081	0.109	0.088	0.074
		3rd volatility level	0.075	0.073	0.089	0.059	0.060
4th volatility level		0.076	0.050	0.109	0.060	0.063	
5th volatility level		0.018	0.046	-0.042	0.005	0.025	

Table 2D.6

Europe		overall market	small defensives	small cyclicals	large defensives	large cyclicals	
Annualized st.d.	Min Variance	0.159					
	CapW-BM	0.219					
	Sampling-based	1st volatility level	0.170	0.183	0.177	0.200	0.189
		2nd volatility level	0.197	0.190	0.192	0.224	0.222
		3rd volatility level	0.222	0.199	0.212	0.250	0.238
		4th volatility level	0.242	0.219	0.244	0.274	0.281
		5th volatility level	0.278	0.248	0.347	0.326	0.344
	Sorting-based	1st volatility level	0.172	0.188	0.179	0.197	0.192
		2nd volatility level	0.186	0.183	0.192	0.226	0.222
		3rd volatility level	0.205	0.197	0.211	0.253	0.240
4th volatility level		0.235	0.219	0.252	0.284	0.270	
5th volatility level		0.281	0.246	0.347	0.330	0.328	

Table 2D.7

Europe		overall market	small defensives	small cyclicals	large defensives	large cyclicals	
Annualized Sharpe Ratios	Min Variance	0.543					
	CapW-BM	0.268					
	Sampling-based	1st volatility level	0.473	0.503	0.651	0.308	0.316
		2nd volatility level	0.398	0.417	0.505	0.326	0.351
		3rd volatility level	0.333	0.367	0.410	0.253	0.302
		4th volatility level	0.279	0.344	0.258	0.217	0.275
		5th volatility level	0.233	0.284	0.096	0.163	0.174
	Sorting-based	1st volatility level	0.573	0.503	0.755	0.327	0.456
		2nd volatility level	0.496	0.442	0.568	0.389	0.335
		3rd volatility level	0.368	0.371	0.422	0.234	0.249
4th volatility level		0.325	0.229	0.430	0.213	0.234	
5th volatility level		0.065	0.187	-0.122	0.016	0.075	

Table 2D.8

USA					
Volatility level	1st	2nd	3rd	4th	5th
Model-based mode	0.051	0.042	0.071	0.057	0.061
Sampling-based average	0.051	0.041	0.071	0.056	0.060
Sorting-based (quintile portfolios)	0.002	0.010	0.004	0	0.001
Europe					
Volatility level	1st	2nd	3rd	4th	5th
Model-based mode	0.049	0.045	0.054	0.061	0.074
Sampling-based average	0.048	0.044	0.053	0.056	0.073
Sorting-based (quintile portfolios)	0	0	0	0.051	0.007

Table 2D.9: For each volatility cluster in Figure 2.5 we fit a log-concave distribution using the non-parametric model in Cule et al. (2010). We report the probability – w.r.t. the log-concave measure obtain from LogConcDEAD package (Cule et al., 2009) – of a small rectangle centered on the model-based mode, the sampling-based average and the sorting-based quintile portfolio. In all cases, the volume of the rectangle is 1% of the volume of the support of the distribution obtained by the log-concave model. See Fig. 2E.13, 2E.14 (Appendix) for a visualization.

2E Figures

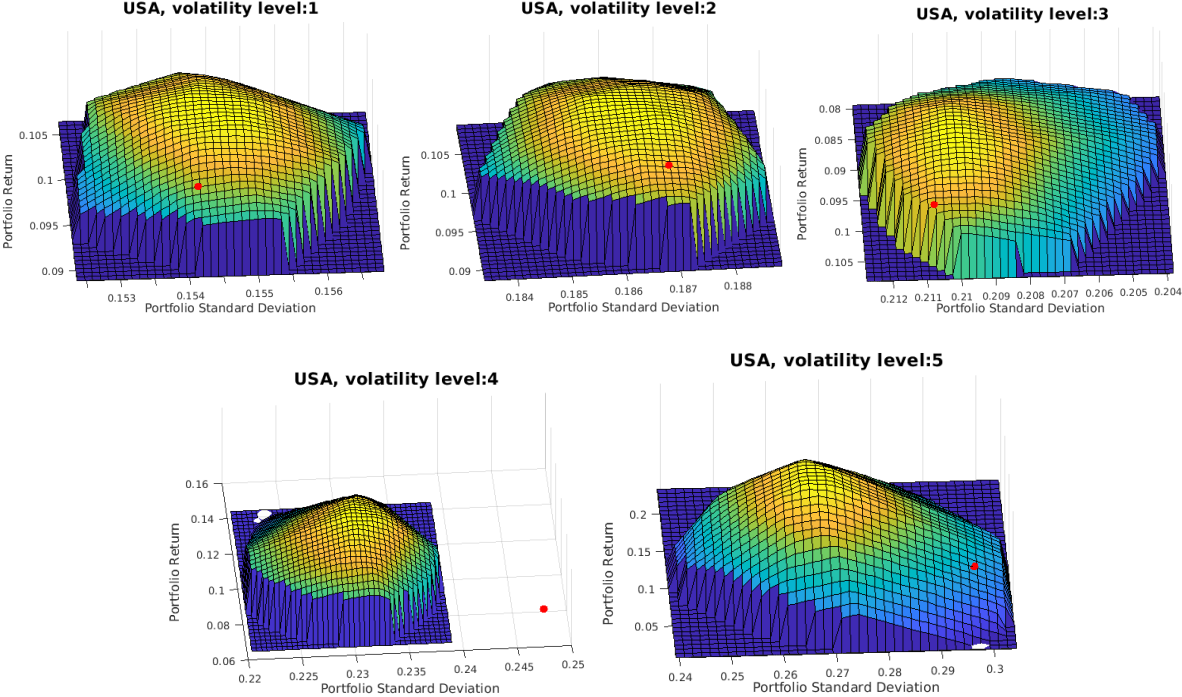


Figure 2E.13: For each volatility cluster in the left plot in Figure 2.5 —U.S. companies— we fit a log-concave distribution using the non-parametric model in Cule et al. (2010). With a red dot we report the sorting-based quintile portfolio.

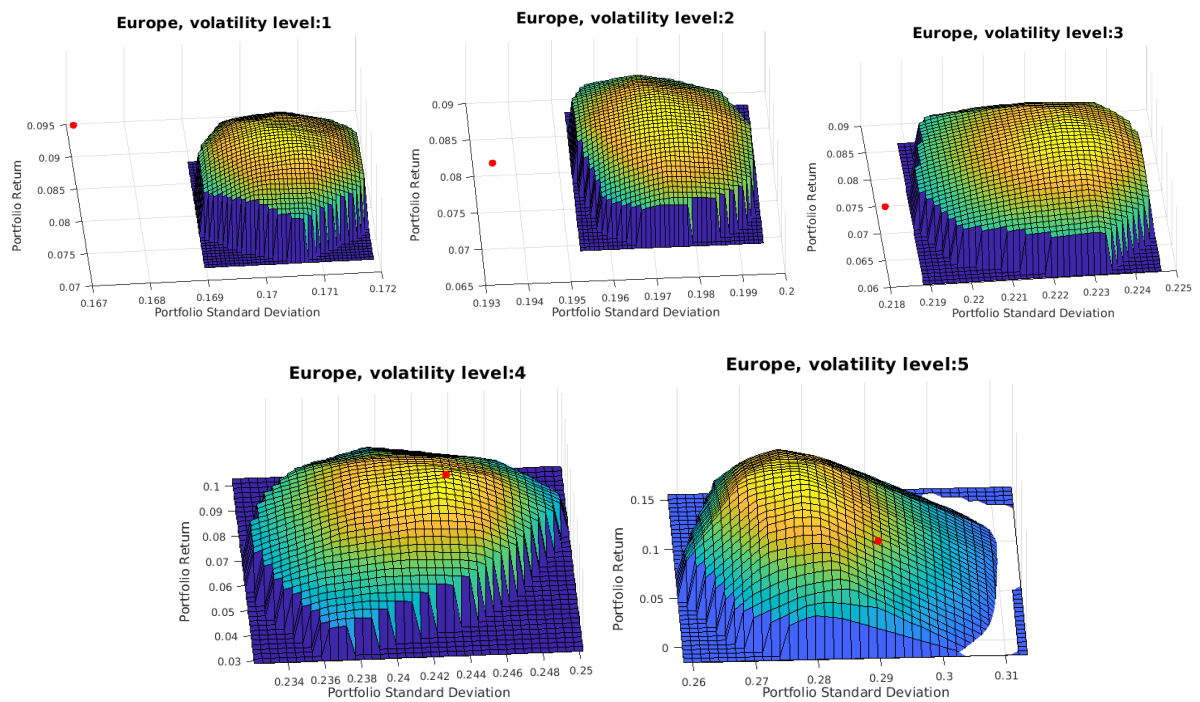


Figure 2E.14: For each volatility cluster in the right plot in Figure 2.5 —European companies— we fit a log-concave distribution using the non-parametric model in Cule et al. (2010). With a red dot we report the sorting-based quintile portfolio.

Chapter 3

Randomized Control in Performance Analysis and Empirical Asset Pricing

Cyril Bachelard, Apostolos Chalkis¹, Vissarion Fisikopoulos¹ and Elias Tsigaridas²

¹ *National & Kapodistrian University of Athens, Greece*

² *Inria Paris and IMJ-PRG, Sorbonne Université and Paris Université*

A version of the chapter, excluding the survey on geometric random walks (Section 4), will be submitted to *Quantitative Finance*. A version of the survey will be submitted to *SIAM Review*.

Abstract

The present article explores the application of randomized control techniques in empirical asset pricing and performance evaluation. It introduces geometric random walks, a class of Markov chain Monte Carlo methods, to construct flexible control groups in the form of random portfolios adhering to investor constraints. The sampling-based methods enable an exploration of the relationship between academically studied factor premia and performance in a practical setting. In an empirical application, the study assesses the potential to capture premias associated with size, value, quality, and momentum within a strongly constrained setup, exemplified by the investor guidelines of the MSCI Diversified Multifactor index. Additionally, the article highlights issues with the more traditional use case of random portfolios for drawing inferences in performance evaluation, showcasing challenges related to the intricacies of high-dimensional geometry.

JEL classification: C15, C63, C88, C58, G12.

Keywords: Empirical asset pricing, random portfolios, MCMC, geometric random walks.

3.1 Introduction

This paper examines the use of randomized control techniques in empirical asset pricing and performance evaluation from various perspectives, including economic, mathematical, and computational considerations. The investigation addresses the role of controlled chance as a tool for performance evaluation and emphasizes the substantial influence of design on inference. A novel approach to both, the design and application of randomized controls is presented. This new methodology serves to evaluate the performance and risk drivers of systematic investment strategies, i.e., portfolios that exhibit certain quantifiable characteristics by construction. Examples of such inherent characteristics include exposures towards classical factors such as size¹ (Banz, 1981), value² (Rosenberg et al., 1985), momentum³ (Jegadeesh and Titman, 1993) or quality⁴ (Asness et al., 2019).

We introduce a way to construct random controls that account for investor constraints, including but not limited to upper bounds, short-selling restrictions, risk limits, and tolerances on factor exposures, which allows to analyze the relationship between portfolio characteristic and portfolio performance within stringent investment guidelines. The suggested approach roots in geometric random walks, a special class of (continuous) Markov Chain Monte Carlo (MCMC) methods. Further, a geometric perspective will help to categorize random controls into one of three groups based on the type of constraints that are to be mapped and on the notion of randomness. Given the portfolio context, a randomized control group is hereafter referred to as a random portfolio (RP).

RPs are commonly utilized within the domain of performance evaluation, an area closely aligned with the intended application scope we are proposing. In the ensuing discourse, we shall integrate this domain. However, it will become apparent that employing RPs for purposes of performance comparisons is fraught with complexities. From our perspective, insufficient attention has been directed toward these challenges in the past. Conversely, RPs prove valuable for examining the relation between systematic portfolio features and performance within real-world scenarios. In particular, the suggested RP

¹Stocks with lower market capitalization tend to outperform stocks with a higher market capitalization in the future.

²Stocks that have a low price relative to their fundamental value, commonly tracked via accounting ratios like price to book or price to earnings outperform high-value stocks.

³Stocks that have outperformed in the past tend to exhibit strong returns going forward.

⁴Stocks which have low debt, stable earnings, consistent asset growth, and strong corporate governance, commonly identified using metrics like return to equity, debt to equity, and earnings variability.

framework based on geometric random walk enables an examination of factor effects within a constrained setup and has the capacity to unveil non-linear and asymmetric relations, if such exist, not only between characteristics and returns but also between characteristics and risks. As such, RPs constitute a novel technique for exploring asset pricing puzzles that is more versatile than classical sorting-based approaches going back to (at least) [Fama and French \(1993\)](#). Further, RPs can be designed to closely resemble a strategy of interest in all aspects, except for one particular characteristic in question, which enables an isolated analysis of the implication of that specific characteristic.

Formally, a RP is a real-valued random vector, denoted ω , having a probability density function (pdf) π with bounded support $\mathcal{K} \subset \mathbb{R}^n$. \mathcal{K} is shaped by investor-imposed constraints and geometrically forms a convex body (which we will explore in detail later on). The components of ω represent the proportions of potential monetary allocations to n investable assets, i.e., the random portfolio weights.

Our primary interest is on statistics derived from the random weights, i.e., the RP's risk, return and characteristic, where the latter typically comes in the form of a factor score. Performance statistics and characteristics are functions of the random weights and of the coefficients obtained from market observations.

Consider the following example. Let $\theta = (\mu, \sigma)$ be the vector of performance parameters, the mean μ and standard deviation σ of a RP; both are (scalar) random variables. At any point in time t , we may estimate the parameters as $\mu_t = \langle \omega, r_t \rangle$, where r_t is the vector of sample mean returns and $\sigma_t = \omega^\top \Sigma_t \omega$, where Σ_t is the sample covariance matrix of securities' returns measured at t over a specific period of stock market history.

The distribution of θ_t , estimated at a given point in time t , serves as a representation of the potential outcomes in terms of the specified performance measures, that could have been achieved over a preceding period, considering a pre-defined universe of investable assets and constraints imposed by investors. This set of counterfactual results and the corresponding probability measure provides a basis for two kinds of inferences.

An option is to employ the performance distribution of the RP as a benchmark for comparing and statistically testing performance. It has been put forward as the null distribution of performance outcomes under the assumption of no skill, thereby framing performance evaluation as a hypothesis testing problem. This is the typical use case endorsed in the literature, see e.g., [Surz \(2006\)](#), [Lisi \(2011\)](#), [Kim and Lee \(2016\)](#). In the

sequel, we demonstrate that this case subsumes the bootstrap approach of [Efron \(1979\)](#), which appears frequently in the context of performance analysis, as a particular instance of a RP. The advantage of a general RP-based analysis over simple resampling procedures is that the former offers the possibility to incorporate investor constraints. This addresses limitations associated with classical benchmark or peer group comparisons which have faced criticism for their susceptibility to biases, such as the inclusion of investment styles that do not align with the strategy of interest or for constituting portfolios that fall outside the investor’s specified constraints ([Surz, 2006](#)).

The narrative supporting the RP-based performance evaluation approach primarily relies on the following two conjectures: (i) A RP forms an obvious choice of control for performance analysis as, by design, it incorporates no penchant to any investment strategy; any portfolio structure is just as likely to occur as any other, and (ii) For (i) to hold, random weights need to follow a uniform distribution over the domain of feasible allocations. Both claims are flawed; even though they are intuitively appealing and despite their many appearances in the literature (we refer the reader to the literature review in [Section 3.1.2](#)). First, the idea that a RP is “analogous to an enumeration of all feasible portfolios” ([Kim and Lee, 2016](#)) is misleading. Although the domain is defined such as to enclose all feasible portfolios, the probability measure over the domain strongly concentrates over a thin shell of typical allocations (the high-volume area of the domain) under a uniform distribution. This concentration phenomenon arises from the intricacies of high-dimensional geometry, which we will discuss in more details in the sequel. Second, the idea that a RP “by construction incorporates no investment strategy, bias or skill” ([Dawson and Young, 2003](#)) is subject to interpretation and depends on whether one views the capitalization-weighted or the equal-weighted portfolio as the unbiased and unskilled reference (see the discussion in [Section 3.2.1](#)).

These two drawbacks, which are elaborated further in [Section 3.2.2](#), lead us to conclude that the use of RP-based statistical significance tests to evaluate the (over- or under-) performance of a particular investment strategy is not appropriate. As a purely descriptive tool, however, we consider RP-based analysis to be valuable, as exemplified in the subsequent [Section](#).

The second use case, a novel concept in our understanding, pertains to exploit a RP for detecting and assessing asset pricing anomalies within real-world scenarios. Specifically,

this entails exploring the interplay between factor exposure and the performance of portfolios that adhere to investor guidelines. At a time t , the factor exposure of the RP is given by $\gamma_{t,h} = \langle \omega, \beta_{t,h} \rangle$, where $\beta_{t,h}$ represents a cross-sectional vector of factor exposures of the firms in the investment universe. These asset-level exposures are (standardized) regression coefficients obtained by regressing the firm’s return series on the series of a factor f_h where the factor is a long-short portfolio formed by sorting stocks according to attribute h . Alternative to regression coefficients, company-specific characteristics could be used, e.g., fundamentals like a firm’s market capitalization or it’s balance sheet ratios. In accordance with the procedure of the index provider MSCI⁵, which serves as the basis for our practical application presented in the empirical part (Section 3.5), we will sometimes employ the term (factor) *score* to be generic. Ultimately, we are interested in estimating the conditional distribution of $\theta_t | \gamma_{t,h}$, which elucidates the interrelation between portfolio score and performance.

Factor scores can also form part of the constraints which define the RP, especially those not integral to the investigated investment style of interest. This grants the researcher complete control over the management and assessment of potentially confounding variables. In Section 3.5, we will employ this approach to examine the persistence of well-studied factor premia in a practical setting.

We would like to stress that our suggested use of RPs is not intended to establish causal relationships. Instead, it aims to identify statistical associations between variables and to evaluate the performance of different investment strategies under various conditions. We therefore pay attention not to confuse between causation and association. The latter is a statistical relationship between two variables, a conditional probability $\mathbb{P}(Y|X)$, while the former means that the exposure produces the effect. Our approach thus distinguishes from the typical application of randomized control in experimental designs like, for instance, pharmaceutical studies, which aim to isolate causal effects by splitting subjects into *treatment* and *control* groups.⁶ In Section 3.5, we emphasize that our use of a randomized control is not to compare treated and non-treated groups, but rather to differentiate portfolios with and without a specific characteristic.

⁵<https://www.msci.com/>

⁶As stressed in de Prado (2022), illuminating the causal nature of things requires more than just an experimental setup for testing but begins with a theory that outlines the causal mechanism(s) which are experimentally falsifiable.

Also, it is not our intention to advocate RPs as a panacea to solve asset pricing puzzles. Rather, we view them as a valuable addition to the arsenal of statistical tools for performance related questions in financial economics. However, it is important to acknowledge the complexities involved in generating RPs that accurately reflect real-world constraints. In a first instance, RPs are theoretical constructs, and the key question is whether one can effectively determine the distribution of performance measures or portfolio characteristics for inferential purposes. In most cases of practical relevance, analytical solutions are not feasible, and we must resort to numerical techniques to obtain the desired results.

3.1.1 Contributions

Our contributions within this article are fourfold. First, we clarify the inherent challenges associated with the conventional practice of utilizing RPs for performance assessment (Section 3.2). Second, we introduce an alternative use case focused on examining asset pricing anomalies within constrained scenarios. This use case is demonstrated in Sections 3.2 and 3.4. Third, we provide an extensive exposition of technical methodologies for generating RPs, covering both exact and approximate sampling-based solutions within a common geometric framework (Sections 3.3 and 3.5). Specifically, we introduce geometric random walks, a class of continuous MCMC methods tailored for high-dimensional constrained scenarios from computational geometry, as an instrument to create a randomized control which is useful to evaluate asset pricing anomalies within constrained setups. We survey all existing random walks, examining the algorithms from both theoretical and practical perspectives, and include complexity results. Lastly, we present an efficient open-source implementation in C++ with an interface in R (Sections 3.5 and 3.5.2), facilitating performance analysis via RPs with ease. With this software, assuming the reader has access to the Wharton Research Data Services (WRDS), our empirical research is fully reproducible.

3.1.2 Previous work

Ever since the claim of the economist Burton Malkiel (Malkiel, 1973) that “a blindfolded monkey throwing darts at a newspaper’s financial pages could select a portfolio that would do just as well as one carefully selected by experts” the concept of a RP has been

used to probe investment skill. Most prominently, the Wall Street Journal’s Dartboard Contest, a monthly column published by the business newspaper between 1988 and 2002, put Malkiel’s claim to the test by letting their staffers (acting as the allegoric monkeys) literally throw darts at a stock table, while investment experts picked their own stocks, always for a holding period of six months⁷. If nothing else, the game added another animal symbolism to the jargon at Wall Street, emblematic in the ongoing debate on active versus passive management and the underlying hypothesis on market efficiency. While the results of the game are not informative, the experimental design, or rather, its deficiencies contain meaningful learnings about the use of a randomized control for research in finance. Several academic studies have examined the game pointing out biases like expert’s tilt towards high risk stocks (Metcalf and Malkiel, 1994), low dividend⁸ yield stocks (Liang, 1999) and high momentum stocks (Pettengill and Clark, 2001).

From a procedural aspect, the dartboard game forms an educative example of what we call a naive RP (a formal definition follows in Section 3.3.1) and sampling from it boils down to a bootstrap exercise in the spirit of Efron (1979). Such bootstrap-like use of RPs is very common in the financial literature. One of the earliest reports we could find is Cohen and Pogue (1967) who used a bootstrap-type of RP in the analysis of mutual fund performances.

Dedicated articles to the topic of RP-based performance analysis are Surz (1994), Surz (2006), Dawson and Young (2003), Burns (2007), Lisi (2011), Billio et al. (2011), Stein (2014), Kim and Lee (2016) and Lee et al. (2018). All articles advocate the benefits of RP-based performance evaluations over traditional approaches and implicitly or explicitly promote the idea to view performance evaluation as a hypothesis test. In Surz (2006) they argue that traditional performance evaluation methods used by the finance community, namely peer group and benchmark comparisons, suffer from inevitable biases and should be replaced by Monte Carlo approaches. In Kim and Lee (2016) they point towards inefficiencies of simulation-based methods and propose a closed-form expression for the probability distribution of the Sharpe ratio of a uniformly distributed RP. Another analytical procedure based on a geometric algorithm was suggested in Cales et al. (2018)

⁷Prior to January 1990, the holding period was for one month. The extension to six months was made to alleviate a possible bias from the price pressure resulting from the announcement effect.

⁸Performances were computed on price series rather than on total return series, thus ignoring the effect of dividends and arguably incentivizing professionals to pick stocks with high growth opportunities.

and applied in [Cales et al. \(2021\)](#) and [Chalkis et al. \(2021a\)](#), again imposing a uniform distribution for a otherwise unconstrained long-only RP. To our knowledge, there is currently no prior literature offering guidance on the creation of a RP with a well-defined distribution within a constrained domain.

Concerning geometric random walk methods, which we posit fill the identified gap, their origin traces to a substantial body of literature which has explored sampling methods for generating randomized approximations to the volumes of polytopes and other convex bodies ([Dyer et al., 1991](#); [Lovász and Simonovits, 1990, 1993](#); [Cousins and Vempala, 2016](#)). Geometric random walk methods possess a distinct lineage compared to the more conventional MCMC methods widely employed in finance, particularly within Bayesian frameworks for approximating posterior distributions. The uniqueness of geometric random walk methods lies in their specialization for approximating distributions characterized by a bounded support.

Geometric random walks have numerous application, including computational biology and medical statistics, where they play an important role in metabolic network studies ([Tian et al., 2009](#); [Nielsen, 2016](#); [Haralbsdóttir et al., 2017](#)). Additionally, geometric random walks prove useful in solving convex programs ([Bertsimas and Vempala, 2004](#); [Kalai and Vempala, 2006](#)) and mixed integer convex programs ([Huang and Mehrotra, 2013](#)). Financial texts that use geometric random walk routines are sparse. In [Chiarawongse et al. \(2012\)](#) they use a geometric random walk algorithm to optimize portfolios under qualitative input. To the best of our knowledge, the only text to use geometric random walk for performance analysis is [Bachelard et al. \(2023\)](#).

The progress towards algorithms for volume computation, random sampling, and integration has developed deep connections between high-dimensional geometry and the efficiency of algorithms and shaped our understanding of convex geometry. Literature on those aspects include [Vempala \(2005\)](#), [Cousins and Vempala \(2016\)](#), [Yuansi et al. \(2018\)](#), and [Mangoubi and Vishnoi \(2019\)](#). Further important work on geometric random walk sampling include [Smith \(1984\)](#), [Lovász and Simonovits \(1990\)](#), [Lovász and Simonovits \(1993\)](#), [Kannan et al. \(1997\)](#) and [Lovász and Vempala \(2006a\)](#). References to the source literature of existing geometric random walk routines are given in [Section 3.4](#).

The remainder of this article is structured as follows. [Section 3.2](#) contextualizes random

portfolios within the realm of performance analysis and factor analysis. In Section 3.3, we delve into the problem of RP generation from a geometric perspective. Following that, Section 3.4 explores geometric random walk methods for constructing RPs tailored to address complex real-world scenarios. This Section provides an overview of possible algorithms, accompanied by descriptions of their properties and complexity, with a focus on their programmatic implementation. Subsequently, Section 3.5 presents empirical experiments using a geometric random walk-based RP to investigate the relationship between factor tilts and performance within the framework of the MSCI Diversified Multifactor index, which we consider representative of a typical setup. Finally, Section 5.5 concludes.

3.2 Use and Abuse of Random Portfolios

3.2.1 A (brief) review of performance analysis

To characterize RP's in a performance evaluation and asset pricing context, we briefly review some classical performance analysis methods.

Performance evaluation is inherently relative. The capitalization-weighted benchmark-relative perspective, arguably predominant in both industry and academia, has deep roots in economic theory (Sharpe, 1964; Lintner, 1965; Mossin, 1966), and forms the blueprint for classical performance analysis. In this context, the available tools try to identify sources of *excess* returns and to attribute them to active bets undertaken by the portfolio manager.

Holdings-based or transaction-based performance attribution tools in the line of Brinsan and Fachler (1985) (but applied to equity-only portfolios), building upon the work of Dietz (1966) and BAI (1968), are widely employed in the industry. This is so because of their explanatory power to simultaneously outline the difference in the allocation structure between a portfolio and a benchmark by grouping stocks into easily interpretable categories like countries, sectors or currencies, and to quantify the individual contributions to overall performance coming from the allocation differences among and within the pre-defined categories. The *determinants* of portfolio performance (Brinsan et al., 1986) are however not to be understood in an etiological sense. Unless the categories, according to which the attribution is done, effectively correspond to deliberate bets undertaken by the portfolio manager which characterize the strategy, allocation and selection effects do not *cause*

over- or underperformance. They can be seen as residuals, meaning that the strategy may generate these effects while pursuing other objectives. They can only identify what we may refer to as the "causa proxima", the immediate cause, while the "causa remota", the remote and perhaps indirect cause, which is encoded in the strategy and ultimately leads to the performance delta, remains unexplored. For instance, the underperformance of a minimum variance strategy following the 2022 energy crisis might be attributed to the strategy's underweighting of the energy sector. However, it is not an inherent characteristic of minimum variance strategies to avoid investments in the energy sector. Rather, the strategy may not have selected energy-related firms due to the high volatility of stocks in that industry in response to the external shock. Thus, the cause of the performance difference is only superficially due to the sector performance but roots in the underlying variance minimizing mechanism that leads to the tendency of avoiding high-volatility assets regardless of their origin or sector.

Another school of performance evaluation, initiated by [Fama and MacBeth \(1973\)](#), subsumed under the term return- or factor-based models, uses regressions to break down observed portfolio returns into a part resulting from a manager's ability to pick the best securities at a given level of risk and a part which is attributable to the dynamics of the overall market as well as to that of further risk factors which are recognized to explain security returns and are associated with a positive premium. This allows for an assessment as to whether the active performance is attributable to a particular investment style.

While the Brinson-type of performance attribution is purely descriptive, the regression setup goes beyond simple performance measurement as it allows for statistical inference in terms of significance testing of the out- or underperformance (alpha) and the loadings on other return drivers (betas) and therefore provides a basis for normative conclusions (assuming that the usual assumptions for linear regression are met). A skilled manager should be able to beat the market (factor) in a statistically significant manner after controlling for alternative betas. In this sense, skill is reserved to the active manager who is able to harvest a positive return premium not explained by known factors. Any benchmark replicating strategy, since it involves no active deviation from the capitalization-weighted allocation structure in a market, is therefore called passive and comes with performance expectations which should match the performance of a market index rather than trying to outperform it.

This view, that the average investor possesses no skill, conflicts (at least semantically) with the concept of a naive RP as the null of no skill. This is because a naive RP averages on the equal-weighted portfolio which forms a natural alternative benchmark. From a portfolio selection perspective the equal-weighted portfolio is rightly called the naive benchmark since the strategy implies no views on market developments. It results as the optimal solution in a Markowitz framework when expectations on returns, variances and correlations are considered constant cross-sectionally. Yet, from an asset pricing perspective, equal weighting provides exposure to systematic investment styles, foremost to the low-size factor, and is thus not so naive after all. The lack of consideration of the different points of view has led to unnecessary misinterpretations in the past. Unnecessary because, as we will show, RPs can easily be constructed to reflect either reference points in expectation, the naive investor who bets on all companies in a universe equally, or the average investor who accounts for the the number of shares a company has issued⁹ and the price they are traded for, i.e., firm capitalization (or actually any other reference point that could be relevant to an investor). In the forthcoming, we use the terms naive RP and basic RP to refer to either cases. A formal definition of the two concepts is given in Section 3.3 and a first educative example follows shortly.

A third and more direct approach of performance analysis comes in the form of hypothesis tests for the difference in performance measures of two strategies; typically, the portfolio in question and a benchmark. For instance, Sharpe ratio tests building upon the work of [Jobson and Korkie \(1981\)](#), the correction of [Mommel \(2003\)](#) and the extension of [Ledoit and Wolf \(2008\)](#) accounting for stylized facts of asset returns. However, at least since [Kasak and Pohlmeier \(2019\)](#) it is known that these tests suffer from low power, that is, they are unlikely to identify superior performance in the data even if there is one.

A randomized procedure for performance analysis replaces the reference point with a reference set (equipped with the empirical probability measure). Using a RP is intuitively appealing as it is a model-free method. In principle, it eliminates the need for parameter estimation, thereby avoiding dependencies on extensive time series data and assumptions about the data generating process. Moreover, it offers flexibility in evaluating different

⁹For the computation of total market value one may want to restrict calculations to shares that are free-floating, i.e., available to public trading, giving rise to a free-float adjusted market capitalization versus a non-adjusted capitalization considering all shares outstanding, i.e., also those that are held by company insiders and which are restricted to trade.

performance measures while considering transaction costs and investment constraints. However, it relies on the assumption that the distribution of no-skill performance, crucial for hypothesis system evaluation, can be objectively approximated through sampling. We contend that such is not unequivocally feasible.

3.2.2 RP-based performance evaluation - An example

Let us illustrate the challenges of performance evaluation and the use as well as the potential for misuse of RP-based inference through a somewhat absurd, yet genuine, example. Imagine a portfolio manager who constructs a portfolio based on the number of times the letter 'Z' appears in the company names of constituents within the S&P 500 index. Intuitively, this approach appears dubious and common sense would advise against entrusting this manager with our savings. However, when comparing a backtest of the manager's strategy over a 22-year period (from 2000-01-01 to 2022-12-31) to the capitalization-weighted parent index (i.e., the S&P 500), the results show an annualized outperformance of 6.2% (based on daily geometric returns) and the Sharpe ratio test of [Ledoit and Wolf \(2008\)](#)¹⁰ rejects the null hypothesis of equal Sharpe ratios at the 5% tolerance level. What is going on? Is there anything special about the letter Z? Of course not. In fact, it happens to be the case that we could have chosen *any* letter of the alphabet and the simulated out-of-sample¹¹ backtest would have shown an outperformance. We could even take the nonsense to the extreme and invert the strategies by investing in all assets except those with a particular letter in the company name. Again, all backtests outperform the benchmark. How should we make sense of this?

Our absurd example is similar to the seemingly paradoxical results presented in [Arnott et al. \(2013\)](#) who find that the arguably nonsensical inverses of sensible investment strategies, i.e., strategies built from well-founded investment beliefs, which outperform the capitalization weighted benchmark, outperform even more. The cause is readily identified by the authors by a tilt towards the size and value factors meaning that both, sensible and senseless strategies outperform for the same (unintended) reasons. In particular, even

¹⁰The test accounts for time series structures in the data by employing heteroscedasticity and autocorrelation consistent (HAC) estimates of standard error.

¹¹Our backtesting procedure ensures that at every point in time, only stocks that have been in the index at that point in time enter the portfolio selection (i.e., there is no look-ahead bias). The allocation is then held for three months, letting the weights float with total return (i.e., dividends are assumed to be reinvested) developments of the underlying stocks, until the portfolios are rebalanced.

randomly generated strategies, i.e., strategies generated from a Monte Carlo simulation (i.e., a RP approach), lead to outperformances for the same causes. All these findings enforce the authors to conclude that “value and size arise naturally in non-price-weighted strategies and constitute the main source of their return advantage” and that, therefore, “a simple performance measure becomes an unreliable gauge of skill”.

Also our inane letter strategies can be rationalized by their exposures to priced factors. For instance, when analyzed through the lens of the Fama-French-Carhart (FFC) 6-Factor model (Fama and French, 2015; Carhart, 1997) one finds that the majority of factor loadings are statistically significant. Coefficients are predominantly positive for size and value as well as for the two quality-related factors –profitability and investment– while all coefficients for momentum are negative. Market betas are, for the most part, below one and as low as 0.82 for strategy Z (and 0.78 when not controlling for other factors). However, alphas remain positive and significant for all strategies (at the 5% level) except for strategy Q. This finding appears robust to the choice of factors, as evidenced by the qualitative consistency of alphas obtained from a regression on the 13 factor themes suggested by Jensen, Kelly, and Pedersen (JKP) (Jensen et al., 2023). Conversely, betas can differ substantially between the two models. Such discrepancies are to be expected, given the methodological disparities between the FFC and JKP models in constructing factors, utilizing underlying data, and selecting control factors.¹² Nevertheless, the observed inferential disparity can be confusing, as in various instances, one model attributes outperformance to a positive exposure to a specific factor, while the other attributes the opposite, both with compelling statistical significance¹³ (especially for profitability, investment, and momentum). To arrive at a conclusive understanding, it might be necessary to forsake the convenience of the regression approach and engage in a more intricate analysis of the fundamental data related to the stocks within the letter portfolios¹⁴.

¹²We found it interesting to study the (cross-sectional) correlation between FFC loadings and JKP loadings as they turn out to be surprisingly low. Over the analyzed period and for constituents of the S&P 500, the average correlation among exposures under the two models are 0.72 for SMB and size, 0.36 for WML and momentum, merely 0.16 for HML and value and are even negative –0.08 RMW and profitability and –0.02 for CMA and investment (the correlations are computed cross-sectionally and the average is taken over time).

¹³Naturally, the regression setups should undergo thorough statistical analysis beyond a simplistic reliance on p-values, but unfortunately, such rigor is seldom practiced in reality.

¹⁴For example, strategy J shows a clearly negative portfolio size score when calculated directly on the basis of company characteristics (we used the logarithm of market capitalization, standardized

Instead, let us explore whether an RP-based approach can provide clarity. The top left chart in Figure 3.1 shows the bivariate risk-return distribution of a RP (red to yellow level sets) together with the corresponding statistics for the 26 letter strategies (grey dots), realizations of the RP (yellow small dots), the capitalization-weighted benchmark given by the S&P 500 index (black dot) and the equal-weighted portfolio of the index constituents, rebalanced monthly (blue dot).

Two immediate observations arise: Firstly, the performances of the letter strategies fall within the high-mass area of the RPs pdf, with Z and Q being outliers. Secondly, most of the mass of the RP return distribution is at levels above the return of the capitalization-weighted benchmark. The first observation suggests that any apparent performance or skill is consistent with chance and that the letter strategies could very well be instances of the RP. Analyzing the factor composition of the RP samples further reinforces this intuition. Figure 3.2 shows the distributions of factor scores of RP samples under the FFC and the JKP models. The marks above the x-axis display the corresponding exposures of the letter strategies. While the divergence of the two models poses some concern, the consistency of the letter strategy scores with the RP distribution under either model individually paints a rather clear picture, which is that the outperformance of the letter strategies may be attributed to factor exposures that happen to be prevalent in the market. By contrast, the factor structure of the capitalization-weighted benchmark can be far-off the bulk of the RP's exposure distributions (this is evident for size where the benchmark must have a negative exposure by construction).

Unlike the letter-based performances, which mostly align with the outcomes of the RP, the capitalization-weighted benchmark exhibits a notable deviation. Here is where one has to be cautious not to jump to conclusions. If we were to statistically test the benchmark return based on the RP return distribution we would conclude that the average investor is rather unskilled since most investment decisions based on chance outperformed the market. This result would be in line with the findings of [Arnott et al. \(2013\)](#) as well as [Clare et al. \(2013\)](#) who locate the return of the capitalization-weighted benchmark far out in the left tail of a RP return distribution over a 49 years (1964 - 2012) and 44 years (1968 - 2011)

cross-sectionally to have mean zero), while the size exposure is positive under both FFC and JKP. For a long-only portfolio to have an exposure to a long-short factor does not necessarily mean that this translates to the portfolio having the corresponding characteristic (a portfolio with strong size exposure may nevertheless contain very large companies).

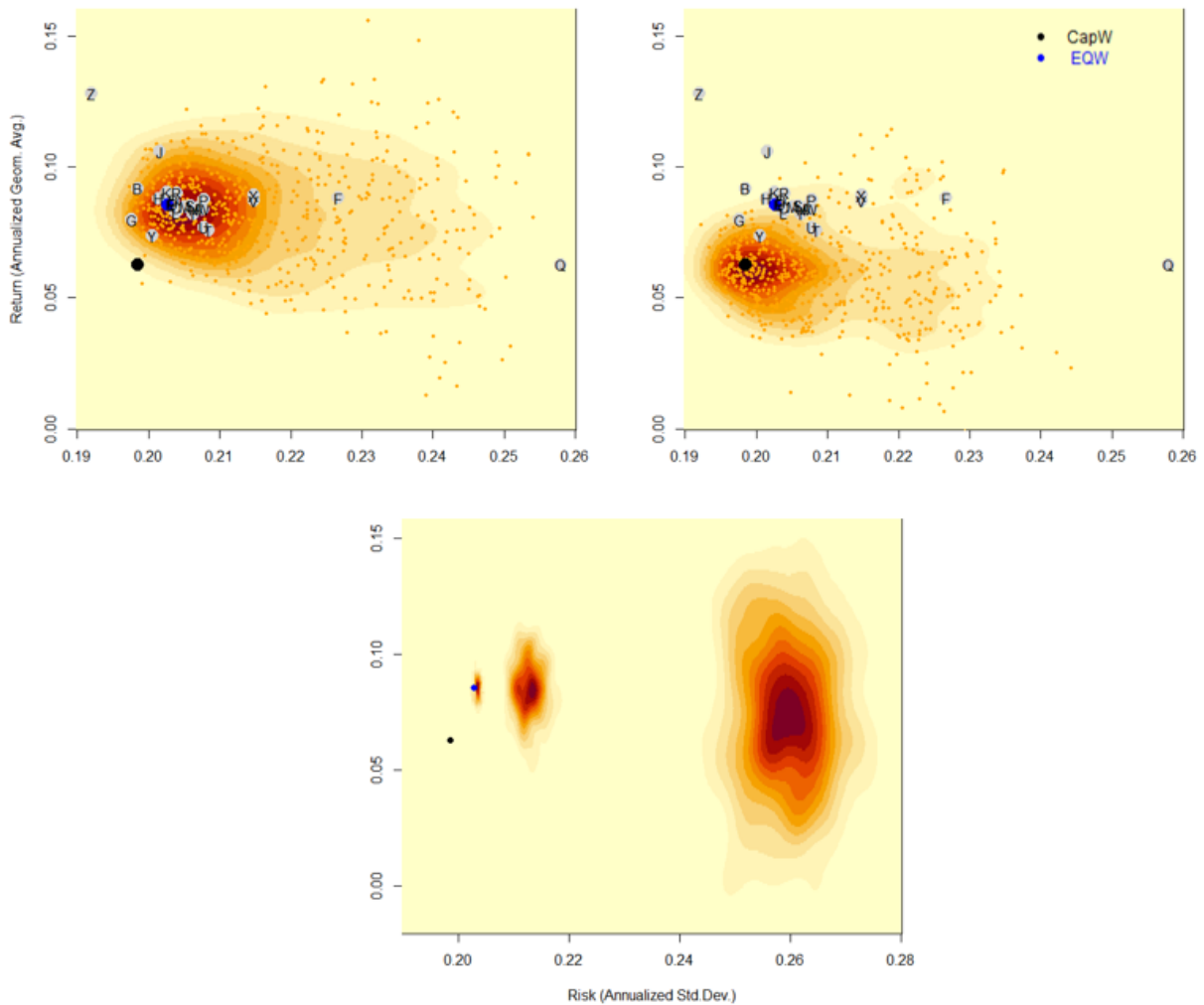


Figure 3.1: Risk-return distribution of different RPs (red to yellow level sets) together with the corresponding statistics for the 26 letter strategies (grey dots), realizations of the RPs (yellow small dots), the capitalization-weighted benchmark given by the S&P 500 index (black dot) and the equal-weighted portfolio of the index constituents. In the top charts, RPs are constructed to center on the equal-weighted (left chart) and capitalization-weighted benchmark (right chart). The bottom chart displays the distributions of three RPs created through the dartboard game approach, selecting 4, 30, and all stocks from the investable universe from right to left.

period respectively. However, one should recall Sharpe’s arithmetic of active management [Sharpe \(1991\)](#) that for every outperforming strategy there needs to be an underperforming one (relative to the capitalization-weighted benchmark). The fact that the RP outperforms the benchmark does not imply that any strategy, or even most strategies, outperform the benchmark. The outcome depends on the way the RP is constructed, circling back to the two previously mentioned issues in the RP narrative regarding the no-skill reference point and the misunderstanding that an RP covers the entirety of outcomes.

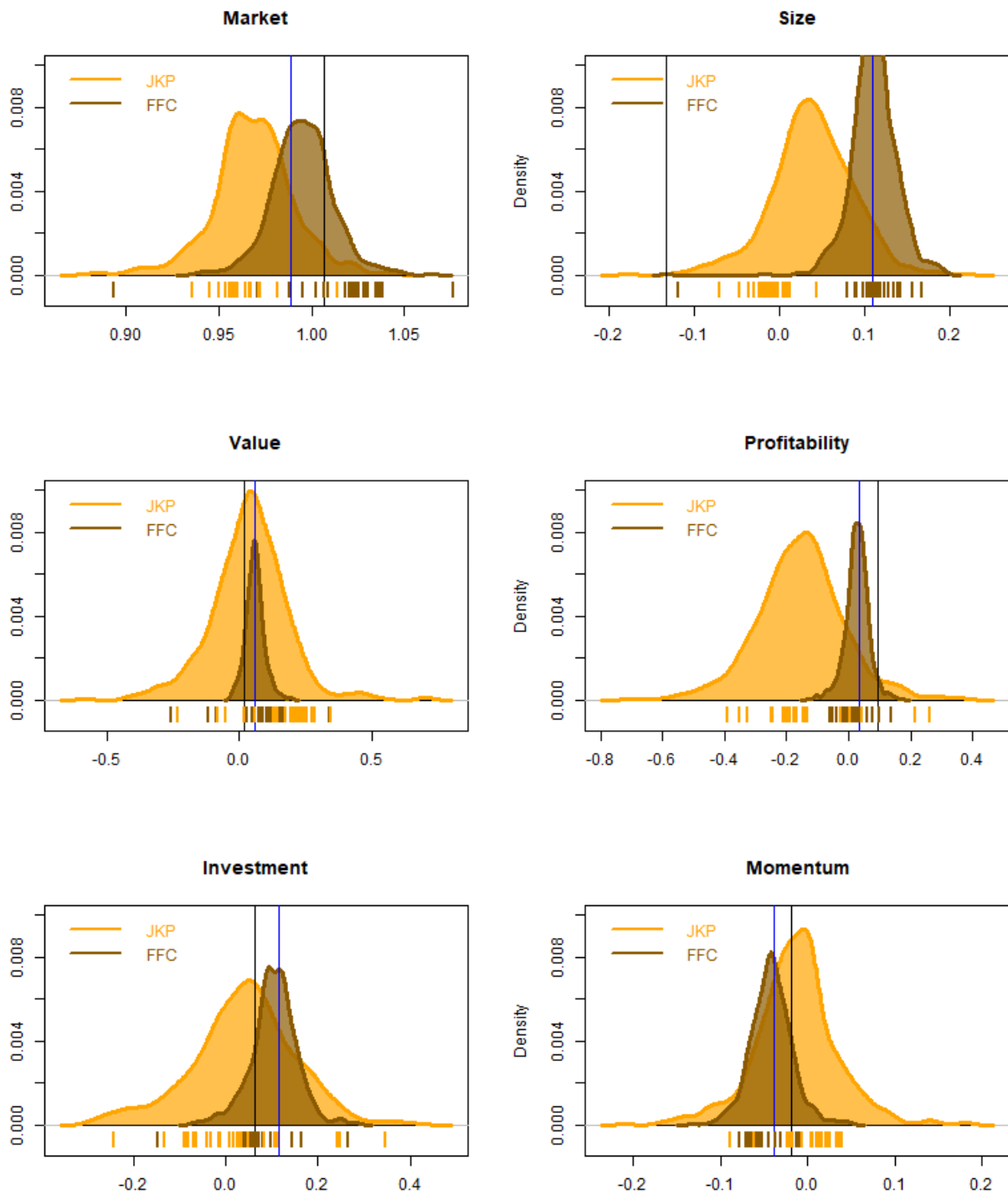


Figure 3.2: Factor exposures

Without delving too deeply into the subject of RP construction (a concise description follows in Sections 3.3 and 3.4), some qualitative aspects need addressing here. The RP depicted in the top left chart of Figure 3.1 is not constructed as having a uniform distribution over the set of feasible assets. Such an approach would yield an inadequate representation of attainable performance, as it would consist solely of homogeneous and

biased portfolios. Instead, it is crafted by convolving 26 individual RPs, each centered on an equally-weighted portfolio while considering the number of assets in the specific letter portfolio at a given rebalancing date.

As the number of assets considered decreases, the risk-return distribution of the RP widens, and the distribution's center shifts right, as depicted in the bottom chart of Figure 3.1. The chart displays three RP densities, constructed using a method commonly found in the literature. The procedure involves repeatedly sampling m stocks uniformly from the set of investable stocks (here, with a set cardinality of $n = 500$) and forming equally-weighted portfolios. This process essentially constitutes an m -out-of- n bootstrap. In the chart, we have used $m = \{500, 30, 4\}$ from left to right.

In the first case where all assets in the index are used, the density merely provides a measure of the uncertainty in the performance of the equal-weighted strategy¹⁵. This is because sampled portfolios are all very close to the equally weighed portfolio (in terms of the weights).

While the uniform measure guarantees that each portfolio has an equal probability of being sampled, the concentration phenomena occurring in higher dimension have the effect that sets of portfolios with specific characteristics appear much more likely than others. With overwhelming probability, samples will be drawn from the high-volume area of the sampling space, which forms a thin shell (the region between two concentric higher dimensional spheres of differing radii)¹⁶. Portfolios within this set show weights whose entropy and L_p norm are close to constants in a given dimension. Sparse or highly concentrated portfolios (such as the capitalization-weighted benchmark) are thus extremely unlikely to be sampled.

In cases where $m < n$, as exemplified by $m = 30$ and $m = 4$, chosen to align with the selections made by Arnott et al. (2013) and in the dartboard game, respectively, the imposed sparsity leads to a broader dispersion of portfolio weights. This translates to a more extensive distribution of performance statistics. However, the resulting distributions remain relatively compact, covering only a small area of the space of feasible solutions, reflecting

¹⁵It can be understood as an estimate of the sampling distribution of the equal-weighted portfolio's performance statistics, or, if one would sample the asset space uniformly, as the posterior density of the risk-return parameters of the equal-weighted portfolio.

¹⁶It can be shown that $\mathbb{E}(\|\omega\|_2^2) = \frac{2n-1}{n^2}$ and that therefore, with $\bar{\omega}$ denoting the equal-weighted portfolio, $\mathbb{E}(\|\omega - \bar{\omega}\|_2^2) = \frac{n^2-n}{n^3}$, which tends to zero with increasing n .

the homogeneity of sampled portfolios. Therefore, such RPs do not form acceptable control groups. Utilizing these distributions for significance testing would likely result in overly optimistic rejections of the Null hypothesis.

Adapting the RP mechanism to center on the capitalization-weighted allocation leads to the risk-return distribution in the top right chart of Figure 3.1. This distribution is more consistent with Sharpe's arithmetic, having mode close to the performance of the market index. In contrast, the profile of the naive index significantly deviates from the center of mass. We could have further expanded or contracted the risk-return distribution by adjusting the variance in the weights, essentially reverse-engineering the random control to achieve a desired statistical test result. A dangerous game. We therefore conclude that a RP, as we have presented it so far, does not provide a statistically acceptable experimental design to probe skill (or lack thereof). Nevertheless, we would argue that, with some care in the construction of the RP, it can serve as a visual aid to characterize the dispersion of performance for non-elaborate strategies (such as letter-based investing). Hopefully, the geometric perspective which we advocate in the subsequent discussion offers some transparency in this context.

The utility of employing an RP lies in the ability to analyze the relationship between portfolio performance and portfolio characteristics. Let's consider the frequency of a letter in a company's name as a stock characteristic and examine whether there is a correlation between the returns of portfolios drawn from the RP and the characteristics of those portfolios¹⁷. If exposure to a particular letter, let's say 'Z', were a rewarded characteristic, this relationship should be reflected in the cross-section of RP samples as a correlation (or more broadly, a relation) between the loading on 'Z' and performance. This (cor-)relation should be present in both RPs used above, the one centered on the equal-weighted and the one centered on the capitalization-weighted benchmark. While there are simpler ways to dismiss letter-based weighting as a viable method for portfolio construction, our point with current example is that the RP approach lends itself to an analysis of the performance impact of any systematic portfolio formation approach. In Section 3.5, we build on this approach to investigate recognized factors and their impact on performance. The analysis aims to detect genuine anomalies and reveal whether certain characteristics

¹⁷For each occurrence of a specific letter in the company name, a score of one is assigned (i.e., a company having letter 'A' appearing three times in it's name gets a score of three). The portfolio characteristics are computed as the weighted sum of stock-level characteristic-scores times the portfolio weights.

are still rewarded in the market when accounting for constraints that many investors need to adhere to.

To finalize our example, Figure 3.3 shows the scatter plot of performance versus characteristics of samples from the RP visualized in the left plot of 3.1. Unsurprisingly, there is no observable correlation for characteristic 'Z', neither between return and characteristic nor between risk and characteristic. In contrast, the right panel of Figure 3.3 illustrates the existence of a fairly strong relationship between momentum¹⁸ and performance at the portfolio level. The chart shows a non-linear (almost quadratic) pattern between risk and momentum. This implies that both, portfolios with the lowest and portfolios with the highest momentum scores exhibit high ex-post volatility, while volatility is low for intermediate-scoring portfolios. The correlation between portfolio's returns and momentum scores is positive (0.31) overall, with steeper segments at the low- and high-exposure endpoints, while the middle part appears to be relatively flat. The graph prompts us to contemplate whether the momentum anomaly would endure in a constrained setting where extreme exposures are no longer feasible due to constraints on asset weights and/or on additional factor exposures. This question is the focus of Section 3.5. Additionally, the Section will provide a detailed description of the backtesting methodology, which we have also applied here. For completeness, the two charts in the third row of Figure 3.3 shows the risk-return patten for the RP samples. The color-coding from red (low exposure) to green (high exposure) helps to identify the unstructured and the partially structured nature of the Z-characteristic versus the momentum characteristic, respectively.

¹⁸we measure momentum by the cumulative return of the portfolio over the last 12 months, omitting the last month.

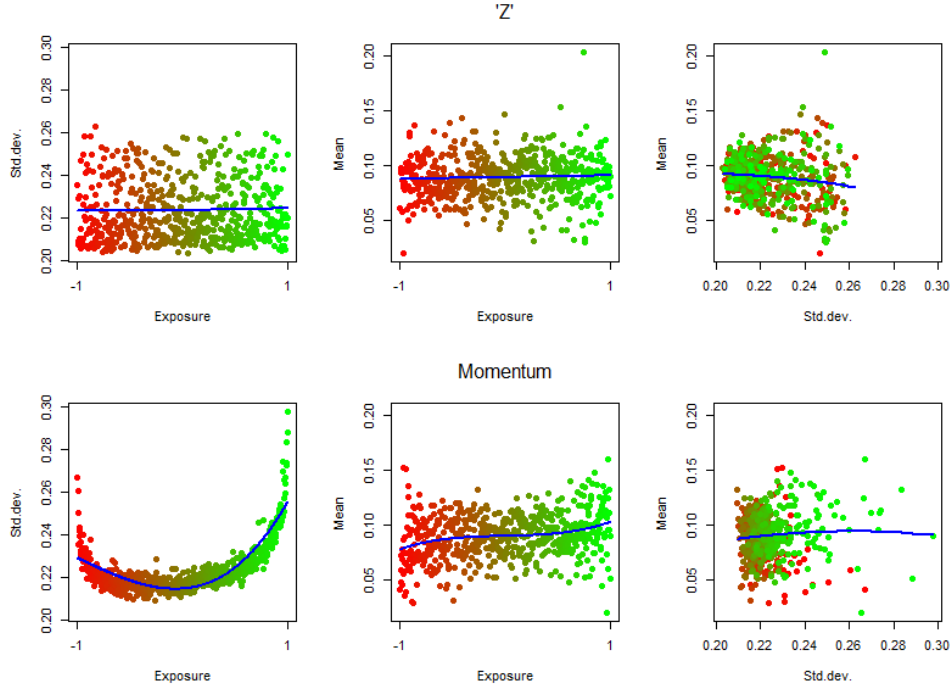


Figure 3.3: The relation between portfolio performance and exposure towards the letter 'Z' in the company names of the portfolio constituents (first row) or towards the momentum factor (second row).

3.3 Random Portfolio Generation

In this Section, we explore the design and construction of a RP. We propose a categorization of RPs into three groups of increasing complexity: naive, basic, and regularized. Among the regularized RPs, we distinguish between simply regularized (where sampling is easy) and generally regularized (where sampling is difficult). The primary objective is to gain insights into the distribution of function values of a RP, such as its return. In certain cases, these distributions have an analytic expression. However, in general, analytic solutions do not exist and sampling-based approximations are necessary.

3.3.1 Naive Random Portfolios

The arguably most intuitive case is to consider non-negative portfolio weights that sum up to one; these are the two defining properties of long-only portfolios. The geometric object which characterizes the space of long-only portfolios is the canonical simplex

$$\mathcal{S} := \{x \in \mathbb{R}^n \mid x_i \geq 0, \sum_{i=1}^n x_i = 1, i = 1, \dots, n\}. \quad (3.1)$$

In practice, mutual funds frequently encounter legal constraints mandating full investment while prohibiting short selling. Consequently, the simplex constraint embodies a legal requirement to which they must adhere. Given its prevalence within the industry, our analysis will concentrate on this typical scenario, presuming the asset space to be defined by the standard simplex or a subset thereof, which arises when additional constraints are imposed. However, the geometric random walk routines which we propose in the sequel are by no means limited to the simplex case but readily extend to accommodate long-short applications.

As for the definition of randomness, we begin with the (arguably) most intuitive case of a uniformly distributed RP, i.e., we focus on the uniform measure over the simplex. Hence, a naive RP may be given by $\omega \sim \text{U}(\mathcal{S})$, which we parametrize it through a flat¹⁹ Dirichlet model; i.e., $\omega \sim \mathcal{D}(\alpha = \mathbf{1})$, where $\mathbf{1}$ is a vector of ones. The Dirichlet pdf is given by

$$\mathcal{D}(\omega; \alpha_1, \dots, \alpha_n) \propto \prod_i \omega_i^{\alpha_i - 1}, \text{ where } \omega_i > 0, \sum_i \omega_i = 1. \quad (3.2)$$

This particular setup allows for an exact characterization of the distribution of linear statistics associated with a RP. For instance, the cumulative distribution function (cdf) of the return of a RP, denoted by μ , has an expression as a ratio of volumes. I.e.,

$$\mathbb{P}(\mu \leq \gamma) = \frac{V(\mathcal{S} \cap \mathcal{H}(r_t, \gamma))}{V(\mathcal{S})}, \quad (3.3)$$

for some scalar γ , where V denotes the volume operator and $\mathcal{H}(r_t, \gamma) = \{h \in \mathbb{R}^n | r_t^\top h \leq \gamma\}$ is the half-space induced by a hyperplane with normal r_t that is the empirical mean vector of the asset return distribution (Bachelard, 2024).

The volume in the numerator of Eq. (8) has a closed form expression (Lasserre, 2015). However, it has been observed in Cales et al. (2018, 2021) that the evaluation of this expression is numerically unstable when the dimension is not small, $n \geq 20$. To overcome this obstacle Chalkis et al Cales et al. (2018, 2021) suggest the use of an efficient and exact geometric algorithm, due to Varsi (1973) and Ali (1973), for the evaluation of the volume(s). This is the preferable implementation strategy of naive RP applications. Details on Varsi’s algorithm, including pseudocode, are given in Section 4.2.1 of Chapter 4.

¹⁹Flat means that the elements of the parameter vector of the Dirichlet distribution, α , are all equal to one.

Another form of a naive RP is given by the mechanism underlying the dartboard game. Dart throwing monkeys are easily digitalized by drawing counts $c = (c_1, \dots, c_n)$ from a Multinomial distribution $\text{Mult}(n, p_1, \dots, p_n)$ with n trials where, with uniform probabilities $p_i = 1/n, i = 1, \dots, n$, which is exactly the setup underlying the classical bootstrap of [Efron \(1979\)](#). Normalizing each count produces (RP) weights $\omega_i = c_i/n$. Unlike the Dirichlet model, the weight space is no longer given by the canonical simplex but consists of a discrete grid over the canonical simplex. As a result, the bootstrap distribution of any derived RP statistic is discrete. The grid node at the centroid of the simplex represents the equal-weighted portfolio and vertices represent single-asset portfolios. Grid nodes located on the boundary of the simplex form sparse portfolios, i.e., allocations where one or more position is exactly zero.

The definition of randomness of the weights depends on the type of resampling plan one chooses in the bootstrap procedure. For instance, the specifications of the dartboard game, i.e., how many darts, m , are thrown at a list of how many company names, n , and whether a name can be hit multiple times or not (i.e., sampling with or without replacement), define the structure of the corresponding grid. In practical applications, resampling plans often involve drawing fewer than n samples ($m < n$), corresponding to an m out of n bootstrap ([Bickel et al., 1997](#)). In this case, the induced grid is biased toward the boundaries of the simplex, resulting in a wider dispersion of the bootstrap distribution of the statistic in question compared to the classical bootstrap RP version. The continuous analogue of this approach involves using a concentration parameter in the Dirichlet distribution such that $\sum_i^n \alpha_i < n$. By introducing a mapping $\lambda(n, m) = \frac{m-1}{n}$, the standard errors of the linear statistic of interest become equivalent under both $\omega \sim \frac{1}{n} \text{Mult}(m, p_1, \dots, p_n)$ and $\omega \sim \text{Dir}(\alpha \lambda(n, m))$ ([Bachelard, 2024](#)). Notice that the flat Dirichlet model also describes a form of bootstrap version, namely the Bayesian bootstrap of [Rubin \(1981\)](#).

To summarize, a *naive* RP conforms to a classical or a Bayesian bootstrap scheme encompassing the m -out-of- n bootstrap, where m can be smaller, larger, or equal to n , and the case $\mathcal{D}(\alpha = \mathbf{1}\lambda), \lambda > 0$ in the Bayesian paradigm²⁰. We call it naive because, in expectation, it recovers the naive benchmark, i.e., $\mathbb{E}(\omega) = \bar{\omega}$, where $\bar{\omega}$ denotes the equally-weighted portfolio.

²⁰Notice that $\lambda > 1$ leads to a more concentrated distribution around the centroid of the simplex. On the other hand, $\lambda < 0$ pushes weights outwards towards the faces, edges and vertices of the simplex (that is compared to uniform density implied by the flat base case $\lambda = 1$).

3.3.2 Basic Random Portfolios

We call a RP basic if, like a naive RP, the geometric representation corresponds to a standard simplex. However, we also impose the condition that $\mathbb{E}(\omega)$ is not equal to the centroid, i.e., one cannot recover the (naive) equally-weighted portfolio in expectation. This occurs whenever some form of asymmetry is introduced in the parametrization of the Dirichlet or the Multinomial model which breaks the symmetry in the distribution of the weights. By that we mean that parameter elements cannot all be equal. For instance, setting $\alpha \propto \omega_{bm}$, where ω_{bm} is the capitalization-weighted allocation, generates a distribution of weights which has center of mass at ω_{bm} .

3.3.3 Regularized Random Portfolios

The simplex condition is typically not the only constraint that asset managers need to adhere to. When there are additional constraint, it becomes quite challenging to obtain analytical results. These additional constraints often originate from regulatory requirements and are designed to limit the risk exposure to individual security issuers or groups of issuers²¹. As a result, upper bounds are imposed on the asset weights, either individually or collectively. Geometrically, these linear restrictions correspond to halfspaces that intersect the simplex, resulting in a polytope

$$\mathcal{P} := \{x \in \mathbb{R}^n \mid Ax \leq b\}, \text{ for some } A \in \mathbb{R}^{m \times n} \text{ and } b \in \mathbb{R}^m. \quad (3.4)$$

Furthermore, managers may impose additional constraints to prevent concentration, or to align with the benchmark allocations, e.g., by setting lower and upper bounds on country or sector exposures relative to the benchmark or in terms of variation of the return difference (tracking-error), to limit transaction costs, to control portfolio characteristics (e.g., factor exposure, sustainability criteria, risk metrics), or to satisfy other specific requirements. From a sampling perspective, the mathematical characterization of constraints plays a crucial role, distinguishing between linear, quadratic, convex, non-linear, and other types. Certain risk measures like Value-at-Risk (VaR) are non-convex, meaning that, if a RP

²¹A common example is the UCITS 5/10/40 rule, which restricts single asset representation to no more than 10% of the fund's assets and limits holdings exceeding 5% to aggregate below 40% of the fund's assets.

is subject to a maximum VaR constraint, it's domain can no longer be represented by a common convex geometric body, which makes sampling extremely hard. Other measures of risk like variance or tracking error are quadratic and geometrically form ellipsoids

$$\mathcal{E} := \{x \in \mathbb{R}^n \mid x^T E x \leq c\}, \text{ for some } E \in \mathbb{R}^{n \times n}, \text{ with } E \succeq 0, \text{ and } c \in \mathbb{R}_+. \quad (3.5)$$

Sampling from (the surface) of an ellipsoid is relatively straightforward. However, when the simplex condition must also be satisfied, sampling from the intersection of the simplex with an ellipsoidal surface becomes a non-trivial task (albeit possible, as demonstrated in [Bachelard et al. \(2023\)](#)).

Under general constraints, the cdf of a RP statistic is not tractable analytically and one has to turn to numerical methods (see Section 3.4). The following special case forms an exception. Recently, [Bachelard \(2024\)](#) demonstrated that Varsi's algorithm can also be applied in the context of a shadow Dirichlet distribution ([Frigyik et al., 2010](#)), which is defined on a linearly constrained simplex. The shadow Dirichlet model allows for the consideration of linear regularizations of the weights, making it highly relevant in practical applications. If $\omega \sim \mathcal{D}(\alpha)$ and M an $n \times n$ left-stochastic matrix (i.e., each column sums to one) of full rank, then $M\omega =: \tilde{\omega} \sim \mathcal{SD}(M, \alpha)$, where \mathcal{SD} denotes the shadow Dirichlet model with pdf

$$\mathcal{SD}(\tilde{\omega}, \alpha) = \frac{1}{|\det(M)| B(\alpha)} \prod_{i=1}^n (M^{-1}\tilde{\omega})_i^{\alpha_i-1}.$$

The normalizing constant is the determinant of matrix M times the normalizing constant of the standard Dirichlet distribution, i.e., the multinomial beta function $B(\alpha) = \int_{\mathcal{S}_{n-1}} \prod_{i=1}^n \omega_i^{\alpha_i-1} d\omega = \frac{\prod_{i=1}^n \Gamma(\alpha_i)}{\Gamma(\alpha_0)}$, where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is the gamma function.

Therefore, in cases where linear constraints can be expressed using the mapping M , the distribution of linear functions of a random portfolio can be precisely obtained using Varsi's algorithm under the uniform measure. These cases are referred to as simply regularized. For all other cases where linear constraints cannot be directly accommodated, sampling methods need to be employed.

Figure 3.4 visualizes the different types of RPs for the case of $n = 3$ assets. Subplot a) shows the uniform density of a naive RP defined over the unit simplex, i.e., $\omega \sim \mathcal{D}(\mathbf{1})$.

Subplot b) also shows another naive RP since, like a), it also centers on the naive $1/n$ portfolio. However, the distribution of the weights is not uniform but follows a concentrated Dirichlet model with $\omega \sim \mathcal{D}(\mathbf{1}\lambda)$; $\lambda = 4$, giving the distribution more mass around the center. If we would have chosen the concentration parameter $\lambda < 1$, the density colors would be reverted (i.e., distributing more mass towards the boundaries of the simplex). Subplot c) shows a simple RP with $\omega \sim \mathcal{D}(\alpha)$; $\alpha = (0.5, 0.3, 0.2)$. d) visualizes the *shadow* of the Dirichlet distribution in c), i.e., $\omega \sim \mathcal{SD}(M, \alpha)$, restricted by a mononotic M which has k -th column $[0 \dots 0 \ 1/(n - k + 1) \dots 1/(n - k + 1)]^\top$. The so constrained weights satisfy the ordering $\omega_1 > \omega_2 > \dots > \omega_n$. Subplots e) and f) show instances of generally regularized RPs defined by the intersection of the simplex with a polytope in the former case, and by the intersection of the simplex with the boundary of an ellipsoid in the latter case. In both cases, the distribution of the weights is a truncated version of the basic RP in c).

3.3.4 Sampling

The task of generating RP variates, i.e., the activity of obtaining realizations of a random composition having a specified distribution over a bounded domain, requires the implementation of a sampling engine. Ideally, the sampler is exact (results in random samples with exactly the desired distribution), efficient (in terms of storage space and execution²² time), robust (the algorithm is efficient for all parameter values), and not too complex (conceptually and also with respect to practical implementation). Whether these expectations can be met depends on the concrete problem specification, i.e., on the distributional assumption and the investment constraints which define the sampling space.

To sample from a naive, a basic or a simply regularized RP one just has to sample a Dirichlet distribution with parameter vector α . To do so, it is enough to sample the marginals from a Gamma distribution with shape parameter $a = \alpha_i$ and fixed rate parameter $b = 1$ and then standardizing by the sum: $\omega_i = \frac{x_i}{\sum_{i=1}^n x_i}$, $x_i \sim \text{Gamma}(a = \alpha_i, b = 1)$. Recall that under the uniform measure, the naive as well as the simply regularized RP allow for an exact solution of the distribution function for linear statistics via Varsi's

²²Execution time has two components: Set-up time and marginal execution time. Set-up time is the time required to do some initial computing depending on the particular problem and marginal execution time is the incremental time required to generate each observation.

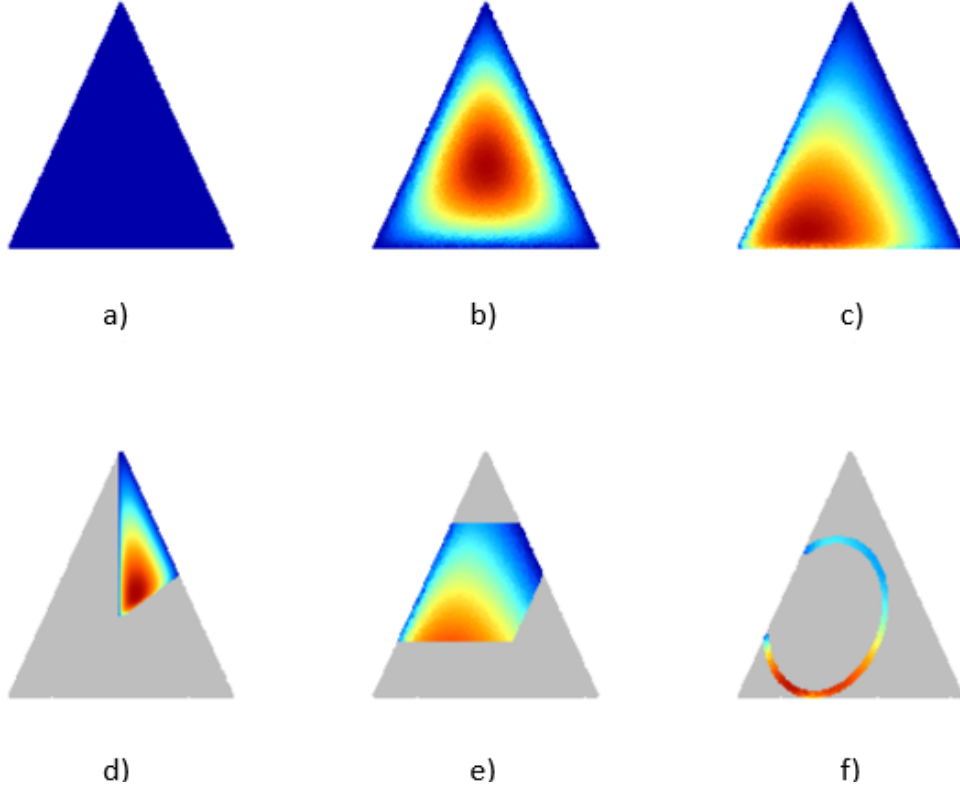


Figure 3.4: a) naive RP with a uniform density defined over the unit simplex, $\omega \sim \mathcal{D}(\mathbf{1})$. b) naive RP with a concentrated Dirichlet model, centered on the naive $1/n$ portfolio, $\omega \sim \mathcal{D}(\mathbf{1}\lambda)$, where $\lambda = 4$ (or $\lambda < 1$ for more mass towards the boundaries). c) Basic RP with weights following a Dirichlet distribution, $\omega \sim \mathcal{D}(\alpha)$, where $\alpha = (0.5, 0.3, 0.2)$. d) Shadow of the Dirichlet distribution in c), denoted as $\omega \sim \mathcal{SD}(M, \alpha)$, where the weights are restricted by a monotonic matrix M with a specific column structure. The constrained weights satisfy the ordering $\omega_1 > \omega_2 > \dots > \omega_n$. e) Generally regularized RP defined by the intersection of the simplex with a polytope, resulting in a truncated distribution of the weights. f) Generally regularized RP defined by the intersection of the simplex with the boundary of an ellipsoid, also leading to a truncated distribution of the weights.

algorithm. Therefore, sampling is actually not needed.

For a generally regularized RP, sampling is the only option. Under the uniform measure, the sampling problem forms an instance of the fundamental problem addressed in the seminal paper of [Smith \(1984\)](#): Given a bounded k -dimensional body $\mathcal{K} \subset \mathbb{R}^n$, where $k \leq n$, find a way to efficiently sample pseudo-random points $(X_1, X_2, \dots, X_d) \in \mathcal{K}$ such that $\mathbb{P}(X \in A \subset \mathcal{K}) = V(A)/V(\mathcal{K})$ with V denoting the k -dimensional content of \mathcal{K} . While [Smith \(1984\)](#) considered the general case of sampling from a generic surface, here, we restrict analysis to sampling from a \mathcal{K} which is either a polytope \mathcal{P} , the interior of an ellipsoid \mathcal{E} or their intersection. Exact uniformity can only be obtained under specific circumstances which require either the applicability of transformation, composition or

acceptance-rejection methods.

The *transformation* technique maps uniformly distributed points from a hypercube C with a smooth deterministic function T onto \mathcal{K} where $T(x)$ has to preserve uniformity. A necessary and sufficient condition for this is that the Jacobian of T is constant over all $x \in C$ (Smith, 1984). In principle, this is a highly efficient method since there exist very efficient pseudo-random number generator to sample from C . The problem is just that T is known only for a very limited class of regions S like spheres or simplices, not so though for general polytopes. Theoretically, one could partition any bounded polytope into a finite union of simplices, and then apply the transformation from the hypercube to each simplex yielding the *composite* technique. Although conceptually sound, the complexity of identifying the simplices is generally such that the approach is not tractable computationally.

The idea behind the *acceptance-rejection* technique is to first find an enclosing set $K \supset \mathcal{K}$ for which efficient sampling algorithms exist, take samples from K and either accept them if they lie within \mathcal{K} or reject them, otherwise. Accepted points will be uniformly distributed in \mathcal{K} because if a point is uniformly distributed within K , then it is conditionally uniformly distributed in \mathcal{K} given it lies in \mathcal{K} . The problem here is that even when K is chosen based on certain optimality conditions, like the smallest enclosing sphere, the number of trial points in K needed to get a point in \mathcal{K} grows, as Smith calls it 'explosively'²³.

As the dimension grows, the (only viable) solution is to sample with geometric random walks. The next Section provides a survey of existing walks that we can use to sample from a RP and, ultimately, to address the finance problems of performance and factor analysis.

3.4 Geometric Random Walks

Geometric random walk algorithms are a specific type of Markov chains that initiate from an interior point within a convex body \mathcal{K} . At each step, they transition to a neighboring point selected from a distribution dependent solely on the current position.

²³As an example, Smith (1984) shows that when \mathcal{K} is a hypercube and K is a circumscribed sphere the expected number of points generated in K needed to find one in \mathcal{K} grows from 1.5 for $k = 2$ to 10^{30} for $k = 100$.

The fundamental concept behind all geometric random walks is to generate a lengthy sequence of points, randomizing their order to render the sequence independent and identically distributed (i.i.d.) over \mathcal{K} . The complexity of the algorithms depends on its *mixing time*, i.e., the number of steps required to bound the distance between the current and the stationary distribution, and on the complexity of the basic geometric operations performed at each step of the walk; the latter is termed *per-step complexity*.

The problem of sampling from a bounded convex body \mathcal{K} is closely related to the problem of approximating the volume of \mathcal{K} . The first celebrated result is given in [Dyer et al. \(1991\)](#) where they sample approximately from the uniform distribution using a grid walk in \mathcal{K} . Since then a great effort has been devoted to geometric random walks; to mention a few seminal papers, in [Smith \(1984\)](#) they introduced the Hit-and-Run (HaR) algorithm, in [Kannan et al. \(1997\)](#) they crucially improve rounding and sampling results using Ball walk (BaW) and in [Lovász and Vempala \(2006b\)](#) they show fast mixing for HaR even when the random walk starts from a corner point in \mathcal{K} . Over the last 35 years various walks have been presented, each possessing distinct advantages and drawbacks. Some routines confine the sampling space to a polytope, while others are more generic, though often, there is a trade-off between flexibility and efficiency. Further, some walks are restricted to uniform sampling while others accommodate more general distributions, although they may need some pre-processing steps where uniform samples are required. The discriminating aspect among the various routines lies in their approach to take the next Markov step, i.e., in the choice of the direction, the step-length, the curvature of the trajectory and behavior when a boundary is hit.

Figure [3.5](#) illustrates the mechanisms of four geometric random walks which are at the basis of various descendent methods. For HaR, the logic is as follows: Start with an arbitrary point x_0 inside the convex body. Then, generate a random direction vector v with each component sampled independently from a standard normal distribution. Compute the intersections between the line defined by the point x_0 and the direction vector v and the boundary of \mathcal{K} . Finally, choose a random point on the segment defined by the two computed boundary points and repeat the process. The choice of the random distribution on the segment needs to be chosen with respect to the target distribution (for instance, uniformly, if the target is to sample uniformly distributed points). Billiard Walk (BiW) ([Gryazina and Polyak, 2014](#)) operates similarly to HaR, but it exclusively moves

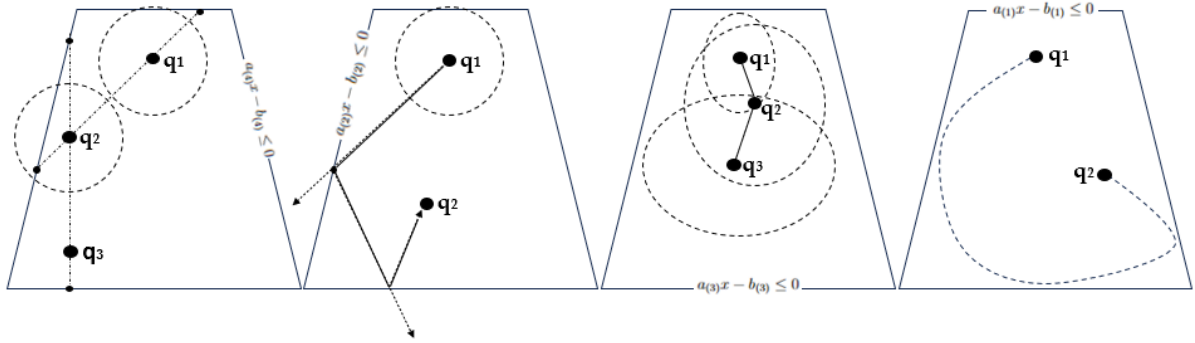


Figure 3.5: Illustration of HaR, BiW, Dikin walk, HMC (from left to right)

in the direction of v and reflects the ray upon encountering the boundary whereby the exit angle matches the entry angle. Both methods work on general \mathcal{K} , but BiW is limited to sampling from the uniform distribution. Dikin walk (Narayanan, 2016) (and other ellipsoidal procedures like Vaidya and John walk) choose their steps by sampling uniformly from an ellipsoid centered at the current point whose shape and size are determined by the shape of \mathcal{K} and the proximity of the current point to the boundary. Those algorithms are very efficient even in high-dimensional cases but only allow for uniform sampling from polytopes. Hamiltonian Monte Carlo (HMC) (Neal, 2011) is a sophisticated technique that employs Hamiltonian dynamics to enhance sampling efficiency. It picks a random velocity according to a local distribution and then walks on a Hamiltonian trajectory, i.e., a trajectory which is given by the Hamiltonian dynamics, to obtain the next Markov point. So generated proposals exhibit less of a random-walk behavior, resulting in more effective and less correlated samples. Various versions of HMC exist, all allowing sampling from general log-concave distributions, though some versions like Riemannian HMC (Kook et al., 2022) are limited to polytopes.

In the following we discuss important aspects that are common to all routines before moving to a detailed description of the individual procedures. We provide a survey of all existing routines. Readers seeking more in-depth understanding of the technicalities are encouraged to consult the source literature for a more comprehensive exploration.

Distribution support. We assume that the support of the target distribution is either a full dimensional convex polytope given by a set of linear inequalities as Eq. (3.4) or the intersection between a polytope and an ellipsoid (see Eq. (3.5)). In either case we may refer to the support as a convex body. In the case where the portfolio domain is given

also by an additional set of equalities –which is always the case in our examples since we require portfolios to be fully invested (simplex condition)– we transform it using an isometric transformation to obtain a full dimensional polytope and ellipsoid. In particular, let, for example, the set $\{x \in \mathbb{R}^n \mid Bx = 0\}$ be defined by the corresponding additional equality constraints. Then, we compute an orthonormal basis of the null space of the matrix B and we project the polytope and the ellipsoid onto the null space to obtain a full dimensional polytope and ellipsoid in the form of \mathcal{P} and \mathcal{E} respectively. Let us demonstrate an example using the Dirichlet model in Eq. 3.2 truncated in the intersection between a convex polytope and an ellipsoid, i.e.,

$$\mathcal{D}(x; \alpha_1, \dots, \alpha_n) \propto \prod_i x_i^{\alpha_i - 1}, \text{ where } x_i > 0, \sum_i x_i = 1, Ax \leq b, x^T E x \leq c, \quad (3.6)$$

$$A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, E \in \mathbb{R}^{n \times n} \text{ pos. def.}, c \in \mathbb{R}_+.$$

Notice that in this case $B = [1, \dots, 1]$. Consider the transformation $\mathcal{T}(x) := N(x - x_0)$, where $N \in \mathbb{R}^{(n-1) \times n}$ is the matrix that generates the null space of B and x_0 a feasible point in the support of \mathcal{D} . When we apply the transformation \mathcal{T} on both \mathcal{D} and its support we obtain the density,

$$[\mathcal{T} \circ \mathcal{D}](y; \alpha_1, \dots, \alpha_n) \propto \prod_i (N_{(i,:)}^T y + x_{0i})^{\alpha_i - 1}, \quad (3.7)$$

since $x = N^T y + x_0$. By replacing x in the constraint equations one could also obtain the support of the transformed density $[\mathcal{T} \circ \mathcal{D}]$. Finally, we sample from the full dimensional body in the null space and we apply the inverse transformation to obtain the sampled portfolios in the initial space.

In the sequel, we consider the case of sampling from a full dimensional convex set given by the intersection between a convex polytope and an ellipsoid,

$$\mathcal{K} = \mathcal{P} \cap \mathcal{E}. \quad (3.8)$$

Computational oracles. Each geometric random walk uses certain sub-routines called *oracles*. An oracle is an algorithm that answers a certain question that is needed for the implementation of a random walk. The oracle that every geometric random walk needs is

the *membership oracle* that answers if a given point belongs or not in \mathcal{K} . The simplest implementation of this oracle is to check the validity of the Eq. (3.4, 3.5).

Another useful sub-routine is the *boundary oracle* that computes the intersection between the boundary of \mathcal{K} , denoted $\partial\mathcal{K}$, and a ray that starts from a point in the interior of \mathcal{K} , let $\ell(t) := \{x + tv \mid x \in \mathcal{K}, v \in \mathbb{R}^n, t \in \mathbb{R}_+\}$. To compute the intersection with $\partial\mathcal{P}$ we have to solve one linear equations per facet and keep the smallest positive root. We have,

$$a_j^T(x + tv) = b_j \Rightarrow t = \frac{b_j - a_j^T x}{a_j^T v}, \quad j = [m], \quad (3.9)$$

where m is the number of facets and $a_j \in \mathbb{R}^n$ are the normal vectors defining the support hyperplane of each facet. The computations in Eq. (3.9) costs $\mathcal{O}(mn)$ operations²⁴. To compute the intersection with $\partial\mathcal{E}$ we have to solve the following equation and keep the smallest positive root,

$$(x + tv)^T E(x + tv) = c \Rightarrow \|v\|_2^2 t^2 + (2x^T E v)t + (x^T E x - c) = 0. \quad (3.10)$$

That is a second order polynomial equation for which we have a closed form to compute its roots. The computations in Eq. (3.10) costs $\mathcal{O}(n^2)$ operations. Clearly, in the case of the body \mathcal{K} we solve both Eq. (3.9, 3.10) and keep the smallest positive root t_+ . Then, $\ell(t_+) \in \partial\mathcal{K}$ is the intersection point.

Some random walks, like e.g., BiW, use a sub-routine called *reflection oracle*. This oracle computes the reflection of a ray $\ell(t)$ when the later hits the boundary of \mathcal{K} so that the reflected ray continues in \mathcal{K} . Typically, this oracle is called after the boundary oracle which computes the boundary point y of ray intersecting $\partial\mathcal{K}$. Let $s \in \mathbb{R}^n$, $\|s\|_2 = 1$ the normal vector defining the tangent hyperplane at y and $v \in \mathbb{R}^n$ the direction vector of the ray $\ell(t)$ until it hits $\partial\mathcal{K}$. Then, the update rule to obtain the direction vector of the reflected ray is the following,

$$v \leftarrow v - 2\langle v, s \rangle s. \quad (3.11)$$

Thus, the reflected direction can be computed after $\mathcal{O}(n)$ operations, given the normal vector s . When $\ell(t)$ hits a facet of \mathcal{P} the vector s is equal to the normal vector of that

²⁴Here and in the remainder of the text we use Bachmann-Landau symbols to express computational complexity (in terms of oracle evaluations and arithmetic operations) as a function of the dimension of the geometric object. Additionally, the $\tilde{\mathcal{O}}(\cdot)$ notation means that we are ignoring polylogarithmic factors.

facet, i.e., equal to the normalized row of the matrix A that corresponds to that facet. When $\ell(t)$ hits $\partial\mathcal{E}$ the vector s is equal to $Ey/\|Ey\|_2$, where y is the boundary point.

Last, certain random walks need to sample uniformly from special sets, namely from the boundary and/or the interior of the unit ball and the interior of a given ellipsoid. To sample from the boundary of the unit ball \mathcal{B}_n we sample n numbers g_1, \dots, g_n from the standard Gaussian distribution $\mathcal{N}(0, 1)$ and then the vector $q = (g_1, \dots, g_n)/\sqrt{\sum g_i^2}$ is uniformly distributed on the boundary of \mathcal{B}_n . Moreover, the point $u^{1/n}q$, where $u \in \mathbb{R}$ is uniformly distributed in $[0, 1]$, is uniformly distributed in \mathcal{B}_n . Last, the point Lx , where x is uniformly distributed in the boundary of \mathcal{B}_n and $E = L^T L$, is uniformly distributed in the ellipsoid \mathcal{E} .

Computing an interior point. To run a geometric random walk it is necessary to compute a point in the interior of the convex body we want to sample from. When the convex body is the convex polytope in Eq. (3.4) one could compute the largest ball inside \mathcal{P} , called *Chebychev ball*. Then, by definition, the center of the ball lies inside \mathcal{P} . To compute the Chebychev ball we have to solve the following linear program,

$$\begin{aligned} \max r \\ \text{subj. to: } a_i^T x + r \|a_i\|_2 \leq b_i, \end{aligned} \tag{3.12}$$

where a_i , $i = [m]$ are the rows of the matrix A in Eq. (3.4). When the convex body is the intersection between a polytope and an ellipsoid, $\mathcal{K} = \mathcal{P} \cap \mathcal{E}$, one could apply on \mathcal{K} the transformation that maps the ellipsoid \mathcal{E} to the unit ball \mathcal{B}_n and consider the intersection of the transformed polytope with \mathcal{B}_n . Then, the largest ball inside the latter body can be computed by solving the following Second-Order Cone Program (SOCP),

$$\begin{aligned} \max r \\ \text{subj. to: } a_i^T x + r \|a_i\|_2 \leq b_i \\ \|x\|_2 \leq 1 - r, \end{aligned} \tag{3.13}$$

where the pairs a_i , b_i define the facets of the transformed polytope. We can obtain a point in \mathcal{K} by applying the inverse transformation on the computed center in Eq. (3.13).

Empirical convergence to the target distribution. In order to evaluate the quality of a sample as an accurate approximation of the target distribution, several convergence diagnostics (Roy, 2020) are available like potential scale reduction factor (PSRF) (Gelman and Rubin, 1992), maximum mean discrepancy (MMD) Gretton et al. (2012) and the uniform tests introduced in Cousins (2017). For a dependent sample, a powerful diagnostic is the effective sample size (ESS). It is the number of effectively independent draws from the target distribution that the Markov chain is equivalent to. In our empirical applications, we ensure that $\text{PSRF} < 1.1$ and $\text{ESS} > 0.95n$, where n reflects the dimensionality of the problem (i.e., in general the number of assets minus the number of equality constraints).

Starting point. A crucial aspect for the efficiency of a random walk is its starting point, which needs to be a point in the interior of the convex body. When the starting point comes from a distribution close to the target distribution, then it is called a *warm start*. The mixing time analysis in the literature usually requires a warm start to bound the rate of convergence. When the starting point is a fixed or a corner point it is called a *cold start*. Regarding uniform sampling, in Lovász and Vempala (2006) they give an algorithm that computes a warm start after $\tilde{O}(n^4)$ calls to the membership oracle of the input convex body.

In practical implementations the starting point usually is a "central" point as the Chebychev or the analytical center of the polytope (Cousins and Vempala, 2016; Chalkis et al., 2023a; Kook et al., 2022). A common practice is to allow for a user defined positive integer that corresponds to the number of Markov points to be ignored before the implementation starts to store the generated points. Several practical methods have been developed to compute a good starting point (Emiris and Fisikopoulos, 2014; Chalkis et al., 2021b; Chevallier et al., 2018; Cousins and Vempala, 2016).

Roundness of the distribution. Another key aspect affecting the performance of a random walk is the roundness of the target distribution, which refers to how closely the covariance of the target distribution aligns with the identity matrix. Regarding uniform sampling, this translates to the roundness of the body, measured by the ratio R/r (Kannan et al., 1997; Lovász and Vempala, 2006b). R and r are the radii of the largest and smallest ball centered at the origin that contains, and is contained, in \mathcal{K} , respectively;

i.e., $rB_n \subseteq \mathcal{K} \subseteq RB_n$. Hence, before the actual sampling process is started, it is crucial to reduce R/r , i.e., to put \mathcal{K} in a well-rounded position, where $R/r = \tilde{\mathcal{O}}(\sqrt{n})$. This is particularly relevant for financial applications like the one we suggest in our empirical study in Section 3.5 because lower and upper bounds on asset weights are typically tight, inducing a skinny polytope.

A powerful approach to obtain well roundness is to put \mathcal{K} in *isotropic position*. In general, $\mathcal{K} \subset \mathbb{R}^n$ is in isotropic position if the uniform distribution over \mathcal{K} is in isotropic position, that is $\mathbb{E}_{X \sim \mathcal{K}}[X] = 0$ and $\mathbb{E}_{X \sim \mathcal{K}}[X^T X] = I_n$, where I_n is the $n \times n$ identity matrix. Thus, to put \mathcal{K} into isotropic position one has to generate a set of uniform points in its interior and apply to \mathcal{K} the transformation that maps the point-set to isotropic position; then iterate this procedure until \mathcal{K} is in $\mathcal{O}(1)$ -isotropic position (Cousins and Vempala, 2016; Lovász and Vempala, 2006b). In Rudelson (1999) they prove that $\mathcal{O}(\log(n))$ iterations and $\tilde{\mathcal{O}}(n)$ uniformly distributed points per iteration suffice to achieve isotropic position. Several algorithms exist which are based on this routine (Kannan et al., 1997; Lovász and Vempala, 2006; Jia et al., 2021). In Jia et al. (2021) they build upon Yuansi (2021) to provide the best algorithm so far that puts a convex body in isotropic position after $\tilde{\mathcal{O}}(n^3)$ membership oracle calls. The practical method in Chalkis et al. (2021b) brings a convex body in near-isotropic position by using BiW with multiple starting points for uniform sampling in each phase. It successfully rounds convex polytopes in a few thousand dimensions.

An alternative notion of well roundness is the John position. A convex body in John position has a sandwiching ratio of $R/r = \tilde{\mathcal{O}}(n)$, which is however worse than that of isotropic position. To put the body in John position one needs to compute the maximum volume ellipsoid (MVE) in it and apply to the body the transformation that maps the ellipsoid to the unit ball. To our knowledge there are specialized results only for the case of a convex polytope \mathcal{P} . In Nemirovski (1999) and Anstreicher (2002) they independently give an algorithm that computes the John ellipsoid of \mathcal{P} in $\tilde{\mathcal{O}}(m^{3.5})$ operations. Interestingly, in Khachiyan and Todd (1993) they provide a linear time transformation of the MVE problem by computing a minimum volume enclosing ellipsoid (MVEE) of a set of points. Thus, the algorithms in Kumar and Yildirim (2005) and in Todd and Yildirim (2007) that solve the MVEE problem can be used to compute the John ellipsoid after $\mathcal{O}(mn^3/\epsilon)$ operations. The practical method in Zhang and Gao (2003) to compute the MVE has been

Year & Authors	Type of rounding	Total cost	Polytope \mathcal{P}	Convex body \mathcal{K}
1997 Kannan et al. (1997)	Isotropic position	$\tilde{\mathcal{O}}(n^5)$ memb. calls	✓	✓
1999 Nemirovski (1999) , Anstreicher (2002)	John position	$\tilde{\mathcal{O}}(m^{3.5})$ operations	✓	✗
2003 Zhang and Gao (2003) , Haraldsdóttir et al. (2017)	*John position	??	✓	✗
2005 Kumar and Yildirim (2005) , Todd and Yildirim (2007)	John position	$\mathcal{O}(mn^3/\epsilon)$ operations	✓	✗
2006 Lovász and Vempala (2006)	Isotropic position	$\tilde{\mathcal{O}}(n^4)$ memb. calls	✓	✓
2016 Cousins and Vempala (2016)	*Isotropic position	??	✓	✓
2021 Jia et al. (2021)	Isotropic position	$\tilde{\mathcal{O}}(n^3)$ memb. calls	✓	✓
2021 Chalkis et al. (2021b)	*Isotropic position	??	✓	✓

Table 3.1: Overview of the algorithms to round a convex body.
*Practical methods.

used in [Haraldsdóttir et al. \(2017\)](#) to bring convex polytopes of thousands of dimensions in John position. However, the practical method in [Chalkis et al. \(2021b\)](#) obtains, in almost the same runtime, both a better sandwiching ratio than [Zhang and Gao \(2003\)](#) –as it brings the polytope to a near isotropic position– and a uniformly distributed sample in \mathcal{P} . Last, in [Cohen et al. \(2019\)](#) and [Song et al. \(2022\)](#) they provide algorithms to compute the John ellipsoid in the special case of a centrally symmetric convex polytope achieving near optimal performance. However, those algorithms can not be used for the purpose of rounding any convex polytope. For an overview in rounding a convex body we refer to [Table 3.1](#).

To round a log-concave distribution with density function π , R is chosen to bound the expected squared distance of the random variable from the centroid of π , i.e., $R^2 \geq \mathbb{E}_\pi(|x - z_\pi|^2)$, where z_π is the centroid. We say that a log-concave density function is well rounded if $R/r = \mathcal{O}(\sqrt{n})$, where r is the radius of the ball contained in a level set of π of constant probability.

In [Lovász and Vempala \(2003b\)](#) they provided the first algorithm to round a log-concave distribution $\pi \propto e^{-f(x)}$ after $\tilde{\mathcal{O}}(n^5)$ membership oracle calls, where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex function. They introduce an iterative algorithm that brings to isotropic position a certain level set of the density function, showing that it suffices to bring it to a near isotropic position. The algorithm can use either HaR or BaW. In [Lovász and Vempala \(2006a\)](#) they give an algorithm that uses HaR and rounds a log-concave distribution after $\tilde{\mathcal{O}}(n^{4.5})$

membership oracle calls; they generalize the algorithm in [Kalai and Vempala \(2006\)](#) that is specialized for the exponential distribution. Moreover, if a near-optimal point of π is given they provide a multi-phase algorithm that rounds π after $\tilde{O}(n^4)$ membership oracle calls. The main idea in [Kalai and Vempala \(2006\)](#); [Lovász and Vempala \(2006a\)](#) is to consider the density function,

$$\pi_\beta \propto e^{-\beta f(x)}, \quad x \in \mathcal{K}, \quad \beta \in \mathbb{R}_+ \quad (3.14)$$

where the parameter $\beta > 0$ controls the variance of π_β . They introduce a new multi-phase algorithm where in each phase they set a different value of β ; the last phase sets $\beta = 1$ and samples from π . Initially, the algorithm invokes the rounding algorithm outlined in [Lovász and Vempala \(2006\)](#) to obtain a warm start for the uniform distribution over \mathcal{K} . In the initial phase, it samples for $\beta = 0$, representing the uniform distribution. As the algorithm progresses to the j -th phase, the parameter β_j incrementally increases to ensure that a sampled point from $\pi_{\beta_{j-1}}$ serves as a warm start for π_{β_j} in accordance with the L_2 norm on density functions. Furthermore, the sample generated in the j -th phase is utilized to estimate the covariance matrix for the subsequent phase. Upon reaching $\beta = 1$, the algorithm concludes by applying a linear transformation to both π and \mathcal{K} , converting the ellipsoid defined by the approximated covariance into the unit ball. Consequently, the resulting density function is well-rounded. Lastly, it is important to state that these rounding results on constrained, log-concave density functions, highlight the importance of HaR since it can mix starting from L_2 warm starts or even from cold start, which can be exploited to reduce the number of phases needed to round π .

Best mixing times. Regarding uniform sampling from a general convex body, best mixing time is achieved by HaR and BaW, that is $\tilde{O}(n^2(R/r)^2)$ steps for a body in general position with sandwiching ratio equal to R/r . However, one could bring the body to a near isotropic position, as a preprocessing step, after $\tilde{O}(n^3)$ membership oracle calls. Then, both HaR and BaW can generate an almost uniformly distributed point after $\tilde{O}(n^2)$ steps. Then, the efficiency per almost uniformly distributed point depends on which oracle is cheaper (i.e., the membership or the boundary oracle). In the case of sampling uniformly from a bounded convex polytope \mathcal{P} , we have that the number facets $m \geq n + 1$

Year & Authors	Random walk	Mixing time	Cost per step in \mathcal{P}	Cost per step in \mathcal{K}
2006 Lovász and Vempala (2006b)	Hit-and-Run	$\tilde{\mathcal{O}}(n^2(R/r)^2)$	$\mathcal{O}(mn)$	$\mathcal{O}(mn + n^2)$
2014 Gryazina and Polyak (2014)	Billiard walk	??	$\mathcal{O}(\rho mn)$	$\mathcal{O}(\rho(mn + n^2))$
2016 Lee and Vempala (2017)	Geodesic walk	$\tilde{\mathcal{O}}(mn^{3/4})$	$\mathcal{O}(mn^{\omega-1})$	–
2016 Narayanan (2016)	Generic Dikin Walk	$\tilde{\mathcal{O}}(nm + n^3)$	$\mathcal{O}(mn^{\omega-1})$	$\mathcal{O}(mn^{\omega-1} + n^{\omega+1})$
2017 Yuansi et al. (2017)	Vaidya walk	$\tilde{\mathcal{O}}(m^{1/2}n^{3/2})$	$\mathcal{O}(mn^{\omega-1})$	–
2018 Yuansi et al. (2018)	Approximate John walk	$\tilde{\mathcal{O}}(n^{2.5})$	$\mathcal{O}(mn^{\omega-1})$	–
2018 Gustafson and Narayanan (2018)	John’s Walk	$\tilde{\mathcal{O}}(n^7)$	$\mathcal{O}(mn^4 + n^8)$	–
2018 Lee and Vempala (2018a)	Uniform Riemannian HMC	$\tilde{\mathcal{O}}(mn^{\frac{5}{3}})$	$\mathcal{O}(mn^{\omega-1})$	–
2019 Kannan et al. (1997)	Ball walk	$\tilde{\mathcal{O}}(n^2R^2/r^2)$	$\mathcal{O}(mn)$	$\mathcal{O}(mn + n^2)$
2020 Laddha et al. (2020)	Dikin walk	$\tilde{\mathcal{O}}(mn)$	$\mathcal{O}(nnz(A) + n^2)$	–
2020 Laddha et al. (2020)	Weighted Dikin Walk	$\tilde{\mathcal{O}}(n^2)$	$\mathcal{O}(mn^{\omega-1})$	–
2021 Chalkis et al. (2021b)	Multiphase Billiard walk	??	$\mathcal{O}((\rho + n)m)$	$\mathcal{O}(\rho(mn + n^2))$
Lee and Vempala (2018b), 2021 Mangoubi and Vishnoi (2019), Yuansi (2021)	\dagger Ball walk	$\tilde{\mathcal{O}}(n^2)$	$\tilde{\mathcal{O}}(m)$	$\mathcal{O}(mn + n^2)$
2022 Yuansi and Eldan (2022)	\dagger Hit-and-Run	$\tilde{\mathcal{O}}(n^2)$	$\mathcal{O}(mn)$	$\mathcal{O}(mn + n^2)$
2022 Kook et al. (2022, 2023)	Riemannian HMC	$\tilde{\mathcal{O}}(mn^3)$	$\mathcal{O}(n^{3/2})$	–
2023 Laddha and Vempala (2023)	Coordinate Hit-and-Run	$\tilde{\mathcal{O}}(n^9(R/r)^2)$	$\mathcal{O}(m)$	$\mathcal{O}(m + n)$

Table 3.2: Overview of all known geometric random walks for uniform sampling. n is the dimension of the convex body; m is the number of facets of the polytope; D is the diameter of the convex body; ω is the matrix multiplication exponent, i.e., $\omega \approx 2.37$; R is the radius of the smallest enclosing ball and r the radius of the largest ball enclosed in the convex body; ρ is the number of boundary reflections; $nnz(\cdot)$ gives the number of non-zero elements of a matrix. The symbols ?? and – respectively denote that the mixing time is unknown and that the method is not applicable to the respective body.

\dagger For a convex body in isotropic position; for Ball walk in the case of an isotropic convex polytope the (amortized) cost per step is $\tilde{\mathcal{O}}(m)$ when an $\mathcal{O}(1)$ -warm start in the interior of the polytope is given.

to guarantee boundness. Therefore, for a convex polytope in a general position, the best mixing time is given between Weighted Dikin walk (Laddha et al., 2020) and Riemannian HMC (Lee and Vempala, 2018a) while their mixing times do not depend on R/r . Since they have the same cost per step, which is the most efficient random walk depends on the number of facets m . Let $m = \mathcal{O}(n^k)$ where $k \geq 1$. For $k \leq \frac{4}{3}$ the most efficient random walk is Riemannian HMC and for $k > \frac{4}{3}$ Weighted Dikin walk achieves a smaller mixing time. However, if one brings \mathcal{P} to a near isotropic position, the most efficient option is BaW since its mixing time is $\tilde{\mathcal{O}}(n^2)$ (same with HaR and Weighted Dikin walk) and its cost per step is the smallest among all uniform samplers, i.e., $\tilde{\mathcal{O}}(m)$ steps.

Considering non-uniform sampling, there are known bounds only for the case of general log-concave distributions. The geometric random walk that achieves the best mixing time is BaW. In particular, for a log-concave in a general position BaW mixes after $\tilde{\mathcal{O}}(n^2D)$ steps where D is the diameter of the support of the target distribution (Lee and Vempala, 2017). If the target distribution is in isotropic position then BaW mixes after $\tilde{\mathcal{O}}(n^2)$ steps.

Year & Authors	Random walk	Mixing time	Cost per step in \mathcal{P}	Cost per step in \mathcal{K}
1987 Smith (1984)	Coordinate Hit-and-Run	??	$\mathcal{O}(m)$	$\mathcal{O}(m+n)$
2006 Lovász and Vempala (2006a)	Hit-and-Run	$\tilde{\mathcal{O}}(n^2 R^2 / r^2)$	$\mathcal{O}(mn)$	$\mathcal{O}(mn+n^2)$
2018 Chevallier et al. (2018)	Exact HMC with reflections	??	$\mathcal{O}(\rho\phi)$	$\mathcal{O}(\rho\phi)$
2019 Lee and Vempala (2017)	Ball walk	$\tilde{\mathcal{O}}(n^2 D)$	$\mathcal{O}(mn)$	$\mathcal{O}(mn+n^2)$
2021 Lee and Vempala (2018b) , Yuansi (2021)	[†] Ball walk	$\tilde{\mathcal{O}}(n^2)$	$\mathcal{O}(mn)$	$\mathcal{O}(mn+n^2)$
2022 Kook et al. (2022, 2023)	Riemannian HMC	$\tilde{\mathcal{O}}(mn^3)$	$\mathcal{O}(n^{3/2})$	–
2023 Chalkis et al. (2023b)	Reflective HMC	??	$\mathcal{O}(Lmn)$	$\mathcal{O}(L(mn+n^2))$

Table 3.3: Overview of all known geometric random walks for general log-concave sampling. ϕ is the cost to compute the intersection between the Hamiltonian trajectory and the boundary of the convex body. D is the diameter of the support of the distribution. L is the number of leapfrog steps in a HMC step. For the rest parameters we refer to the Table 3.2.

[†] For a log-concave distribution in isotropic position.

However, when the starting point is a corner point BaW is known to mix slowly, while HaR is known to mix rapidly from any given interior point, including corner points ([Lovász and Vempala, 2006a](#)). However, both Reflective and Riemannian HMC have shown superior performance in practice ([Chalkis et al., 2023b](#); [Kook et al., 2022](#)). For a quick overview on the random walks that allow for non-uniform sampling we refer to Table 3.3.

Geometric random walks in practice. Considering uniform sampling in practice, there are several areas of application with the most effort being devoted to sampling steady states of metabolic networks in biology ([Haraldsdóttir et al., 2017](#); [Chalkis et al., 2021b](#); [Jadebeck et al., 2020](#)) and practical volume calculation of convex polytopes ([Emiris and Fisikopoulos, 2014](#); [Cousins and Vempala, 2016](#); [Chalkis et al., 2023a](#)). Until recently the dominant paradigm for random walks was Coordinate Direction Hit-and-Run (CDHR) ([Kaufman and Smith, 1998](#)), which is a version of HaR that uses directions parallel to the axes in each step instead of random ones. The mixing time of CDHR for uniform sampling has been bounded by $\tilde{\mathcal{O}}(n^9(R/r)^2)$ ([Laddha and Vempala, 2023](#); [Narayanan and Srivastava, 2020](#)) which is the worst bound among uniform geometric random walk samplers. Interestingly, experiments ([Emiris and Fisikopoulos, 2014](#); [Haraldsdóttir et al., 2017](#)) indicate that CDHR achieves a similar mixing rate with both HaR and BaW. Thus, the faster step of CDHR compared to HaR’s and BaW’s— $\mathcal{O}(m)$ vs. $\mathcal{O}(mn)$ —is the main reason why CDHR overshadowed, until recently, all other random walks in practical computations on convex polytopes. However, in [Chalkis et al. \(2021b\)](#) a new multi-phase sampler based on

BiW proved to be much faster in practice than CDHR for uniform sampling over convex polytopes. Moreover, in [Chalkis et al. \(2023a\)](#) they again show the superior performance of BiW against HaR and CDHR by developing a new practical algorithm to approximate volume of polytopes. The Riemannian HMC in [Kook et al. \(2022\)](#) is specialized for convex polytopes and its mixing rate does not depend on the roundness of the polytope. In their experiments on metabolic networks they show that it is faster than CDHR for convex polytopes in a general position. However, there are not any experimental results that compare this Riemannian HMC sampler with the multi-phase sampler in [Chalkis et al. \(2021b\)](#) for rounded and non-rounded polytopes.

Regarding non-linear convex bodies, in [Chalkis et al. \(2022\)](#) they provide computational oracles to implement several geometric random walks for the case of spectrahedra (the feasible region of Semidefinite Programs). Their experiments show that BiW outperforms HaR, CDHR and BaW despite the fact that its cost per step is bigger comparing to all of the rest cost per steps. Regarding non-uniform sampling from non-linear convex bodies, in [Chalkis et al. \(2023b\)](#) they show that the Reflective HMC is the best performer for isotropic log-concave distributions. Last, for the case of log-concave sampling over a convex polytope again in [Chalkis et al. \(2023b\)](#) they show that the Reflective HMC outperforms both HaR and CDHR. The Riemannian HMC in [Kook et al. \(2022\)](#) appears to have a relatively small cost per step and exhibits mixing times seemingly unaffected by the roundness of the distribution. These characteristics suggest that it could serve as a competitive random walk for this particular scenario. Nevertheless, there is currently a lack of empirical experiments evaluating its performance specifically on log-concave distributions.

3.4.1 Ball walk (BaW)

BaW can be seen as a special instance of the Metropolis-Hastings algorithm ([Hastings, 1970](#)) for constrained sampling in convex domains. Initially, it was introduced for uniform sampling over convex bodies by [Kannan et al. \(1997\)](#) in the context of volume approximation and rounding of convex bodies. They prove a bound on the mixing time of $\tilde{O}(n^2 R^2 / r^2)$, which translates to $\tilde{O}(n^3)$ for an isotropic convex body. For a general isotropic log-concave distribution π constrained in a convex body the mixing time of BaW is $\tilde{O}(n^2 / \psi(\pi)^2)$,

where $\psi(\pi)$ is the *isoperimetric coefficient* of π (Kannan et al., 1997).

In Kannan et al. (1997) they bound $\psi(\pi)$ from below with $n^{1/2}$, thus resulting in the mixing time of $\tilde{\mathcal{O}}(n^3)$ for isotropic log-concave distributions. Improving this bound on $\psi(\pi)$ has been a significant open problem in convex geometry for many years, with bounding $\psi(\pi)$ by a constant being famously known as the *Kannan-Lovász-Simonovits (KLS) conjecture* (Lee and Vempala, 2018b). Furthermore, it has been demonstrated (Lee and Vempala, 2017) that, if the KLS conjecture holds true, the mixing time of BaW reduces to $\mathcal{O}(n^2)$ when the radius of the ball is $\delta = \mathcal{O}(1/\sqrt{n})$. However, for larger values of δ , the rejection probability escalates in each step, leading to an increase in the mixing time. Hence, a lower bound of $\Omega(n^2)$ remains valid for the mixing time of BaW.

In Lee and Vempala (2017) they improve the bound on $\psi(\pi)$ to $n^{1/4}$ which leads to a mixing time of $\tilde{\mathcal{O}}(n^{2.5})$ for isotropic log-concave distributions. They further extend this result and prove that the mixing time of BaW is $\tilde{\mathcal{O}}(n^2 D)$, where D represents the diameter of the convex body.

Finally, in Yuansi (2021) an almost constant bound on $\psi(\pi)$ is given, implying that as the dimension n increases, the mixing time of BaW converges to $\tilde{\mathcal{O}}(n^2)$ for isotropic log-concave distributions – which is also a tight bound on its mixing time.

An intriguing observation in Mangoubi and Vishnoi (2019) relates to the uniform distribution case and an isotropic convex polytope. They introduce an implementation of BaW steps that demand $\mathcal{O}(m)$ operations following an initial step that costs $\mathcal{O}(mn)$ operations. Consequently, the total expense for generating an almost uniformly distributed point decreases by a factor of n in the context of an isotropic convex polytope.

Step overview. In each step, BaW picks uniformly a point y from a ball of radius δ centered on the current Markov point p . Then, if the point lies outside of \mathcal{K} the random walk stays on p ; otherwise, it applies a Metropolis filter to decide moving to y or not. In theory the radius δ should be around $\tilde{\mathcal{O}}(1/\sqrt{n})$ (Kannan et al., 1997; Cousins and Vempala, 2015) to achieve the optimal mixing time for log-concave sampling. Since the implementation of BaW requires only membership oracle calls the cost per step is $\mathcal{O}(mn)$ in the general case. In the case of uniform sampling from an isotropic polytope \mathcal{P} in Mangoubi and Vishnoi (2019) they reduce the cost per step to $\tilde{\mathcal{O}}(m)$ operations by conducting the following analysis: First, for all the sampled points the starting point is

ALGORITHM 3: Ball_walk($\mathcal{K}, p, \pi, \delta$)

Input : Convex body \mathcal{K} ; point $p \in \mathcal{K}$, π probability distribution function over \mathcal{K} .

Output: A point q in \mathcal{K}

Pick a uniform random point y from the ball of radius δ centered at p ;

if $y \notin \mathcal{K}$ **then** $q \leftarrow p$ **end**

else if $y \in \mathcal{K}$ **then**

$\alpha = \min \left\{ 1, \frac{\pi(y)}{\pi(p)} \right\};$
 $u \sim \mathcal{U}[0, 1];$
if $u \leq \alpha$ **then** $q \leftarrow y$ **else** $q \leftarrow p$;

end

return q ;

always the $\mathcal{O}(1)$ -warm start given as input. Then, they exploit the property that any Markov chain whose stationary distribution is close to the uniform distribution on the polytope will spend on average at least $\frac{1}{n}$ of its time a distance $\frac{1}{n}$ from the boundary of the polytope. They compute a lower bound on the distance of the current Markov point to each one of the facets. Thus, they check a certain linear constraint $a_i^T p \leq b_i$ only if the lower bound on the distance of the point p to the i -th facet is $\mathcal{O}(\frac{1}{n})$. Finally, they prove that with at least probability $1 - \epsilon$ the BaW will not violate any linear constraint, where the small $\epsilon > 0$ is a parameter used to bound the size of BaW's step towards any facet.

Practical performance. In Cousins and Vempala (2016) they provide a BaW implementation for spherical Gaussian sampling to approximate the volume of a polytope. The mixing time in practice is –as expected– $\tilde{\mathcal{O}}(n^2)$. After an experimental evaluation, they set the radius in each step to $\delta = 4r / \max\{1, \alpha\}n$, where $\alpha = 1/2\sigma$ and r is the radius of the largest inscribed ball in the convex body. However, due to the ignored constant terms in the big-O notation, the number of steps BaW needs to mix in practice is higher than the number of steps that both HaR and CDHR need. Thus, computations carried out with HaR are faster by a constant than the same computations carried out by BaW. CDHR is faster as the dimension increases than BaW since it has a smaller cost per step. Considering uniform sampling, BaW's mixing time in practice is also $\tilde{\mathcal{O}}(n^2)$ (Chalkis and Fisikopoulos, 2021) while the radius of each step is $\delta = 4r/\sqrt{n}$.

Software. The MATLAB library `cobra` features an implementation of BaW for Gaussian sampling, built upon the findings in Cousins and Vempala (2016). This implementation has been replicated in the C++ library `volesti`. Additionally, the C++ package `PolytopeWalk` (Yuansi et al., 2020) implements the algorithms outlined in Yuansi et al. (2018). Concerning the reduced cost per step in the uniform case, as far as our knowledge extends, there hasn't been any practical implementation leveraging this result to create faster software based on the BaW uniform polytope sampler described in Mangoubi and Smith (2017).

3.4.2 Hit-and-Run (HaR)

HaR was first defined by Smith in Smith (1984) for uniform sampling from bounded convex bodies, and extended in Bélisle et al. (1993) for any density function over a (non-)convex body. HaR's mixing time has been studied extensively over the past decades. In Lovász and Vempala (2003a) and Lovász and Vempala (2006b) they show that HaR mixes after $\tilde{\mathcal{O}}(n^2(R/r)^2)$ steps for log-concave distributions (including the uniform distribution) and for both a warm and a cold start. That implies a mixing time of $\tilde{\mathcal{O}}(n^3)$ for rounded target distributions. In the case of sampling uniformly an isotropic convex body, in Yuansi and Eldan (2022) they prove that its mixing time is $\tilde{\mathcal{O}}(n^2)$ building upon the near optimal bound implied by the KLS conjecture. Several variants of HaR have been developed. For instance, in Kaufman and Smith (1998) they study optimal direction choices instead of picking a uniformly random direction in each step. It turns out that the optimal direction choice is to pick a direction from the Gaussian distribution $\mathcal{N}(0, \Sigma)$ where Σ is the covariance of the target log-concave distribution. This approach is equivalent to sampling from an isotropic distribution (Kalai and Vempala, 2006).

An important property of HaR is that it mixes from a warm start that comes from a distribution with a bounded L_2 norm from the target distribution. This allows to develop fast multi-phase polynomial randomized algorithms for several important problems in computational statistics and optimization as one would need a smaller number of phases comparing to random walks that mix from an $\mathcal{O}(1)$ warm start, i.e., BaW. Thus, HaR is the random walk of choice to round a log-concave distribution, to approximate an integral over a convex set or solving a convex problem (Lovász and Vempala, 2006a). Regarding

additional applications, a specific version of HaR, i.e. *artificial centering hit and run*, has been widely used to study constrained based models in biology (Megchelenbrink et al., 2014; Saa and Nielsen, 2016). HaR has been used for volume computation (Lovász and Vempala, 2006), sampling steady states of metabolic networks (Megchelenbrink et al., 2014; Saa and Nielsen, 2016), convex optimization (Kalai and Vempala, 2006) and other applications (Berbee et al., 1987; Haraldsdóttir et al., 2017). Its practical performance established this random walk as one of the main paradigms for constrained log-concave sampling.

Step overview. At each step the algorithm picks a uniformly distributed unit vector that defines a line through the current Markov point and computes the intersection of the line with $\partial\mathcal{K}$ by calling the boundary oracle of \mathcal{K} obtaining a segment in the interior of \mathcal{K} . Then, the next Markov point is chosen according to π_ℓ , i.e., the constraint of π over that segment. To sample from π_ℓ one could use a univariate Metropolis Hastings algorithm for $\mathcal{O}(1)$ iterations, or perform exact sampling if there is a closed form of π_ℓ at hand.

Practical performance. Extended experiments (Cousins and Vempala, 2016; Haraldsdóttir et al., 2017) have shown that HaR mixes after $\tilde{\mathcal{O}}(n^2)$ steps for well-rounded convex polytopes; this result agrees with the latest theoretical bounds on the mixing time (Yuansi and Eldan, 2022). Moreover, the experimental results in Cousins and Vempala (2016) show a convergence after $\tilde{\mathcal{O}}(n^2)$ steps for the Gaussian distribution. Interestingly, the constant in the $\tilde{\mathcal{O}}(\cdot)$ notation is many orders of magnitude smaller than in theory. For example, in Haraldsdóttir et al. (2017) they propose a step of $8n^2$.

Software. The current (competitive) implementations of HaR include, the C++ library `volesti` (Chalkis and Fisikopoulos, 2021; Chalkis et al., 2024), the MATLAB library `cobra` (Heirendt et al., 2019), the C++ library `HOPS` (Jadebeck et al., 2020), as well as the R packages `HitAndRun` (G. van Valkenhoef, T. Tervonen, 2022) and `walkr` (Yao, 2020). The most efficient ones are those in `volesti` and `HOPS` as they are both optimized C++ implementations.

ALGORITHM 4: Hit_and_Run(\mathcal{K}, p, π)

Input : Convex body $\mathcal{K} \subset \mathbb{R}^n$, point $p \in \mathcal{K}$, π probability distribution over \mathcal{K}

Output: A point $q \in \mathcal{K}$

Pick a line ℓ through p ;

return a random point q on the chord $\ell \cap \mathcal{K}$ chosen from the distribution π_ℓ ;

3.4.3 Coordinate Directions Hit-and-Run (CDHR)

CDHR can be used to sample from any distribution π over a convex body \mathcal{K} . It was first introduced for uniform sampling in [Smith \(1984\)](#).

Despite remaining relatively obscure in practical applications for many years, with a few exceptions ([Berbee et al., 1987](#)), CDHR eventually found utility in practical volume approximation algorithms ([Emiris and Fisikopoulos, 2014](#); [Cousins and Vempala, 2016](#)), for uniform and spherical Gaussian sampling. Notably, in both scenarios, CDHR demonstrated superior performance compared to HaR. This success story continued with [Haraldsdóttir et al. \(2017\)](#), where CDHR is employed to uniformly sample steady states of metabolic networks, wherein the sampled region forms a convex polytope \mathcal{P} . In this work, CDHR is combined with a preprocessing step that transforms the polytope into John position, giving rise to an algorithmic pipeline dubbed Coordinate Hit-and-Run with Rounding (CHRR). Subsequently, CDHR became the primary paradigm for sampling from polytopes until the emergence of a new implementation of BiW, as presented in [Section 3.4.4](#).

Expanding its utility to sampling from nonlinear constrained regions, [Chalkis et al. \(2022\)](#) implemented CDHR for log-concave sampling from the feasible regions of Semidefinite programs. In their experimental analysis, they demonstrate that CDHR outperforms HaR but falls short of BiW's performance in the case of uniform sampling.

The practical success of CDHR spurred increased research efforts to study its mixing time, with [Laddha and Vempala \(2023\)](#) and [Narayanan and Srivastava \(2020\)](#) providing the first upper bound guarantees for uniform sampling.

Step overview. In each step we uniformly pick an orthonormal vector $e_k \in \mathbb{R}^n$ that defines the line ℓ passing through the current Markov point p (see [Alg. 5](#)). When the sampled body is a convex polytope, the boundary oracle in the very first step costs $O(md)$ as showed above. For the step j -th step with $j > 1$, to compute the intersection between ℓ

ALGORITHM 5: Coordinate_Directions_Hit_and_Run(\mathcal{K}, p, π)

Input : Convex body $\mathcal{K} \subset \mathbb{R}^n$, point $p \in \mathcal{K}$, π probability distribution over \mathcal{K}

Output: A point $q \in \mathcal{K}$

Pick a standard orthonormal vector e_i , $i \in [n]$;

Let the line $\ell = \{x \in \mathbb{R}^n \mid x = p + te_i, t \in \mathbb{R}\}$ through p ;

return a random point q on the chord $\ell \cap \mathcal{K}$ chosen from the distribution π_ℓ ;

and $\partial\mathcal{P}$ we compute the following roots,

$$a_i^T(p + te_k) = b_i \Rightarrow t = \frac{b_i - a_i^T p}{a_i^T e_k}, \quad i \in [m], \quad (3.15)$$

and we keep the smallest positive and the largest negative. Notice that $a_i^T p$ is already computed in the $(j-1)$ -th step and $a_i^T e_k$ takes 1 operation. Sampling from the constrained π_ℓ takes $\tilde{\mathcal{O}}(1)$ operations and, moreover, we have to update just one coordinate per step. Thus, the amortized cost per step of CDHR for polytopes is $O(m)$. It has the smallest per-step cost among all the known geometric random walks together with Ball walk in the case of convex polytopes.

When the sampled body is an intersection between a polytope and an ellipsoid $\mathcal{K} = \mathcal{P} \cap \mathcal{E}$ the cost of computing the intersection of the line ℓ with the boundary of the ellipsoid \mathcal{E} is added to the cost per step of CDHR. For this we have,

$$(p + te_k)^T E(p + te_k) = c \Rightarrow E_{kk}t^2 + (2p^T E e_k)t - c = 0, \quad (3.16)$$

where the dominant computation is the dot product between p and the k -th column of the matrix E which takes $\mathcal{O}(n)$ operations. Then, the real roots in the Eq.(3.16) correspond to the intersection $\ell \cap \partial\mathcal{K}$. Thus, the amortized cost per step of CDHR for sampling \mathcal{K} is $\mathcal{O}(m+n)$.

Practical performance. Extended experiments (Cousins and Vempala, 2016; Haraldsdóttir et al., 2017) have shown that CDHR mixes after $\tilde{\mathcal{O}}(n^2)$ steps achieving the same mixing rate with HaR and BaW. In Haraldsdóttir et al. (2017) they propose a step of $8n^2$ to sample from polytopes in John position. Since CDHR achieves the same mixing rate as HaR in practice, it outperforms HaR because of its cheaper cost per step.

Software. CDHR is implemented in the C++ library `volesti`, the MATLAB library `cobra` and the C++ library `HOPS`. The most efficient ones are those in `volesti` and `HOPS` as they are both optimized C++ implementations.

3.4.4 Billiard walk (BiW)

BiW is designed for sampling exclusively from the uniform distribution over a convex (or non-convex) body. It was first introduced in [Van den Meersche et al. \(2009\)](#) and later independently in [Gryazina and Polyak \(2014\)](#) where they prove asymptotic convergence for convex and non-convex bodies. Nevertheless, its mixing time remains an open problem. In both, [Van den Meersche et al. \(2009\)](#) and [Gryazina and Polyak \(2014\)](#), the authors experimentally showcase that the mixing rate of BiW surpasses that of HaR for a limited class of bodies.

To our knowledge, considering applications that require uniform sampling, BiW has been used to address important applications in [Chalkis et al. \(2021b\)](#) and in [Chalkis et al. \(2023a\)](#). In [Chalkis et al. \(2021b\)](#) they introduce an enhancement in the cost per step of BiW for convex polytopes, demonstrating superior performance over both HaR and CDHR when sampling the largest human metabolic network known at that time. The authors also present a multi-phase sampling pipeline that enhances the mixing rate of BiW from one phase to the next, ultimately providing both a rounded body and a uniformly distributed sample in the initial body upon termination. In [Chalkis et al. \(2023a\)](#), the improved version of BiW is utilized to estimate polytope volumes in thousands of dimensions for the first time.

Step overview. BiW exploits linear trajectories and boundary reflections. In each step, BiW uniformly picks a unit direction, akin to HaR. However, being at p , the random walk then moves only forward. Upon encountering the boundary, the trajectory undergoes reflection, with a maximum of ρ reflections permitted in total. The step concludes once the trajectory has traveled a distance L , drawn from a predetermined distribution. If the trajectory exceeds ρ reflections before reaching distance L , the random walk remains at point p .

In the case of a convex polytope \mathcal{P} , in [Chalkis et al. \(2021b\)](#) they reduce the cost per step by inducting the following analysis: To compute the intersection of the ray

ALGORITHM 6: Billiard_Walk($\mathcal{K}, p, \rho, \tau$)

Input : Convex body \mathcal{K} ; point $p \in \mathcal{K}$; upper bound on the number of reflections ρ ; length of trajectory parameter τ .

Output: A point q in \mathcal{K}

Length of the trajectory $L \leftarrow -\tau \ln \eta$, $\eta \sim \mathcal{U}(0, 1)$;

Current number of reflections $i \leftarrow 0$;

Initial point of the step $p_0 \leftarrow p$;

Uniformly distributed direction $v \in \mathcal{S}_{n-1}$;

do

Segment $\ell \leftarrow \{x_i + tv_i, 0 < t \leq L\}$;

if $\partial P \cap \ell = \emptyset$ **then**

$p_{i+1} \leftarrow p_i + Lv_i$;

$i \leftarrow i + 1$;

break;

end

$p_{i+1} \leftarrow \partial \mathcal{K} \cap \ell$;

The inner vector, s , of the tangent plane on p , s.t. $\|s\| = 1$;

$L \leftarrow L - |P \cap \ell|$;

$v_{i+1} \leftarrow v_i - 2\langle v_i, s \rangle s$;

$i \leftarrow i + 1$;

while $i \leq \rho$;

$q \leftarrow p_i$;

if $i = \rho + 1$ **then**

$q \leftarrow p_0$;

end

return q ;

$\ell := \{p + tv, t > 0\}$ with $\partial \mathcal{P}$ in the very first step, BiW uses the boundary oracle of \mathcal{P} which costs $\mathcal{O}(mn)$ operations. Let p_i and v_i denote the point and the direction vector of each reflection in a BiW step, while $i > 1$. To compute the next intersection we have to solve,

$$a_j^T(x_i + tv_i) = b_j \Rightarrow a_j^T(x_{i-1} + t_{i-1}v_{i-1}) + ta_j^T(v_{i-1} - 2\langle v_{i-1}, a_k \rangle a_k) = b_j, \quad j = [m] \quad (3.17)$$

and

$$v_{i+1} = v_i - 2\langle v_i, a_l \rangle a_l, \quad (3.18)$$

where a_k, a_l are the normal vectors of the facets that ℓ hits at iteration $i - 1$ and i respectively. Index l is computed by solving all the m equations in (3.17) while each one of them is solved after $O(1)$ operations as we use the stored values from the previous iteration.

Also the inner product $\langle v_i, a_l \rangle$ in Eq. (3.18) is stored in the previous iteration. After computing all $\langle a_i, a_j \rangle$ as a preprocessing step, which takes m^2n operations, the per-step cost of BiW is $\mathcal{O}((n + \rho)m)$.

In the case of the convex body \mathcal{K} , which is a polytope intersected with an ellipsoid, the computations outlined in Eq. (3.17) cannot be completed within $\mathcal{O}(m)$ operations, as is the case with a polytope alone. In particular, when the direction v_{i-1} is reflected from the boundary of the ellipsoid, a constant number of dot products must be computed in the j -th equation, resulting in a total of $\mathcal{O}(mn)$ operations in Eq. (3.17). Since the intersection between a ray and the boundary of the ellipsoid is computed after $\mathcal{O}(n^2)$ operations, the total cost per step in this case becomes $\mathcal{O}(\rho(mn + n^2))$.

Practical performance. In Chalkis et al. (2023a) they experimentally show that BiW mixes after $\tilde{\mathcal{O}}(1)$ steps for well-rounded convex bodies. Moreover, in both, Chalkis et al. (2021b) and Chalkis et al. (2023a), they show that setting $\rho = \mathcal{O}(n)$ suffices to both achieve this mixing rate and to avoid pathological cases where the trajectory gets stuck in a corner. The latter implies that in practice, BiW achieves a $\tilde{\mathcal{O}}(1)$ mixing time while the cost per step is $\mathcal{O}(mn)$ for well rounded polytopes. In the case of the intersection between a polytope and an ellipsoid the cost per step becomes $\mathcal{O}(n^2)$.

Software. BiW is implemented in the C++ library `volesti` that provides both the improved version of BiW and the multi-phase sampling method given in Chalkis et al. (2021b). BiW is also implemented in the R package `Xsample` (Van den Meersche et al., 2009) where it was first introduced as a random walk.

3.4.5 Dikin walk

The Dikin walk method is exclusively applicable for uniformly sampling from a convex body. Here we only consider the case of a convex polytope, denoted as \mathcal{P} . Initially introduced in Kannan and Narayanan (2012), this method utilizes interior point techniques revolving around Dikin ellipsoids, which by definition are enclosed in \mathcal{P} . An intriguing property of the Dikin walk is its invariance under affine transformations of the polytope, implying that its mixing rate remains unaffected by the polytope’s shape characteristics, such as the ratio between the enclosing and inscribing radii (R/r), as observed in prior random

walk approaches. In [Kannan and Narayanan \(2012\)](#), the authors establish a mixing time bound of $\tilde{\mathcal{O}}(mn)$ steps, assuming a warm start. A shorter proof for a Gaussian variant is provided in [Sachdeva and Vishnoi \(2016\)](#). When initiated from the Chebychev center, it is demonstrated that convergence to the target distribution necessitates $\tilde{\mathcal{O}}(mn^2)$ steps. Additionally, the Dikin walk is utilized for solving linear programs within $\tilde{\mathcal{O}}(mn^2)$ steps. To define the Dikin ellipsoids they use the Hessian of the simple log-barrier of the polytope,

$$H(x) = \sum_{1 \leq i \leq m} \frac{a_i a_i^T}{(1 - a_i^T x)^2}, \quad x \in \mathcal{P}. \quad (3.19)$$

In [Laddha et al. \(2020\)](#) they introduce the notion of strong self-concordance of a barrier function. They prove that Dikin walk mixes in $\tilde{\mathcal{O}}(n\hat{\nu})$ steps from a warm start in a convex body using a strongly self-concordant barrier with symmetric self-concordance parameter $\hat{\nu}$. They show that the strong self-concordance of Lee-Sidford barrier is bounded from the standard self-concordance parameter $\nu = \mathcal{O}(n)$. Thus, they prove that Dikin walk, in a polytope \mathcal{P} , with Lee-Sidford barrier mixes after $\tilde{\mathcal{O}}(n^2)$ steps.

Finally, in [Narayanan \(2016\)](#) they extend the results in [Kannan and Narayanan \(2012\)](#) for an intersection between a polytope and a spectrahedron, that is the feasible region of a Semidefinite Program, i.e., the convex set given by a Least Matrix Inequality (LMI). They prove that Dikin walk mixes after $\tilde{\mathcal{O}}(mn + n\nu_h)$, where ν_h is the size of the matrices in LMI.

Step overview. Dikin walk uses the Hessian of a convex barrier function to define an ellipsoid centered at the current Markov point, which is contained in the body. In each step a uniformly distributed point in the ellipsoid is sampled. The new point is accepted with a probability that depends on the ratio of the volumes of the ellipsoids centered at the two points. The latter filter is applied to guarantee the convergence to the uniform distribution.

Since the implementation of each step requires to compute the volumes of two ellipsoids, which involves the computation of the determinant of a positive matrix, the cost per step of Dikin walk is $\mathcal{O}(mn^{\omega-1})$. In [Laddha et al. \(2020\)](#) they show that the step of Dikin walk with the simple log-barrier can be implemented after $\mathcal{O}(\text{nnz}(A) + n^2)$, where $\text{nnz}(\cdot)$ denotes the number of non-zero elements in matrix A .

ALGORITHM 7: DIKIN_WALK $(\mathcal{P}, p, H(\cdot), r)$

Input : polytope \mathcal{P} ; positive definite matrix $H(x)$ for each point $x \in \mathcal{P}$;
 $p \in \mathcal{P}$; radius r .

Output: A point q in \mathcal{P}

Let the ellipsoid $E_p(r) = \{x \in \mathbb{R}^n \mid (x - p)^T H(p)(x - p)\}$;

Pick y uniformly from $E_p(r)$;

$$\alpha = \min \left\{ 1, \frac{\text{vol}(E_p(r))}{\text{vol}(E_y(r))} \right\};$$

$u \sim \mathcal{U}[0, 1]$;

if $u \leq \alpha$ **then** $q \leftarrow y$ **else** $q \leftarrow p$;

return q ;

Practical performance. See Section 3.4.7

Software. Dikin walk is implemented in the C++ package `PolytopeWalk`. The same implementation is integrated in `volesti`. Last, there is an R implementation provided by the R package `walkr` (Yao, 2020).

3.4.6 Vaidya walk

Vaidya walk is closely related to Dikin walk and goes back to Vaidya (1996). It can be used to sample from the uniform distribution over a convex polytope \mathcal{P} . The point of difference is the ellipsoid it uses in each step. Instead of a Dikin ellipsoid it uses a Vaidya ellipsoid centered on the current Markov point \mathcal{P} . In Yuansi et al. (2018), they establish an upper bound of $\tilde{O}(m^{1/2}n^{3/2})$ for the mixing time of the Vaidya walk, starting from a warm start. It's worth noting that comparing the Vaidya walk with the Dikin walk reveals a trade-off between the number of facets m and the dimension n . When $m \gg n$, the Vaidya walk proves more efficient than the Dikin walk. However, for small m , the mixing rate of the Dikin walk surpasses that of the Vaidya walk.

The efficiency of the Dikin walk hinges on the representation of \mathcal{P} , particularly on the presence of redundant inequalities. Although adding redundant inequalities doesn't alter the polytope P itself, they heavily impact the Hessian of f , notably reducing the size of the Dikin ellipsoid and consequently affecting the mixing time of the Dikin walk. To address this issue and enhance the dependence of the Dikin walk on the number of facets m , the Vaidya walk introduces a solution. Its core concept involves employing a barrier with unequal weights for the logarithmic terms in the simple log-barrier function,

ALGORITHM 8: VAIDYA_WALK (P, p, r)

Input : polytope \mathcal{P} ; point $p \in \mathcal{P}$; radius r .

Output: A point q in \mathcal{P}

$C \sim$ fair coin;

if C head **then** do nothing, go to the next step;

Pick $\xi \sim \mathcal{N}(0, I_n)$;

Propose $y = p + \frac{r}{(mn)^{1/4}} V_p^{-1/2} \xi$;

if $y \notin \mathcal{P}$ **then** $q \leftarrow p$ **end**

else if $y \in \mathcal{P}$ **then**

$\alpha = \left\{ 1, \frac{g_y(p)}{g_p(y)} \right\}$;

$u \sim \mathcal{U}[0, 1]$;

if $u \leq \alpha$ **then** $q \leftarrow y$ **else** $q \leftarrow p$;

end

return q ;

or equivalently, using a matrix with unequal weights to define the ellipsoid in each step of the random walk, as outlined in Eq. (3.19). Such modifications, previously utilized in optimization contexts (Vaidya, 1996; Lee and Sidford, 2014), have shown substantial improvements.

Step overview. Instead of uniformly distributed proposals from state-dependent ellipsoids, the Vaidya walk utilizes a Gaussian proposal centered at the current Markov point p of the random walk, with the covariance being dependent on the point, given by

$$V_p = \sum_{i=1}^m (\sigma_{p,i} + \beta_\nu) \frac{a_i^T a_i}{(1 - a_i^T p)^2} \quad (3.20)$$

where $\beta_\nu = n/m$ and the scores,

$$\sigma_p := \left(\frac{a_1^T H(p) a_1}{(1 - a_1^T p)^2}, \dots, \frac{a_m^T H(p) a_m}{(1 - a_m^T p)^2} \right). \quad (3.21)$$

A single step of Vaidya walk with radius parameter $r > 0$ is given by the pseudocode of Alg. 8. With $g_p(z)$ we write the probability density function of the Gaussian with covariance V_p and centered at $p \in \mathcal{P}$,

$$g_p(z) = \sqrt{\det V_p} \left(\frac{mn}{2\pi r^2} \right)^{n/2} \exp \left(- \frac{mn}{2r^2} (z - p)^T V_p (z - p) \right). \quad (3.22)$$

Practical performance. See Section 3.4.7

Software. Vaidya walk is implemented in the C++ package PolytopeWalk which is based on the algorithms given in Yuansi et al. (2018). The same implementation is integrated in volesti.

3.4.7 John walk

John walk was first introduced for uniform sampling from a convex polytope in Gustafson and Narayanan (2018) and latter an improved version was given in Yuansi et al. (2018) introducing Approximate John walk. It goes one step further than Vaidya walk and establishes a sublinear dependence on the number of facets m in the mixing time. In Gustafson and Narayanan (2018) they compute the exact largest ellipsoid centered on the current Markov point p , namely the exact John ellipsoid. Their analysis gives a mixing time of $\tilde{O}(n^7)$ steps. Interestingly, in Yuansi et al. (2018) they employ the analysis in Lee and Sidford (2014) that use John’s weights to improve the performance of interior point methods to solve linear programs by employing approximate John ellipsoids. In Yuansi et al. (2018) they integrate that analysis to improve the mixing time of John walk and prove a bound of $\tilde{O}(n^{2.5})$ steps which depends logarithmic on the number of facets m .

Step overview. In each step being at a point $p \in \mathcal{P}$ John walk approximates the largest inscribed ellipsoid in \mathcal{P} centered at p and employs the covariance matrix that defines the ellipsoid to compute a Gaussian proposal. The inverse covariance matrix underlying the John walk is given by,

$$J_p = \sum_{i=1}^m \zeta_{p,i} \frac{a_i a_i^T}{(1 - a_i^T p)^2}. \quad (3.23)$$

The weight vector $\zeta_p \in \mathbb{R}^m$ is obtained by solving the convex program,

$$\zeta_p = \arg \max_{w \in \mathbb{R}^m} \left\{ \sum_{i=1}^m w_i - \frac{1}{a_J} \log \det(A^T S_p^{-1} W^{a_J} S_p^{-1} A) - \beta_J \sum_{i=1}^m \log w_i \right\}, \quad (3.24)$$

where $\beta_J := d/2m$, $a_J := 1 - 1/\log_2(1/\beta_J)$, matrix $W = \text{diag}(w)$ and S_p the slackness matrix at p . The convex program of Eq. (3.24) was introduced in Lee and Sidford (2014)

ALGORITHM 9: JOHN_WALK (\mathcal{P}, p, r)

Input : polytope \mathcal{P} ; point p ; radius r of the John ellipsoid.

Output: A point q in \mathcal{P}

$C \sim$ fair coin;

if C head **then** do nothing, go to the next step;

Propose $y \sim \mathcal{N}\left(x, \frac{r^2}{(d)^{3/2}} J_x^{-1}\right)$;

if $y \notin \mathcal{P}$ **then** $q \leftarrow p$ **end**

else if $y \in \mathcal{P}$ **then**

$\alpha = \left\{ 1, \frac{g_y(x)}{g_x(y)} \right\}$;

$u \sim \mathcal{U}[0, 1]$;

if $u \leq \alpha$ **then** $q \leftarrow y$ **else** $q \leftarrow p$;

end

return q ;

and ζ_p are called approximate John weights. It is also closely related with the problem of computing the largest ellipsoid in P centered at p , which was first studied in [J. \(2014\)](#). However, John walk step is costlier than that of Dikin and Vaidya walk by a constant factor ([Yuansi et al., 2018](#)).

Practical performance. In [Yuansi et al. \(2018\)](#) they compare Vaidya walk against Dikin and John walk by performing experiments on well-rounded polytopes. Both Dikin and Vaidya presented to achieve a better performance than John walk. However, no extensive experimental evaluation of John walk took place until now.

Software. John walk is implemented in the C++ package `PolytopeWalk`. The same implementation is integrated in `volesti`.

3.4.8 Hamiltonian Monte Carlo (HMC)

HMC can be used to sample from any probability density function $\pi(\cdot)$ in \mathbb{R}^n . It simulates a particle moving in a conservative field determined by $-\log \pi(p)$ and $-\nabla \log \pi(p)$. Being at $p \in \mathcal{K}$, HMC introduces an auxiliary random variable $v \in \mathbb{R}^d$, called the *momenta*, and generates samples from the joint density function

$$\pi(p, v) = \pi(v|p)\pi(p). \tag{3.25}$$

Then, marginalizing out v recovers the target density $\pi(p)$. Usually we consider the momenta v to follow the standard Gaussian distribution $\mathcal{N}(0, I_n)$. Then, the probability density function,

$$\pi(p, v) \propto e^{-H(p, v)}, \quad (3.26)$$

where $H(p, v) = -\log \pi(p, v) = -\log \pi(p) + \frac{1}{2}|v|^2$ is the corresponding *Hamiltonian function*. In each step being at p , HMC draws a value for the momentum, $v \sim \mathcal{N}(0, I_n)$. Then, (p, v) is given by the Hamilton's system of ordinary differential equations (ODE),

$$\begin{aligned} \frac{dp}{dt} &= \frac{\partial H(p, v)}{\partial v} \\ \frac{dv}{dt} &= -\frac{\partial H(p, v)}{\partial p} \end{aligned} \Rightarrow \begin{cases} \frac{dp(t)}{dt} = v(t) \\ \frac{dv(t)}{dt} = -\nabla \log \pi(p) \end{cases} \quad (3.27)$$

The solution of the ODE in Eq. (3.27) gives a trajectory $p(t) \in \mathbb{R}^n$, where, ideally, the next Markov point is randomly picked from it. When a closed form solution is not known, the ODE is solved by the leapfrog method so that the convergence is guaranteed. Thus, to discretize the Hamiltonian Dynamics we use the leapfrog integrator,

$$\hat{v}_{i+1} = v_i + \frac{\eta}{2} \nabla \log(\pi(p_i)), \quad p_{i+1} = p_i + \eta \hat{v}_{i+1}, \quad v_{i+1} = \hat{v}_{i+1} + \frac{\eta}{2} \nabla \log(\pi(p_{i+1})),$$

where $\eta > 0$ is the leapfrog step. The new position is accepted with a probability obtained by a regular Metropolis filter. In general, the step length and the number of steps until we obtain the next Markov point are inputs. However, there are several heuristics to determine those two parameters (Hoffman and Gelman, 2014; Chalkis et al., 2023b) to improve the mixing rate of the random walk.

In the case where π is restricted over a convex body \mathcal{K} there are two versions of HMC that can sample from the target distribution. The first is the Reflective HMC that employs boundary reflections in each leapfrog step so that the random walk stays inside the body. The second is the Riemannian HMC that uses the log-barrier of the convex body to satisfy the constraints in each step.

3.4.9 Reflective NUTS HMC (ReNHMC)

The Reflective HMC using a leapfrog integrator was introduced for log-concave sampling from linear and non-linear convex bodies (Afshar and Domke, 2015; Chalkis et al., 2023b). The mixing time is unknown and an open problem. ReNHMC is a Reflective HMC sampler that employs the NUTS criterion (Hoffman and Gelman, 2014). NUTS exploits the fact that a Hamiltonian trajectory tends to fill the space by forming a spiral. Thus, the optimal number of leapfrog steps is that one when the trajectory starts making a U-turn and approaches the initial point. NUTS is a heuristic that checks the inner products of the position and momenta between the current and the initial position to declare stopping.

Step overview. When the position is updated, i.e., $p_{i+1} = p_i + \eta \hat{v}_{i+1}$, it might lie outside of \mathcal{K} . In this case, ReHMC reflects the ray $\{p_i + t \hat{v}_{i+1} \mid t \geq 0\}$ so that the new proposal is inside \mathcal{K} . Notice, that the ray may need to be reflected more than once in the case where the step length η again leads outside of \mathcal{K} after a certain reflection. The proposal is again accepted after applying a Metropolis filter. In each leapfrog step we check the NUTS criterion given in Hoffman and Gelman (2014).

Practical performance. In Chalkis et al. (2023b) they show that ReHMC outperforms both HaR and CDHR for log-concave sampling from convex polytopes and spectrahedra in thousands of dimensions. ReNHMC has never been implemented for truncated sampling before this paper.

Software. ReNHMC is implemented in the C++ library `volesti`.

3.4.10 Constrained Riemannian HMC (CoRHMC)

CoRHMC introduced in Kook et al. (2022) can be used for log-concave sampling over a convex polytope \mathcal{P} . In general, Riemannian HMC uses a Riemannian manifold over the convex body \mathcal{K} . For this, it defines a local distance and integrate it in the Hamiltonian function, i.e.,

$$H(p, v) = -\log \pi(p) + \frac{1}{2} v^T M(p)^{-1} v + \frac{1}{2} \log \det M(p), \quad (3.28)$$

where $M(p) \in \mathbb{R}^{n \times n}$ is a position-dependent positive definite matrix. The, the ODE solved in each step becomes,

$$\begin{aligned} \frac{dp}{dt} &= \frac{\partial H(p, v)}{\partial v} = M(p)^{-1}v \\ \frac{dv}{dt} &= -\frac{\partial H(p, v)}{\partial p} = -\nabla \log \pi(p) - \frac{1}{2} \text{Tr} \left[M(p)^{-1} \frac{\partial M(p)}{\partial p} \right] + \frac{1}{2} p^T M(p)^{-1} \frac{\partial M(p)}{\partial p} M(p)^{-1} p, \end{aligned} \tag{3.29}$$

where $\text{Tr}[\cdot]$ is the regular matrix trace. Unfortunately, the discrete leapfrog integrator does not guarantee convergence to the target distribution. To solve this ODE one should use a symplectic numerical integrator ([Girolami and Calderhead, 2011](#)). However, this task seems quite challenging in practice. The main difficulty is to achieve the numerical accuracy required when the Markov point is close enough to the boundary so that the computations would not lead to numerical overflow.

CoRHMC is a specialized sampler for convex polytopes. In [Kook et al. \(2023\)](#) they provide efficient methods to solve the ODE in Eq. (3.29) by achieving a cost per step of $\mathcal{O}(n^{3/2})$. In [Kook et al. \(2023\)](#) they show that CoRHMC mixes after $\tilde{\mathcal{O}}(mn^3)$ steps which is worse than the mixing time of both HaR and BaW for rounded distributions. However, this sampler performs significantly well in practice by outperforming both HaR and CDHR as shown in [Kook et al. \(2022\)](#). Its main advantage is that its mixing rate does not depend on the roundness of the distribution; thus, CoRHMC does not require any rounding preprocess for skinny distributions (or skinny polytopes in the case of uniform distribution). Thus, it can sample efficiently from hard instances in thousands of dimensions within minutes ([Kook et al., 2022](#)). Notably, it has never been performed a comparison between CoRHMC and the MMCS with BiW given in [Chalkis et al. \(2021b\)](#) for uniform sampling over a convex polytope.

Software. CoRHMC is implemented in the MATLAB package `PolytopeSamplerMatlab` ([Kook et al., 2021](#)). `volesti` also contains an implementation of CoRHMC in C++.

3.5 Empirical Study

We employ geometric random walks to explore factor anomalies in constrained long-only portfolios. To do so, we follow the investment guidelines of a prominent market index which we believe accurately represents the typical framework followed by many mutual funds and institutional investors. The index that we look at is the MSCI Diversified Multiple-Factor index²⁵ (hereinafter referred to as the DMF index) for the U.S. market. The DMF index aims to create systematic exposure to the four factors: value, momentum, quality, and size, while maintaining a risk profile similar to that of the underlying capitalization-weighted parent index, which is the MSCI USA index. The allocation of the DMF index is determined through an optimization which maximizes the exposure to a combination of the four factors, subject to a risk tolerance set equal to the ex-ante variance of the parent index at the time of calculation.

MSCI's prospectus explains that the DMF index is designed to systematically capture the additional sources of systematic return associated with the four factor investing styles. However, the MSCI approach significantly differs from the typical academic literature on factor premia (measured by the performance difference between a long position in a portfolio of companies with the most favorable factor characteristics and a short position in a basket of assets with the least favorable factor characteristic) in that their model operates in a long-only setup with stringent constraints on asset weights, sector allocations, risk tolerance, and bandwidths on additional factor exposures, all relative to the parent index. In particular, asset weights are bound to the allocation of the parent index by a bandwidth of $\pm 2\%$ -points per stock and of $\pm 5\%$ -points on sector level. Within those limits, MSCI tries to harvest the premias by tilting the capitalization-weighted allocation towards stocks with elevated factor exposure.

A pertinent question arises concerning the possibility of capturing academically investigated factor premia within such a strongly constrained setup. To find out we employ a geometric random walk to sample portfolios from within the set of constraints imposed by the DMF index and analyze whether the sampled portfolio's out-of-sample²⁶ performances varies in a systematic way with their factor scores. Do portfolios with higher value-

²⁵See <https://www.msci.com/diversified-multi-factor-index>.

²⁶Out-of-sample means that, at every point of the back-testing procedure, we only use information that was effectively available at that point in time.

momentum-, quality-, size- or volatility-scores deliver higher, lower, or equal return and risk than their low-scoring counterparts?

The traditional method of factor analysis involves forming percentile portfolios based on a sorting of companies idiosyncratic factor scores. Backtests are then conducted where each percentile portfolio is weighting its constituents equally or according to capitalization. Factor anomalies are identified through the outperformance of the top percentile portfolio relative to the bottom percentile. Interaction with other factors is examined through regression analysis. In contrast, our RP-based approach offers anomaly detection within investor-defined constraints, and factor interactions can be directly controlled by incorporating factor constraints, eliminating the need for indirect regression methods.

Backtesting framework. In Section 3.3, we outlined the process of constructing a RP by initially defining investment constraints geometrically and then deriving the distribution of performance statistics either analytically or through sampling and simulation. However, this construction assumes the availability of clean time series data for all stocks, of equal length, to estimate market parameters. Consequently, the described RPs represent snapshots. For an extended analysis of a stock universe over time, adjustments must consider changes in index composition and corporate actions affecting individual stocks (such as splits, mergers, and dividends). This complicates matters as both the number of investable assets and the geometric body induced by the constraints evolve over time. Therefore, akin to quintile portfolios, RPs must undergo regular rebalancings, precluding analytical solutions even for the naive case.

The method which we pursue is the following: Every month we construct a new RP by casting the constraints employed by the DMF index to a convex body. The linear constraints, which include the lower and upper bounds on asset, sector, and factor exposures, establish a polytope, and the constraint on variance shapes an ellipsoid. These bounds shift from one month to the next, adapting to changes in the allocation and risk level of the parent index. Subsequently, we sample from the convex body using a geometric random walk. Because of the quadratic constraint, the choice of sampler is limited to either HaR, CDHR, BaW, BiW or ReNHMC. Among those, only HaR, CDHR, BaW and ReNHMC allow for a distribution other than the uniform. Since ReNHMC is considerably

faster than the alternatives (Chalkis et al., 2023b; Cousins and Vempala, 2016), ReNHMC is our method of choice (without the variance constraint we would use CoRHMC). As probability model we use the Dirichlet distribution with a parameter vector equal to the weights of the capitalization-weighted parent index, i.e., $\alpha = \omega_{bm}$. This generates samples which have $\mathbb{E}(\omega) = \omega_{bm}$, yet are close to the boundary of the constraints set²⁷.

For each sampled portfolio, we then simulate the cumulative returns path for the subsequent month using the returns of the portfolio’s constituents. At the next rebalancing, the process is repeated and the simulated price paths are concatenated to consecutive time series spanning the entire period of analysis.

Concatenation from one period to the next follows a similar logic to the formation of quintile portfolios. Portfolios are matched based on factor scores, where the portfolio with the highest exposure to a certain factor in one period is combined with the portfolio scoring highest in the following period, and so forth, down to the lowest scoring pair. Sorting the time-series by factor exposure allows for an analysis of the correlation between factor rank and performance over a long-term period. For example, if there is a monotonic relation between momentum score and return, this should be evident in the backtests through a decreasing out-of-sample return with the simulation index. In other words, the first simulation, forming a concatenation of best-momentum portfolios, should yield the largest cumulative return over the entire backtesting period, with performance gradually decreasing for subsequent simulations.

To evaluate the influence of the MSCI constraints, we iterate the backtesting exercise twice, initially applying only the simplex condition before introducing the MSCI constraints in a second round.

3.5.1 Data

The implementation of the MSCI Multifactor framework requires a comprehensive dataset of stock price series from companies covered by the MSCI USA index, which includes large and mid-cap equities traded in the U.S.²⁸. As the composition of the MSCI USA index is proprietary, we suggest utilizing data from the Center for Research in Securities Prices

²⁷Recall that sampling with a parameter vector α with sum less than n results in weights towards the boundaries. We want points to be close to the boundary since the solution of the optimization problem that the DMF index is solving is always a boundary point.

²⁸See <https://www.msci.com/our-solutions/indexes/developed-markets>.

(CRSP) as an alternative data source. We select a subset of stocks listed on the NYSE, AMEX, and NASDAQ indexes, focusing solely on ordinary common shares of companies incorporated in the U.S. (CRSP shares code 10 and 11). Additionally, we exclude foreign shares, certificates, American depository receipts, shares of beneficial interest, depository units, American trust components, closed-end funds, and real estate investment trusts from our dataset. We found that a good approximation to mimick the MSCI universe is to filter for the largest stocks listed on the three exchanges such that in aggregate they cover 85% of total market capitalization. On average, over the backtesting period, this translates to approximately 600 securities per rebalancing. While this approach does not precisely replicate the MSCI methodology, a comparative analysis using both datasets reveals qualitatively identical results. The primary goal is not to achieve an exact replication but rather to establish a representative framework. Through the utilization of CRSP data, all our findings can be reproduced and validated.

Furthermore, to be included in our study, stocks need a consistent price history of five years without any gaps larger than two weeks. This is needed to compute the covariance matrix in the quadratic constraint bounding portfolio variance to that of the parent index. Moreover, illiquid stocks are removed from the investable universe. As a threshold, we require a median trading volume over the previous 252 trading days to be above USD 1.5 million. We do this because, on the one hand, such illiquid stocks are not easily tradable and therefore would lead to a large implementation shortfall (i.e., the difference between a simulated performance and one obtained from real investments) and, on the other hand, such companies display artificially low volatility due to a lack of trading and not because they are not risky. The cleaning process is necessary to ensure that at every point in time, the investable universe only contains information that was effectively available at that point in time and to avoid any positive survivorship bias.

All estimations are based on discrete daily total²⁹ returns using closing prices denoted in USD, covering the period 01.01.1995 to 31.12.2022. Backtests begin on 03.01.2000. All descriptive statistics are calculated on annualized discrete monthly returns, as is customary in the financial industry.

²⁹Returns, i.e., the percentage changes in prices from time $t - 1$ to t , are termed *total* when adjusted for dividends (i.e., dividends are re-invested).

Factors. The DMF index actively manages its exposure to specific target and non-target style factors relative to the parent index, not only via a weighting in the objective function, but also by setting benchmark-relative constraints. Book-to-Price, Earnings Yield, Earnings Quality, Investment Quality, Profitability, and Momentum exposures are required to fall within the range of $[0.1, 0.6]$, Earnings Variability, Leverage, and Size within $[-0.6, -0.1]$. Moreover, the index limits its exposure to non-target style factors, namely Beta, Residual Volatility, Growth, and Liquidity to remain within a range of ± 0.1 standard deviations relative to the parent index.

For value and quality, MSCI computes scores by a weighting of sub-scores. Value, for instance, consists of one-third Price-to-Book and two-thirds Earnings Yield. The scores are further sector standardized using the GICS³⁰ classification system. Details can be taken from the MSCI methodology paper.

Table 3A.4 in Appendix 3A shows how we have mapped the MSCI factors to JKP factors. All scores (i.e., exposures) are ultimately standardized to z-scores and winsorized at ± 3 . Note that the standardization is a monotonous transformation which does not affect the ordering of the scores. Winsorization, on the other hand, does have an influence.

On a portfolio level, the factor score i.e., the exposure of a portfolio ω to some factor f_s is simply taken to be the weighted sum of the portfolio weights times the company factor scores, i.e., $\sum_i \omega_i \beta_{i,s}$. We notice that, because of the standardization and winsorization steps, the weighted sum method is not exactly the same as when we would regress the portfolio returns on the factors.

3.5.2 Implementation

We use the C++ library `volesti` (Chalkis and Fisikopoulos, 2021; Chalkis et al., 2024), an open-source library for high dimensional MCMC sampling and volume approximation. `volesti` provides the implementations of all the geometric random walks in Section 3.4. The core of its implementation is in C++ to optimize performance while the user interface is implemented in R. The library employs `eigen` (G. et al., 2010) for linear algebra and `boost` (Maurer and Watanabe, 2017) for random number generation.

³⁰See <https://www.msci.com/our-solutions/indexes/gics>

3.5.3 Results

Figures 3.6 and 3.7 contain scatter-plots depicting the relation between portfolio factor exposure and portfolio risk (first column), portfolio factor exposure and portfolio return (second column) and portfolio risk and portfolio return (third column). Factor exposure levels are scaled for all factors to live on the interval $[-1, 1]$. Figure 3.6 reflects the case where no constraints are applied other than the simplex condition. The plots are overlaid by grey points, labeled one to five, indicating the performance statistics of five sorting-based quintile portfolios. The blue line indicates the fit of a polynomial regression (of degree two). Figure 3.7 shows the results when the constraints imposed by the DMF index are met.

We note the following observations: In the initial backtests, where only the simplex constraint is enforced, the results affirm a positive relationship between portfolio factor scores and returns across all factors except size. Interestingly, for size, the relationship appears slightly negative. This could be attributed to the asset universe defined by MSCI, which primarily consists of large and medium-sized companies. Notably, value and momentum exhibit non-linearities, with slopes becoming steeper at both ends of the spectrum. The relationship between portfolio factor scores and risk exhibits a U-shaped pattern, indicating that portfolios with particularly high and particularly low factor exposure tend to have higher variance compared to their unexposed counterparts. This U-shaped pattern is observed across all four factors. The results align closely with classical sorting-based simulations, which are superimposed in Figure 3.6 as grey dots labeled 1 to 5, indicating the corresponding (equally-weighted) quintile portfolios. The primary benefit of the sampling-based approach over the sorting-based method lies in its independence from the weighting within percentile portfolios, eliminating potential strong influences on the results. Additionally, the sampling method provides greater granularity by illustrating the uncertainty in the outcomes of percentile portfolios. Also, it can be seen that some of the quantile portfolio are rather outlying relative to the conditional performance distributions. However, we consider this initial application more as a proof-of-concept. The true advantage becomes apparent when additional constraints are introduced, rendering sorting-based methods impractical.

The addition of benchmark-relative bounds used by the DMF index has varying effects.

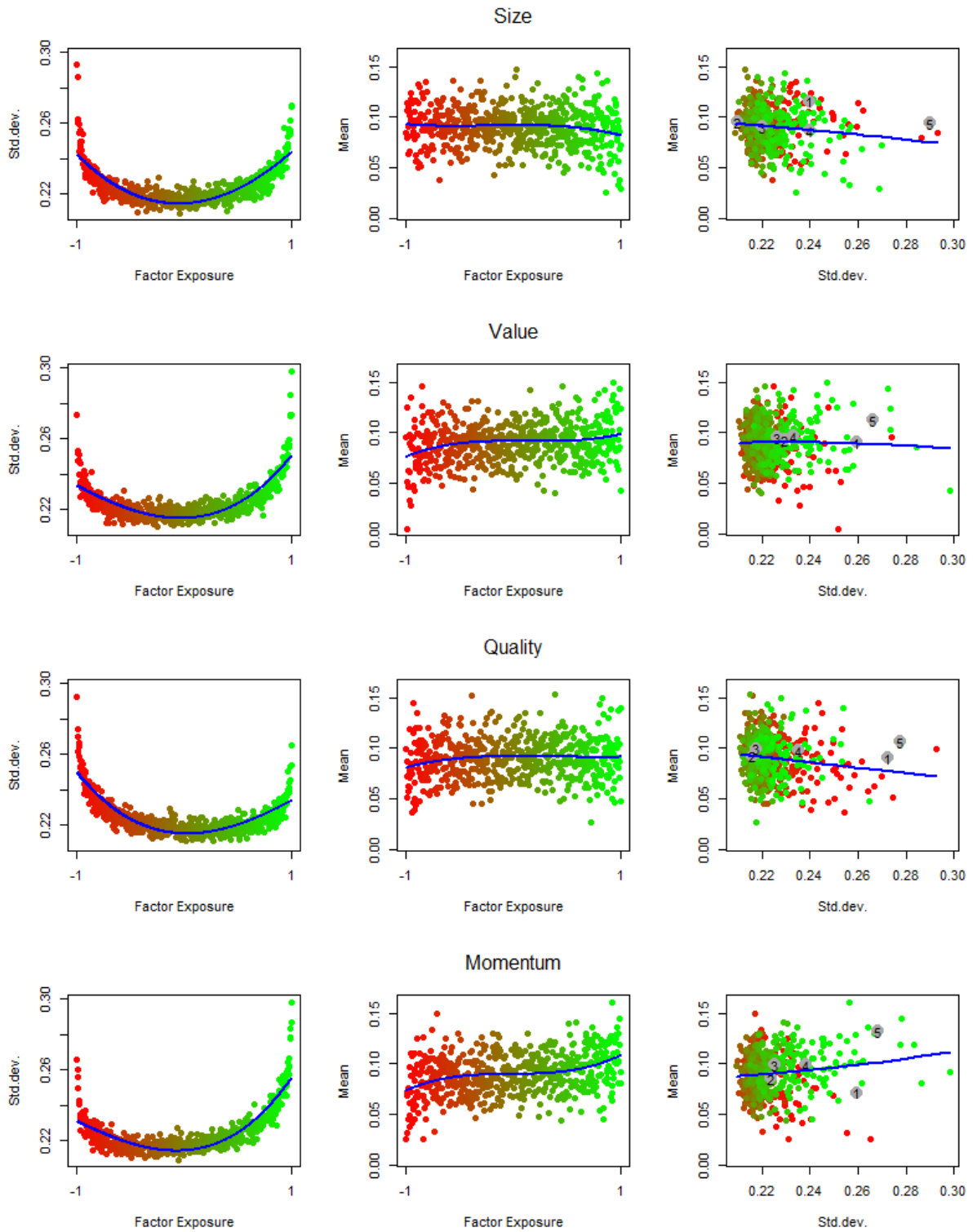


Figure 3.6: Performance vs. factor exposure - Long Only.

For momentum, the relationship between factor exposure and return remains clearly positive, with a correlation coefficient of 0.229, compared to 0.251 for the long-only case. For value, the correlation remains positive but weakens from initially 0.266 to 0.128. For

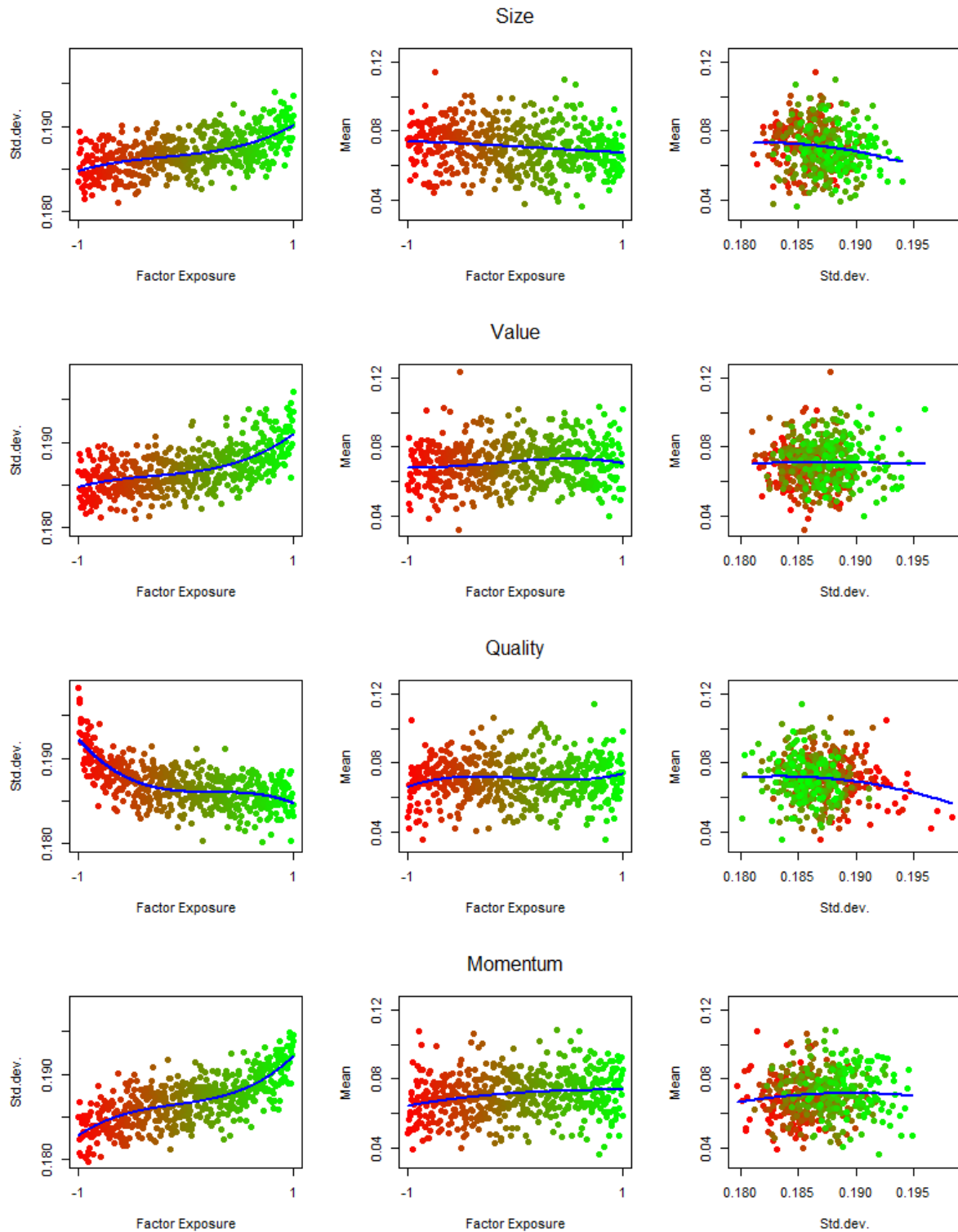


Figure 3.7: Performance vs. factor exposure - Within the constraints of the MSCI Diversified Multifactor Index.

quality, the positive correlation of 0.189 vanishes to 0.038. For size, where the relation was negative, the effect is further strengthened, showing in correlations dropping from

initially -0.106 to -0.168. The non-linear relationships between factor scores and risk are significantly influenced. The graphs indicate that imposing the DMF constraints limits allocations in such a way that one ends up in one of the two legs of the previously identified U-shaped patterns. For size, value, and momentum, this corresponds to the right leg, exhibiting positive (near-linear) relations with correlations of 0.553, 0.693, and 0.813, respectively. Conversely, for quality, it corresponds to the left leg, displaying a negative correlation of -0.580.

Given the observable patterns in the constrained setup we draw the following conclusion: Tilts towards value and momentum are rewarded with higher returns but come at the price of increased volatility. The tilt towards quality has a risk reducing effect without impacting returns, thus increasing overall efficiency. The size tilt should not be actively sought as the relation between size exposure and both, risk and return, is unfavorable. However, we have to keep in mind that, due to the concentration of the capitalization-weighted allocation, any deviation from it will likely increase the size exposure. Value, momentum and quality tilts each imply a size tilt, though without the performance drag of an isolated size tilting. Overall, factor patterns appear to be present even within the very strong bounds implied by the DMF index and an active investment approach trying to harvest them seems justified.

Our method for sampling portfolios within given constraints would allow us to analyse the impact of specific constraints in isolation. For instance, it would be interesting to repeat the analysis after removing one of the constraints, like the bound on variance, or a group of constraints, like the control on non-target factor exposures. We leave this idea for future applied research.

To finalize, let us challenge once more the option to use the RP for statistical testing; here, to test the outperformance of the DMF index relative to its parent. With an Information ratio (IR) of 0.42, the DMF index resides at the 0.025 quantile of the distribution of IR's measured from the RP backtests. We could consider this as indication of skill for MSCI's factor team. However, the quantile value is in direct dependence of the parameter of the Dirichlet distribution which we chose to sample weights from. Although our choice of parametrization is economically justified, the sensitivity on inferential results is hardly acceptable. Conversely, the relation between factor exposure and performance is robust to a wide range of parameter values (we have tested $\alpha = \omega_{bm} \times \{0.01, 0.1, 1, 5, 10\}$,

all giving qualitatively similar results. Only with a scaling factor > 10 do the resulting samples become too homogenous).

3.6 Conclusion

We discussed the use of randomized control techniques in the context of empirical asset pricing and performance evaluation. We examined the scenario where the performance distribution of a random portfolio, serving as the random control group, forms the basis for assessing the performance of an elaborated investment strategy relative to the (null) distribution of performances under chance. A random portfolio was defined as a portfolio whose weights constitute a random vector with a specific probability distribution supported on a bounded space, reflecting the investor’s investment constraints, such as bounds on asset weights or risk limits.

We explained that commonly employed random portfolios do not offer a statistically acceptable experimental design for performance comparisons and cautioned the reader against such a use case with the aid of an illustrative example. However, under consideration of the random portfolios construction methodology, we found random portfolios to be a helpful tool for visual and descriptive perform analysis. Furthermore, we suggested a novel use case, involving the investigation of the relationship between portfolio factor scores and performance in a constrained setup, which allows to empirically test for the presence of factor anomalies under stringent investment guidelines.

We proposed a categorization of random portfolios into three groups — naive, basic, and regularized — based on weight distribution and geometric structure of the distribution’s support. Additionally, we introduced geometric random walks, a class of Markov chain Monte Carlo methods, as a novel and powerful paradigm to sample random portfolio weights adhering to linear and quadratic constraints under various models. We provided a survey of all existing geometric random walks, explaining their strengths, limitations, and computational complexities.

Ultimately, we performed experiments on real-world datasets using the open-source software `volesti` which implements various geometric random walk samplers. We replicated the investment guidelines of the MSCI Diversified Multi-Factor index, a well-known index in the industry. This index aims to harvest sources of active returns by tilting the allocation

of the capitalization-weighted parent index towards the four factors: value, quality, size, and momentum, while respecting bounds on risk, weights, and factor exposures relative to the parent index.

Our study revealed diverse impacts of DMF index constraints on factor-exposure-performance relations, compared to an otherwise unconstrained long-only framework. Momentum maintained a positive relation to returns, nearly as strong as in the long-only case. Value showed a diminishing positive relation, while quality's positive relation vanished in the constrained scenario. Size exhibited a negative relation, which intensified with constraints. Regarding risk, there was a truncation of the initial U-shaped relation, with momentum, value, and size showing increased volatility with higher factor exposures, while quality exhibited a negative correlation with volatility.

In summary, tilts towards value and momentum offered higher returns with increased volatility, quality tilts reduced risk without affecting returns, and size tilts led to lower returns and higher risk. Our findings suggest that factor patterns persist even within stringent constraints, justifying an active investment approach to capitalize on the positive ones.

References

- Afshar, H. and Domke, J. (2015). Reflection, refraction, and Hamiltonian Monte Carlo. In *Advances in neural information processing systems*, pages 3007–3015.
- Ali, M. M. (1973). Content of the frustum of a simplex. *Pacific Journal of Mathematics*, 49:313–322.
- Anstreicher, K. M. (2002). Improved complexity for maximum volume inscribed ellipsoids. *SIAM Journal on Optimization*, 13(2):309–320.
- Arnott, R., Hsu, J., Kalesnik, V., and Tindall, P. (2013). The surprising alpha from malkiel’s monkey and upside-down strategies. *The Journal of Portfolio Management*, 4:91–105.
- Asness, C. S., Frazzini, A., and Pedersen, L. H. (2019). Quality minus junk. *Review of Accounting Studies*, 24:34–112.
- Bachelard, C. (2024). On the geometry of the bootstrap. *Unpublished Manuscript*.
- Bachelard, C., Chalkis, A., and Tsigaridas, V. F. E. (2023). Randomized geometric tools for anomaly detection in stock markets. In Ruiz, F., Dy, J., and van de Meent, J.-W., editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 9400–9416. PMLR.
- BAI (1968). Measuring the investment performance of pension funds for the purpose of inter-fund comparison. *Bank Administration Institute*.
- Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9:3–18.
- Berbee, H. C. P., Boender, C. G. E., Rinnooy Ran, A. H. G., Scheffer, C. L., Smith, R. L., and Telgen, J. (1987). Hit-and-run algorithms for the identification of nonredundant linear inequalities. *Mathematical Programming*, 37(2):184–207.
- Bertsimas, D. and Vempala, S. (2004). Solving convex programs by random walks. *J. ACM*, 51(4):540–556.

- Bickel, P. J., Götze, F., and van Zwet, W. R. (1997). Resampling fewer than n observations. *Statistica Sinica*, 7:1–31.
- Billio, M., Calès, L., and Guegan, D. (2011). A cross-sectional score for the relative performance of an allocation. *International Review of Applied Financial Issues and Economics*, 3.
- Brinsan, G. P. and Fachler, N. (1985). Measuring non-us. equity portfolio performance. 11:73–76.
- Brinsan, G. P., Jr., L. R. H., and Beebower, G. L. (1986). Determinants of portfolio performance. *Financial Analysts Journal*, 41:39–44.
- Burns, P. (2007). Random portfolios for performance measurement. In Kontoghiorghes, E. and Gatu, C., editors, *Optimisation, Econometric, and Financial Analysis*, pages 227–249, New York. Springer.
- Bélisle, C. J. P., Romeijn, H. E., and Smith, R. L. (1993). Hit-and-run algorithms for generating multivariate distributions. *Mathematics of Operations Research*, 18(2):255–266.
- Cales, L., Chalkis, A., and Emiris, I. (2021). On the cross-sectional distribution of portfolios’ returns. *arXiv:2105.06573*.
- Cales, L., Chalkis, A., Emiris, I., and Fisikopoulos, V. (2018). Practical volume computation of structured convex bodies, and an application to modeling portfolio dependencies and financial crises.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52:57–82.
- Chalkis, A., Christoforou, E., Emiris, I. Z., and Dalamagas, T. (2021a). Modeling asset allocation strategies and a new portfolio performance score. *Digital Finance*, 3:333–371.
- Chalkis, A., Emiris, I. Z., and Fisikopoulos, V. (2023a). A practical algorithm for volume estimation based on billiard trajectories and simulated annealing. *ACM J. Exp. Algorithmics*, 28.

- Chalkis, A., Emiris, I. Z., Fisikopoulos, V., Repouskos, P., and Tsigaridas, E. (2022). Efficient sampling in spectrahedra and volume approximation. *Linear Algebra and its Applications*, 648:205–232.
- Chalkis, A. and Fisikopoulos, V. (2021). volesti: Volume Approximation and Sampling for Convex Polytopes in R. *The R Journal*, 13(2):642–660.
- Chalkis, A., Fisikopoulos, V., Papachristou, M., and Tsigaridas, E. (2023b). Truncated log-concave sampling for convex bodies with reflective hamiltonian monte carlo. *ACM Trans. Math. Softw.*, 49(2).
- Chalkis, A., Fisikopoulos, V., and Tsigaridas, E. (2024). volesti. <https://github.com/GeomScale/volesti>.
- Chalkis, A., Fisikopoulos, V., Tsigaridas, E., and Zafeiropoulos, H. (2021b). Geometric Algorithms for Sampling the Flux Space of Metabolic Networks. In Buchin, K. and Colin de Verdière, E., editors, *37th International Symposium on Computational Geometry (SoCG 2021)*, volume 189 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 21:1–21:16, Dagstuhl, Germany. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- Chevallier, A., Pion, S., and Cazals, F. (2018). Hamiltonian Monte Carlo with boundary reflections, and application to polytope volume calculations. Research Report RR-9222, INRIA Sophia Antipolis, France.
- Chiarawongse, A., Kiatsupaibul, S., Tirapat, S., and Roy, B. V. (2012). Portfolio selection with qualitative input. *Journal of Banking & Finance*, 36:489–496.
- Clare, A., Motson, N., and Thomas, S. (2013). An evaluation of alternative equity indices - part 2: Fundamental weighting schemes. *Cass Consulting*, March 2013.
- Cohen, K. J. and Pogue, J. A. (1967). An empirical evaluation of alternative portfolio-selection models. *The Journal of Business*, 40:166–193.
- Cohen, M. B., Cousins, B., Lee, Y. T., and Yang, X. (2019). A near-optimal algorithm for approximating the john ellipsoid. In Beygelzimer, A. and Hsu, D., editors, *Proceedings of*

- the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 849–873. PMLR.
- Cousins, B. (2017). *Efficient high-dimensional sampling and integration*. PhD thesis, Georgia Institute of Technology, Georgia, U.S.A.
- Cousins, B. and Vempala, S. (2015). Bypassing KLS: Gaussian cooling and an $O^*(n^3)$ volume algorithm. In *Proc. ACM STOC*, pages 539–548.
- Cousins, B. and Vempala, S. (2016). A practical volume algorithm. *Mathematical Programming Computation*, 8:133–160.
- Dawson, R. and Young, R. (2003). Near-uniformly distributed, stochastically generated portfolios.
- de Prado, M. L. (2022). Causal factor investing: Can factor investing become scientific? *ADIA Research Lab*.
- Dietz, P. O. (1966). Pension funds: Measuring investment performance.
- Dyer, M., Frieze, A., and Kannan, R. (1991). A random polynomial-time algorithm for approximating the volume of convex bodies. *J. ACM*, 38(1):1–17.
- Efron, B. (1979). The bootstrap: Another look at the jackknife. *The Annals of Statistics*, 7:1–26.
- Emiris, I. and Fisikopoulos, V. (2014). Practical polytope volume approximation. *ACM Transactions of Mathematical Software, 2018*, 44(4):38:1–38:21. Conf. version: Proc. Symp. Comp. Geometry, 2014.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56.
- Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116:1–22.
- Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81:607–636.

- Frigyik, B., Gupta, M., and Chen, Y. (2010). Shadow dirichlet for restricted probability modeling. *Advances in Neural Information Processing Systems 23*, pages 613–621.
- G., G., J., B., et al. (2010). *Eigen v3*.
- G. van Valkenhoef, T. Tervonen (2022). *hitandrun: "Hit and Run" and "Shake and Bake" for Sampling Uniformly from Convex Shapes*. R Foundation for Statistical Computing.
- Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472. Publisher: Institute of Mathematical Statistics.
- Girolami, M. and Calderhead, B. (2011). Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214.
- Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012). A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773.
- Gryazina, E. and Polyak, B. (2014). Random sampling: Billiard walk algorithm. *European Journal of Operational Research*, 238:497–504.
- Gustafson, A. and Narayanan, H. (2018). John’s walk. *arXiv preprint arXiv:1803.02032*, page NA.
- Haraldsdóttir, H. S., Cousins, B., Thiele, I., Fleming, R. M. T., and Vempala, S. (2017). CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics*, 33(11):1741–1743.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.
- Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdottir, H. S., Wachowiak, J., Keating, S. M., Vlasov, V., Magnusdottir, S., Ng, C. Y., Preciat, G., Zagare, A., Chan, S. H., Aurich, M. K., Clancy, C. M., Modamio, J., Sauls, J. T., Noronha, A., Bordbar, A., Cousins, B., Assal, D. C. E., Valcarcel, L. V., Apaolaza, I., Ghaderi, S., Ahookhosh, M., Guebila, M. B., Kostromins, A., Sompairac, N., Le,

- H. M., Ma, D., Sun, Y., Wang, L., Yurkovich, J. T., Oliveira, M. A., Vuong, P. T., Assal, L. P. E., Kuperstein, I., Zinovyev, A., Hinton, H. S., Bryant, W. A., Artacho, F. J. A., Planes, F. J., Stalidzans, E., Maass, A., Vempala, S., Hucka, M., Saunders, M. A., Maranas, C. D., Lewis, N. E., Sauter, T., Palsson, B., Thiele, I., and Fleming, R. M. (2019). Creation and analysis of biochemical constraint-based models: The cobra toolbox v3.0. *Nature Protocols*, 14:639–702.
- Hoffman, M. D. and Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623.
- Huang, K.-L. and Mehrotra, S. (2013). An empirical evaluation of walk-and-round heuristics for mixed integer linear programs. *Computational Optimization and Applications*, 55:545–570.
- J., F. (2014). Extremum Problems with Inequalities as Subsidiary Conditions. In Giorgi, G. and Kjeldsen, T. H., editors, *Traces and Emergence of Nonlinear Programming*, pages 197–215. Springer, Basel.
- Jadebeck, J. F., Theorell, A., Leweke, S., and Nöh, K. (2020). HOPS: high-performance library for (non-)uniform sampling of convex-constrained models. *Bioinformatics*, 37(12):1776–1777.
- Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48:65–91.
- Jensen, T. I., Kelly, B., and Pedersen, L. H. (2023). Is there a replication crisis in finance? *The Journal of Finance*, 78:2465–2518.
- Jia, H., Laddha, A., Lee, Y. T., and Vempala, S. (2021). Reducing isotropy and volume to KLS: An $O^*(n^3\psi^2)$ volume algorithm. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, page 961–974, New York, NY, USA. Association for Computing Machinery.
- Jobson, J. D. and Korkie, B. M. (1981). Performance hypothesis testing with the sharpe and treynor measures. *The Journal of Finance*, 36:889–908.

- Kalai, A. T. and Vempala, S. (2006). Simulated annealing for convex optimization. *Mathematics of Operations Research*, 31(2):253–266.
- Kannan, R., Lovász, L., and Simonovits, M. (1997). Random walks and an $O(n^5)$ volume algorithm for convex bodies. *Random Structures and Algorithms*, 11:1 – 50.
- Kannan, R. and Narayanan, H. (2012). Random walks on polytopes and an affine interior point method for linear programming. *Mathematics of Operations Research*, 37:1 – 20.
- Kasak, E. and Pohlmeier, W. (2019). Testing out-of-sample portfolio performance. *International Journal of Forecasting*, 35:540–554.
- Kaufman, D. E. and Smith, R. L. (1998). Direction choice for accelerated convergence in hit-and-run sampling. *Operations Research*, 46(1):84–95.
- Khachiyan, L. G. and Todd, M. J. (1993). On the complexity of approximating the maximal inscribed ellipsoid for a polytope. *Mathematical Programming*, 61(1):137–159.
- Kim, W. C. and Lee, Y. (2016). A uniformly distributed random portfolio. *Quantitative Finance*, 16:297–307.
- Kook, Y., Lee, Y. T., Shen, R., and Vempala, S. (2021). Polytopesamplermatlab. <https://github.com/ConstrainedSampler/PolytopeSamplerMatlab>.
- Kook, Y., Lee, Y. T., Shen, R., and Vempala, S. (2022). Sampling with riemannian hamiltonian monte carlo in a constrained space. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A., editors, *Advances in Neural Information Processing Systems*, volume 35, pages 31684–31696. Curran Associates, Inc.
- Kook, Y., Lee, Y. T., Shen, R., and Vempala, S. (2023). Condition-number-independent convergence rate of riemannian hamiltonian monte carlo with numerical integrators. In Neu, G. and Rosasco, L., editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 4504–4569. PMLR.
- Kumar, P. and Yildirim, E. A. (2005). Minimum-volume enclosing ellipsoids and core sets. *Journal of Optimization Theory and Applications*, 126(1):1–21.

- Laddha, A., Lee, Y. T., and Vempala, S. (2020). Strong self-concordance and sampling. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, page 1212–1222, New York, NY, USA. Association for Computing Machinery.
- Laddha, A. and Vempala, S. (2023). Convergence of gibbs sampling: Coordinate hit-and-run mixes fast. *Discrete & Computational Geometry*.
- Lasserre, J. B. (2015). Volume of slices and sections of the simplex in closed form. *Optimization Letters*, 9:1263–1269.
- Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15:850–859.
- Lee, Y., Kwon, D.-G., Kim, W. C., and Fabozzi, F. J. (2018). An alternative approach for portfolio performance evaluation: enabling fund evaluation relative to peer group via malkiel’s monkey. *Applied Economics*, 50:4318–4327.
- Lee, Y. T. and Sidford, A. (2014). Path Finding Methods for Linear Programming: Solving Linear Programs in \tilde{O} (vrank) Iterations and Faster Algorithms for Maximum Flow. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 424–433. ISSN: 0272-5428.
- Lee, Y. T. and Vempala, S. (2017). Eldan’s stochastic localization and the KLS hyperplane conjecture: An improved lower bound for expansion. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 998–1007.
- Lee, Y. T. and Vempala, S. (2017). Geodesic walks in polytopes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 927–940, New York, NY, USA. ACM.
- Lee, Y. T. and Vempala, S. (2018a). Convergence rate of riemannian hamiltonian monte carlo and faster polytope volume computation. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, page 1115–1121, New York, NY, USA. Association for Computing Machinery.

- Lee, Y. T. and Vempala, S. (2018b). The Kannan-Lovasz-Simonovits conjecture. *arXiv preprint arXiv:1807.03465*, page NA.
- Liang, B. (1999). Price pressure: Evidence from the dartboard column. *The Journal of Business*, 72:119–134.
- Lintner, J. (1965). Security prices, risk and maximal gains from diversification. *The Journal of Finance*, 20:587–615.
- Lisi, F. (2011). Dicing with the market: Randomized procedures for evaluation of mutual funds. *Quantitative Finance*, 11:163–172.
- Lovász, L. and Simonovits, M. (1990). Mixing rate of markov chains, an isoperimetric inequality, and computing volume. *Proceedings 31st Annual Symposium on Foundations of Computer Science, 22.-24. Oct. 1990*, pages 346–355.
- Lovász, L. and Simonovits, M. (1993). Random walks in a convex body and an improved volume algorithm. *Random Structures and Algorithms*, 4:359–412.
- Lovász, L. and Vempala, S. (2003a). Hit and run is fast and fun. *Technical Report MSR-TR-2003-05, Microsoft Research, Redmond.*, page NA.
- Lovász, L. and Vempala, S. (2003b). Logconcave functions: geometry and efficient sampling algorithms. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 640–649.
- Lovász, L. and Vempala, S. (2006a). Fast algorithms for logconcave functions: Sampling, rounding, integration and optimization. *Proceedings 47th Annual IEEE Symposium on Foundations of Computer Science, 21- 24. Oct 2006*.
- Lovász, L. and Vempala, S. (2006b). Hit-and-run from a corner. *SIAM J. Comp.*, 35(4):985–1005.
- Lovász, L. and Vempala, S. (2006). Simulated annealing in convex bodies and an $O^*(n^4)$ volume algorithms. *J. Computer and System Sciences*, 72:392–417.
- Malkiel, B. G. (1973). A random walk down wall street.

- Mangoubi, O. and Smith, A. (2017). Rapid mixing of hamiltonian monte carlo on strongly log-concave distributions.
- Mangoubi, O. and Vishnoi, N. (2019). Faster algorithms for polytope rounding, sampling, and volume computation via a sublinear "ball walk". In *FOCS*.
- Maurer, J. and Watanabe, S. (2017). Boost random number library. Software.
- Megchelenbrink, W., Huynen, M., and Marchiori, E. (2014). optgpsampler: An improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PLOS ONE*, 9(2):1–8.
- Memmel, C. (2003). Performance hypothesis testing with the sharpe ratio. *Finance Letters*, 1:21–23.
- Metcalf, G. E. and Malkiel, B. G. (1994). The wall street journal contests: the experts, the darts, and the efficient market hypothesis. *Applied Financial Economics*, 4:371–374.
- Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, 34:768–783.
- Narayanan, H. (2016). Randomized interior point methods for sampling and optimization. *Annals of Applied Probability*, 26(1):597–641. Publisher: Institute of Mathematical Statistics.
- Narayanan, H. and Srivastava, P. (2020). On the mixing time of coordinate hit-and-run.
- Neal, R. M. (2011). *MCMC Using Hamiltonian Dynamics*, chapter 5. CRC Press.
- Nemirovski, A. (1999). On self-concordant convex-concave functions. *Optimization Methods & Software - OPTIM METHOD SOFTW*, 11.
- Nielsen, S. P. (2016). ll-ACHRB: a Scalable Algorithm for Sampling the Feasible Solution Space of Metabolic Networks. *Bioinformatics*, 32:2330–2337.
- Pettengill, G. N. and Clark, J. M. (2001). Estimating expected returns in an event study framework: Evidence from the dartboard column. *Quarterly Journal of Business and Economics*, 40:3–21.

- Rosenberg, B., Reid, K., and Lanstein, R. (1985). ersuasive evidence of market inefficiency. *The Journal of Portfolio Management*, 11:9–16.
- Roy, V. (2020). Convergence Diagnostics for Markov Chain Monte Carlo. *Annual Review of Statistics and Its Application*, 7(1):387–412.
- Rubin, R. (1981). The bayesian bootstrap. *The Annals of Statistics*, 9:130–134.
- Rudelson, M. (1999). Random vectors in the isotropic position. *J. Funct. Anal*, pages 60–72.
- Saa, P. A. and Nielsen, L. K. (2016). ll-ACHRB: A Scalable Algorithm for Sampling the Feasible Solution Space of Metabolic Networks. *Bioinformatics*, 32(15):2330–2337.
- Sachdeva, S. and Vishnoi, N. K. (2016). The mixing time of the dikin walk in a polytope—a simple proof. *Operations Research Letters*, 44(5):630–634.
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19:425–444.
- Sharpe, W. F. (1991). The arithmetic of active management. *The Financial Analysts’ Journal*, 47:7 – 9.
- Smith, R. L. (1984). Efficient monte carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32(6):1296–1308.
- Song, Z., Yang, X., Yang, Y., and Zhou, T. (2022). Faster algorithm for structured john ellipsoid computation.
- Stein, R. (2014). Not fooled by randomness: Using random portfolios to analyse investment funds. *Investment Analysts Journal*, 43:1–15.
- Surz, R. J. (1994). Portfolio opportunity distributions. *The Journal of Investing*, 3:36–41.
- Surz, R. J. (2006). A fresh look at investment performance evaluation: Unifying best practices to improve timeliness and reliability. *The Journal of Portfolio Management*, 32:54–65.

- Tian, G.-L., Fang, H.-B., Tan, M., Qin, H., and Tang, M.-L. (2009). Uniform distributions in a class of convex polyhedrons with applications to drug combination studies. *Journal of Multivariate Analysis*, 100:1854–1865.
- Todd, M. J. and Yildirim, E. A. (2007). On khachiyan’s algorithm for the computation of minimum-volume enclosing ellipsoids. *Discrete Applied Mathematics*, 155(13):1731–1744.
- Vaidya, P. M. (1996). A new algorithm for minimizing convex functions over convex sets | SpringerLink.
- Van den Meersche, K., Soetaert, K., and Van Oevelen, D. (2009). `xsample()`: An r function for sampling linear inverse problems. *Journal of Statistical Software, Code Snippets*, 30(1):1–15.
- Varsi, G. (1973). The multidimensional content of the frustum of the simplex. *Pacific Journal of Mathematics*, 46:303–314.
- Vempala, S. (2005). Geometric random walks: A survey. *Combinatorial and Computational Geometry*, 52:577–616.
- Yao, A. (2020). `walkr`. <https://github.com/andy Yao95/walkr>.
- Yuansi, C. (2021). An almost constant lower bound of the isoperimetric coefficient in the KLS conjecture. *Geometric and Functional Analysis*, 31(1):34–61.
- Yuansi, C., Dwivedi, R., Wainwright, M. J., and Bin, Y. (2017). Vaidya walk: A sampling algorithm based on the volumetric barrier. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1220–1227.
- Yuansi, C., Dwivedi, R., Wainwright, M. J., and Bin, Y. (2018). Fast MCMC Sampling Algorithms on Polytopes. *Journal of Machine Learning Research*, 19(55):1–86.
- Yuansi, C., Dwivedi, R., Wainwright, M. J., and Bin, Y. (2020). Polytopewalk. <https://github.com/yuachen/polytopewalk>.
- Yuansi, C. and Eldan, R. (2022). Hit-and-run mixing via localization schemes.
- Zhang, Y. and Gao, L. (2003). On numerical solution of the maximum volume ellipsoid problem. *SIAM Journal on Optimization*, 14(1):53–76.

Appendix

3A Factor mapping

MSCI factor	JKP factor
Book-to-Price	Book-to-market equity
Earnings Yield	Earnings-to-price
Earnings Quality	Total accruals
Investment Quality	Return on equity
Profitability	Gross profits-to-assets
Momentum	Price momentum t-12 to t-1
Earnings Variability	Earnings variability
Leverage	Book leverage
Size	Market Equity
Beta	Market Beta
Residual Volatility	Idiosyncratic volatility from the CAPM (252 days)
Growth	Sales Growth (1 year)
Liquidity	Liquidity of market assets

Table 3A.4: Mapping of MSCI factors to JKP factors.

Chapter 4

On The Geometry of the Bootstrap

Cyril Bachelard and Olivier Gallay

A version of the article will be submitted to *European Journal of Operations Research*.

Abstract

The study examines nonparametric bootstrap techniques from a geometric perspective, based on a unified representation of classical and Bayesian frameworks using random weights. It establishes a clear relationship between resampling plans and the distribution of the weights, and presents a direct computational method for the exact Bayesian bootstrap distribution without resorting to sampling. Moreover, the study connects the exact moments of bootstrap distributions to the L_p norm of the random weights and proposes ways to incorporate prior information into bootstrap procedures through a parametrization of the weight distribution. Finally, applications are discussed where it is common to use bootstrap standard errors, which by the geometric perspective, can be determined without the need for resampling.

Keywords: Nonparametric bootstrap, bayesian bootstrap, informative priors, concentration phenomena, bootstrap regression.

4.1 Introduction

Consider a sample of n independent and identically distributed (iid) realizations $x = \{x_1, \dots, x_n\}$ of a random variable X having unknown probability distribution F defined on a space \mathcal{X} ¹. Let a statistic of interest be of the form $\hat{\theta} = n^{-1} \sum_{i=1}^n z_i$ where $z_i = g(x_i)$.² The (nonparametric) bootstrap technique constructs an approximation of the sampling distribution for the estimate by resampling, uniformly with replacement, from z and repeatedly recomputing the statistic on the resampled data $z^* = \{z_1^*, \dots, z_n^*\}$. Notice that a bootstrap sample z^* is an iid sample drawn from the nonparametric maximum likelihood estimator for F , i.e., the empirical probability distribution \hat{F}_n putting mass $1/n$ on each element z_i . There exist n^n distinct ways to generate bootstrap samples. However, since the calculation of $\hat{\theta}$ is symmetric in its arguments (i.e., the operation is insensitive to the order in which the arguments are presented), the total number of distinct samples (i.e., up to permutation) is $K = \binom{2n-1}{n}$. In principle, all K combinatorial possibilities and their corresponding occurrence probabilities can be enumerated leading to an exact bootstrap distribution

$$D(\gamma) = \mathbb{P}(\hat{\theta} \leq \gamma) = \sum_{k=1}^K p_k \mathbb{1}\{\hat{\theta}^{(*,k)} \leq \gamma\} \quad (4.1)$$

with $\hat{\theta}^{(*,k)} = n^{-1} \sum_{i=1}^n z_i^{(*,k)}$, $-\infty < \gamma < \infty$ and p_k denoting the probability for $z^{(*,k)}$ to be sampled. In practice, however, the computation of (4.1) by enumeration is only possible for small n (see Section 4.4.2 for more details). It belongs to the computational complexity class of #P-hard problems (Bertsimas and Sturt, 2020), i.e., a class of counting problems which cannot generally be solved in a polynomial time (unless $P = NP$).

Efron (1979), in his seminal paper, introduced the bootstrap as a generalization of the Jackknife (Quenouille, 1949; Tukey, 1958), replacing the deterministic leave-one-out scheme of the Jackknife with a randomization method by using sampling with replacement. The introduction of randomization was motivated by the computational intractability of

¹In this analysis, we restrict our attention to the case where \mathcal{X} denotes an interval of the real line. It is nevertheless noteworthy that the bootstrap method is applicable to distributions defined over broader domains, such as the plane or function spaces.

²The transformation g is introduced to stress that our results extend to statistics beyond the sample mean. As an example, take $z_i = (x_i - \bar{x})^r$ with \bar{x} denoting the arithmetic mean of the sample x , then $\hat{\theta} = n^{-1} \sum_{i=1}^n z_i$ defines the r -th sample central moment of x . Further examples are given in Section 4.5, which discusses applications.

exact calculations. Rather than exhaustively enumerating all K combinatorial possibilities, the bootstrap approximates the distribution by resampling R times from the dataset. Typically, $R \ll K$. The approximation then reads

$$D(\gamma) = \mathbb{P}(\hat{\theta} \leq \gamma) \approx \frac{1}{R} \sum_{r=1}^R \mathbb{1}\{\hat{\theta}^{(*,r)} \leq \gamma\} \quad (4.2)$$

with $\hat{\theta}^{(*,r)} = n^{-1} \sum_{i=1}^n z_i^{(*,r)}$.

The bootstrap approximation has been extensively studied and is well understood in terms of asymptotic first and higher order properties (see e.g., [Singh \(1981\)](#) and [Weng \(1989\)](#)). From its core idea, many variants of the bootstrap of [Efron \(1979\)](#) (henceforth referred to as the classical bootstrap, abbreviated as B), have been derived. Of particular relevance for the current exposition are the generalizations proposed by [Zheng \(1987\)](#) and [Mason and Newton \(1992\)](#), which utilize a random weighting scheme. Additionally, the Bayesian bootstrap, introduced by [Rubin \(1981\)](#) and hereinafter referred to as BB , along with its variants such as the weighted likelihood bootstrap by Newton and [Newton and Raftery \(1994\)](#), and adaptations of the Bayesian interpretation by [Boos and Monahan \(1986\)](#) using informative priors, are noteworthy.

Common to the classical bootstrap and its descendants is their numerical, random and approximative characteristic. Motivated by the inaccuracy resulting from randomization, which can be substantial in certain applications, [Bertsimas and Sturt \(2020\)](#) proposed a deterministic approach. The authors presented a fully polynomial time approximation scheme (FPTAS) to compute confidence intervals for (4.1) with guaranteed error bounds as well as an exact algorithm based on polynomial multiplication allowing for exact computation of bootstrap confidence intervals which is described as "practically tractable for data sets of values represented with several significant digits".

In the subsequent discussion, we examine the bootstrap method through a geometric lens, unveiling novel sampling-free approaches - both exact and approximate. Specifically, our analysis reveals the existence of an efficient deterministic algorithm for computing the exact Bayesian bootstrap distribution of [Rubin \(1981\)](#), along with certain derivative methods that incorporate prior information. Additionally, we present analytical findings pertaining to the central moments of the exact bootstrap distribution as in (4.1), with the sample size n held constant. We explore these results across various bootstrap procedures,

including classical and Bayesian methods, with and without informative priors.

The remainder of this document is structured as follows. Section 4.2 presents the nonparametric bootstrap from a geometric perspective and shows that the BB distribution relates to the volume of the intersection of a simplex with a half-space, which can be computed exactly and efficiently. The performance of the method is illustrated using a numerical example borrowed from [Bertsimas and Sturt \(2020\)](#). Section 4.3 establishes a connection between the central moments of the exact bootstrap distribution (classical and Bayesian) and geometric properties of random weights. For the classical bootstrap, those properties rest on combinatorial aspects which are briefly reviewed. The results are generalized to the cases with informative priors in Section 4.4. Finally, Section 4.5 suggests practical applications. Section 5.5 concludes and suggests potential directions for future research.

4.2 Geometry of the bootstrap and an exact algorithm

Sampling with replacement from \hat{F}_n means drawing counts $c = \{c_1, \dots, c_n\}$ from a Multinomial distribution $\text{Mult}(\pi_1, \dots, \pi_n)$ with n trials where all probabilities are equal, i.e., $\pi_i = 1/n$, $i = 1, \dots, n$. Normalizing each count produces weights $\omega_i = c_i/n$. Instead of recomputing the statistic in question on resampled data z^* , the bootstrap statistic can be computed on the weighted data, i.e., as $\hat{\theta}^* = \sum_{i=1}^n \omega_i^* z_i$, where the weights ω^* are drawn from the scaled Multinomial distribution, denoted $M_{\mathbf{1}/n}$, $\mathbf{1}$ being a vector of ones, therefore satisfying $\omega_i^* \geq 0$ and $\sum_{i=1}^n \omega_i^* = 1$. In this sense, the data, z , are fixed³ whereas the weights, ω , are random and averaging across bootstrap samples effectively means averaging over the distribution of the weights (i.e., $\mathbb{E}(\hat{\theta}) = \sum_{i=1}^n \mathbb{E}_{M_{\mathbf{1}/n}}(\omega)_i z_i$). As such, the resampling plans of various bootstrap procedures can be expressed in terms of a distributional model of the weights.

The geometric properties of the distribution of the weights induced either implicitly or explicitly by different resampling plans reveal much about the properties of the corresponding bootstrap distributions and inferential conclusions. For instance, exact standard errors depend on the expected squared L_2 -norm of the weights and exact higher central moments on higher-order expected L_p -norms under various models (see Section

³The nonparametric bootstrap is always conditional on the observed sample. Special to our proposed geometric approach is that we treat z as a constant vector defining a hyperplane.

4.3). Further, the exact BB distribution relates to a ratio of volumes which can also be computed exactly (see Section 4.2.1).

Geometrically, the set of possible weights implied by the classical bootstrap forms a discrete grid of nodes, denoted G_n , over the standard $(n-1)$ -simplex⁴ $\mathcal{S}^{n-1} := \{s \in \mathbb{R}^n | s_i > 0, \sum_i s_i = 1; i = 1, \dots, n\}$. The grid contains a total of $|G_n| = K = \binom{2n-1}{n}$ nodes induced by the set of all normalized partitions of integer n . For instance, the case $n = 3$ yields $\mathcal{P}(3) = \{\rho^{(1)}, \rho^{(2)}, \rho^{(3)}\} = \frac{1}{3}\{(3, 0, 0), (2, 1, 0), (1, 1, 1)\}$ with distinct permutations $Q = \{Q_1, Q_2, Q_3\} = \frac{1}{3}\{\{(3, 0, 0), (0, 3, 0), (0, 0, 3)\}, \{(2, 1, 0), (1, 2, 0), (0, 1, 2), (0, 2, 1), (2, 0, 1), (1, 0, 2)\}, \{(1, 1, 1)\}\}$. We say that a node is activated whenever a corresponding sample $z^{(*,k)}, k = 1, \dots, K$ is drawn. In this sense, the classical bootstrap corresponds to a probability weighting of z by the Cartesian coordinates of the activated grid nodes⁵.

All permutations within a partition share the same activation probabilities, forming what we call activation probability level sets. Those level sets can be enumerated (in principle) and therefore give us the total probability for the sets. We go through the combinatorial exercise of computing them in Section (4.2.3) in order to strengthen the geometric perspective. For large n , $|\mathcal{P}(n)| = L \ll K = |G_n|$.

For given n , the grid is fully determined and, no matter how long the bootstrap approximation is run (i.e., how large one chooses R), off grid weights will never be sampled. This is where the BB generalizes the classical bootstrap. Instead of sampling from the discrete Multinomial, the BB samples from the continuous Dirichlet distribution $\text{Dir}(\alpha = (\alpha_1, \dots, \alpha_n))$ (see Definition 7 below).

The BB uses as non-informative (and improper) prior the limiting Dirichlet distribution with $\alpha \rightarrow 0$ which puts all mass on the vertices of the simplex. Since the Dirichlet distribution is the conjugate prior to the Multinomial distribution, the (proper) posterior is again Dirichlet distributed with altered parameter $\alpha' = \alpha + c = \mathbf{1}$. This so called flat parametrization distributes weights uniformly over the probability simplex.

⁴We use the superscript $(\cdot)^{n-1}$ to stress that the simplex is a $n - 1$ dimensional body embedded in n dimensional (Euclidean) space.

⁵We use the terms weights and grid nodes interchangeably depending on whether we talk about the computation of weighted averages or to give a geometric intuition. An alternative term would be that of a probability mass function (pmf). However, we prefer not to use the latter because it would restrict the discussion to the case where weights live on the probability simplex whereas some of our results generalize to other geometric bodies as the support for the distribution of the weights.

Figure 4.1 provides an illustration of the weights-space for the cases $n = 3$ and $n = 4$. The grey bodies correspond to the 2-dimensional simplex in $n = 3$ (left figure) and the projection of the 3-dimensional simplex in $n = 4$ (right figure). In both figures, black squares represent the nodes of the grid G_n .

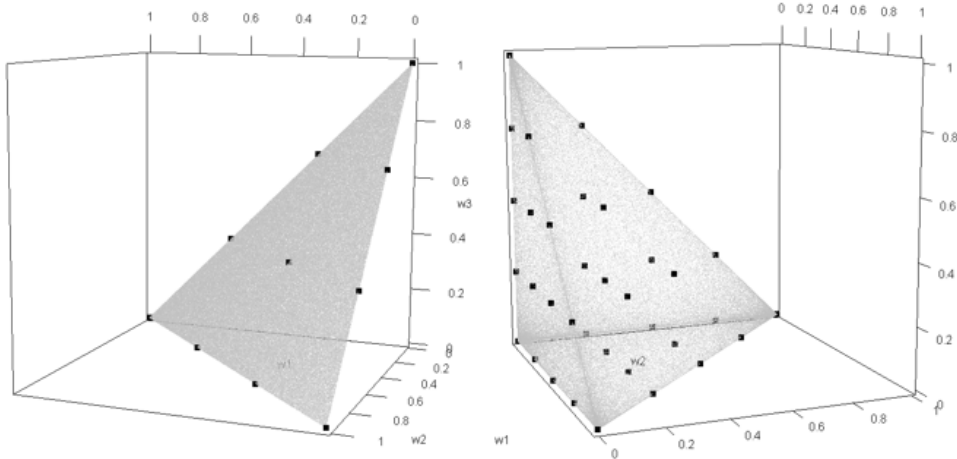


Figure 4.1: Left: Simplex and grid for $n = 3$. Right: Projection of the simplex and grid for $n = 4$.

The main difference between the two bootstrap methods then lies in their interpretation. Instead of simulating the sampling distribution of a statistic (approximating a parameter), as is done in the classical bootstrap, the BB simulates the posterior distribution of the parameter.

Zheng (1987) reinterpreted Rubin's results from a frequentist perspective, introducing the term random weighting bootstrap (RWB). They argued that the conditional distribution obtained from the BB methodology would also approximate the sampling distribution, demonstrating consistency under $\omega \sim D_\alpha$. Mason and Newton (1992) further generalized the random weighting idea, naming it the generalized bootstrap (GB), only assuming that weights are exchangeable, positive, sum to one, and that $\sum i = 1^n (\omega_i - 1/n)^2 \xrightarrow{P} c > 0$ as $n \rightarrow \infty$, encompassing the flat Dirichlet model or $\omega \sim D_4$, as proposed in Weng (1989), as special cases. They provided conditions on the weights to guarantee consistency.

In the subsequent discussion, we abstract from interpretational positions (Bayesian versus frequentist) and focus solely on the procedural aspect, considering the BB, RWB, and the special case of the GB with $\omega \sim D_\alpha$ as equivalent.

In terms of inference, the Bayesian and the frequentist bootstrap methods are well known to produce similar results, as emphasized by Rubin (1981) and thoroughly investigated

by [Lo \(1987\)](#) and [Weng \(1989\)](#). We introduce a fresh perspective on this matter by examining the geometric properties of the weights distribution. Contrary to the common notion that the analysis of the bootstrap and its properties necessitates asymptotic theory ([Bickel et al., 1997](#)), the geometric perspective enables drawing conclusions for any finite n .

Before delving into the comparison between classical and Bayesian bootstrap, we demonstrate how the exact BB distribution for $\hat{\theta}$ can be computed efficiently using a geometric algorithm. Subsequently, inferential similarity justifies utilizing the exact BB distribution as an approximation, free of randomization error, to the exact classical bootstrap distribution.

4.2.1 Computing the exact Bayesian bootstrap distribution

The following result builds upon the fact that the flat Dirichlet distribution spreads uniformly over the simplex. For ease of reference, the following definitions are recalled.

Definition 6 (Uniform Distribution).

Let $S \subset \mathbb{R}^n$ be a bounded d -dimensional surface, where $d \leq n$, and let V denote Lebesgue measure. We say that the vector $X = (X_1, X_2, \dots, X_n) \in S$ is uniformly distributed, or that elements X_1, X_2, \dots, X_n are jointly uniformly distributed over S with positive Lebesgue measure, if $\mathbb{P}(X \in A \subset S) = V(A)/V(S)$.

Definition 7 (Dirichlet Distribution).

Let $\omega = (\omega_1, \dots, \omega_n)$ be a vector with n components, where $\omega_i \geq 0$ for $i = 1, 2, \dots, n$ and $\sum_{i=1}^n \omega_i = 1$. The Dirichlet probability density function (pdf) is

$$f(\omega_1, \dots, \omega_n) = \frac{1}{B(\alpha)} \prod_{i=1}^n \omega_i^{\alpha_i - 1},$$

$$B(\alpha) = \int_{S^{n-1}} \prod_{i=1}^n \omega_i^{\alpha_i - 1} d\omega = \frac{\prod_{i=1}^n \Gamma(\alpha_i)}{\Gamma(\alpha_0)}$$

with $\alpha_0 = \sum_{i=1}^n \alpha_i$ and $\alpha_i \geq 0$ for $i = 1, \dots, n$. $B(\alpha)$ denotes the multinomial beta function acting as the normalizing constant which can be represented with the gamma function $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$.

In the symmetric case, where all elements of the parameter vector α are equal to 1, Dirichlet random variables are uniformly distributed over the simplex. When $n = 2$, one

finds the Beta(α_1, α_2) distribution as a special case.

The marginal distribution of ω_i are Beta($\alpha_i, \alpha_0 - \alpha_i$) with expectation $\mathbb{E}(\omega_i) = \alpha_i/\alpha_0$, variance $\mathbb{V}(\omega_i) = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)}$, covariance $\mathbb{C}(\omega_i, \omega_j) = \frac{-\alpha_i\alpha_j}{\alpha_0^2(\alpha_0 + 1)}, i \neq j$, and $\mathbb{E}(\omega_i, \omega_j) = \frac{\alpha_i\alpha_j}{\alpha_0(\alpha_0 + 1)}, i \neq j$.

Note that the Dirichlet distribution is supported by the simplex \mathcal{S}^{n-1} which has Lebesgue measure zero and is therefore singular with respect to (w.r.t.) Lebesgue measure in \mathbb{R}^n . The simplex constraint is sometimes made explicit by specifying the density over $n - 1$ free variables as $f(\omega) = \frac{1}{\mathbb{B}(\alpha)} \prod_{i=1}^{n-1} \omega_i^{\alpha_i-1} (1 - \sum_{i=1}^{n-1} \omega_i)^{\alpha_i-1}$.

Building upon the above definition we can formulate the exact BB distribution as

$$C(\gamma) = \mathbb{P}(\hat{\theta} \leq \gamma) = \int_{\mathcal{S}^{n-1}} f(\omega) \mathbb{1}\{z^\top \omega \leq \gamma\} d\omega. \quad (4.3)$$

Proposition 8. *Let $H(z, \gamma) := \{h \in \mathbb{R}^n | z^\top h \leq \gamma\}$ denote a half-space with normal z and scalar γ . Then, $C(\gamma) = \mathbb{P}(\hat{\theta} \leq \gamma) = \frac{V(\mathcal{S}^{n-1} \cap H(z, \gamma))}{V(\mathcal{S}^{n-1})}$.*

Proof.

$$\begin{aligned} C(\gamma) &= \mathbb{P}(\hat{\theta} \leq \gamma) = \int_{\mathcal{S}^{n-1}} f(\omega) \mathbb{1}\{z^\top \omega \leq \gamma\} d\omega \\ &= \int_{\mathcal{S}^{n-1}} \left(\frac{\Gamma(\alpha_0)}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=1}^n \omega_i^{\alpha_i-1} \right) \mathbb{1}\{z^\top \omega \leq \gamma\} d\omega \\ &= \frac{\Gamma(\alpha_0)}{\prod_{i=1}^n \Gamma(\alpha_i)} \int_{\mathcal{S}^{n-1}} \left(\prod_{i=1}^n \omega_i^{\alpha_i-1} \right) \mathbb{1}\{z^\top \omega \leq \gamma\} d\omega \\ &= \frac{1}{\int_{\mathcal{S}^{n-1}} \prod_{i=1}^n \omega_i^{\alpha_i-1} d\omega} \int_{\mathcal{S}^{n-1}} \left(\prod_{i=1}^n \omega_i^{\alpha_i-1} \right) \mathbb{1}\{z^\top \omega \leq \gamma\} d\omega \\ &= \frac{\int_{\mathcal{S}^{n-1} \cap H(z, \gamma)} \prod_{i=1}^n \omega_i^{\alpha_i-1} d\omega}{\int_{\mathcal{S}^{n-1}} \prod_{i=1}^n \omega_i^{\alpha_i-1} d\omega} \\ &= \frac{V(\mathcal{S}^{n-1} \cap H(z, \gamma))}{V(\mathcal{S}^{n-1})} \end{aligned}$$

where the last step follows directly from the definition of uniformity and because $\omega_i^{\alpha_i-1} = 1$ for the symmetric case where all $\alpha_i = 1$. \square

Computation of the ratio $\frac{V(\mathcal{S}^{n-1} \cap H(z, \gamma))}{V(\mathcal{S}^{n-1})}$ has been subject to several studies (Varsi, 1973;

Ali, 1973; Cho and Cho, 2001; Lasserre, 2015; Dirksen, 2017). Lasserre (2015) provided a closed form expression to the problem. Unfortunately, their solution becomes numerically unstable when $n \geq 20$, as noted in Cales et al. (2018) and Cales et al. (2021).

Varsi (1973) proposed an algorithm to compute the (nonempty) content of the simplex cut off by a hyperplane (the frustum of the simplex) by means of a dissection technique. Ali (1973) shortly after proved that the algorithm is equivalent to divided differences using the Fourier-Stieltjes transform. For convenience, we reproduce the structure of Varsi’s algorithm in Appendix 4A the way it is presented in Ali (1973) with slightly adapted notation. The interested reader is referred to Varsi (1973) and Ali (1973) for the proof.

Varsi’s algorithm requires $O(n^2)$ computations to get the level of the cumulative (exact bootstrap) distribution function (cdf) at a particular value. In order to evaluate the full cdf, the algorithm has to be evaluated for a sequence of threshold values γ , i.e., for a series of parallel shifted hyperplanes. The pdf may be obtained by numerical derivation, or, as suggested in Cales et al. (2021), by use of B-splines (Curry and Schoenberg, 1966) and making use of the Boor-Cox recursion formula (Boor, 1972; Cox, 1972). Cales et al. (2021) further provide the derivation of the closed-form expression for the ratio of volumes given in Lasserre (2015), yet note that the computation inherits the numerical instability.

Varsi’s algorithm allows for the computation of the exact bootstrap distribution of linear statistics under the BB paradigm of Rubin (1981). In Section 4.4.3 it is shown that the approach generalizes to cases where the support of the weights is a subset of the simplex (which, to the best of our knowledge, has not been considered as a bootstrap option before), as long as there exists a one-to-one mapping between the initial simplex and the codomain.

4.2.2 Numerical example

Bertsimas and Sturt (2020) provide an example comparing confidence intervals obtained from 100 runs of the randomized algorithm of the classical bootstrap with their proposed deterministic and exact procedures in terms of accuracy and tractability. They consider a dataset $z = (1010, 1020, \dots, 1070, 1, \dots, 1)$ with $n = 81$ elements. Their experiment illustrates a significant randomization error of the traditional method even when R is very large (they set $R = 10^6$; common is to use $R = 10^3$). Out of the 100 bootstrap runs, more

than 25% produced an estimate of less than 25 for the 0.025-quantile, which is way off the true quantile value of 37.08 (rounded to two digits) which they identify with their exact method.

By their deterministic method, the exact quantile value can be approximated to arbitrary precision. However, the proposed method does not scale well with dimension and data complexity.

Varsi's algorithm does not suffer from those limitations. For the current example, Varsi's algorithm returns a value of 38.30 which, not surprisingly, is fairly close to the true value of the exact classical bootstrap. Recall that the value returned by Varsi's algorithm is not just a fast deterministic approximation to the quantile value of the exact classical bootstrap, but corresponds to the true quantile value of the exact Bayesian bootstrap.

4.2.3 Concentration phenomena

Due to well-known concentration phenomena⁶ appearing in high dimensions (concentration of measure (Milman and Schechtman, 1986), concentration of distance (Aggarwal et al., 2001)), uniform weights on the simplex strongly concentrate in a high-volume area of the simplex with increasing n . The same holds for the grid nodes embedded in \mathcal{S}^{n-1} . A basic geometric question therefore arises about the expected length of random weight vectors and concentration bounds that set probabilistic deviation limits. We first analyze this question for the discrete case $\omega \in G_n$, going through the underlying combinatorics step-by-step and making a numerical example.

Partitions and node activation probabilities

As previously stated in Section 4.2, node activation probabilities of the grid induced by the resampling plan of the classical bootstrap follow a multinomial law and many nodes share the same activation probabilities. To gain a clear understanding of this effect, consider the following computational procedures:

First, enumerate all partitions $\mathcal{P}(n)$ for integer dimension n . A partition ρ of a positive integer n is a non-increasing sequence of positive integers $\rho_1, \rho_2, \dots, \rho_n$ such that

⁶The phenomena occur geometrically only in high dimensions, or probabilistically for a large number of random variables with sufficient independence between them. For a standard reference, consult Ledoux (2001).

$\sum_{i=1}^n \rho_i = n$ (see Andrews (1998) for a standard reference). Ordering does not matter, i.e., partitions are identical up to permutation (e.g., there are 30 possible distinct reorderings of vector $(5, 1, 0, 0, 0, 0)$ which we consider identical). The elements of a partition indicate the number of times a certain value is sampled during one bootstrap draw: $\#\{z_i^* = z_i\}$. Each partition $\rho^{(l)} \in \mathcal{P}(n)$, $l = 1, \dots, L$, defines a probability level set. As an example, we set $n = 6$ and show all distinct partitions in Table 4.1. For $n = 6$, there are $K = \binom{2n-1}{n} = 11!/(5!6!) = 462$ grid nodes clustered in $L = 11$ partitions.

The partitions can be thought of as sets of leaves of a probability tree with n layers and n branches at each layer. The number of possible paths through the tree that lead to a specific level set is given by the multinomial coefficient $a_l = n! / \prod_{i=1}^n \rho_i^{(l)}$. The number of leaves for each level set is $b_l = n! / \prod_{j=1}^{J_l} c(\rho^{(l)})_j!$, where $c(\rho^{(l)})$ denotes the number of counts for all J_l distinct values in $\rho^{(l)}$. For instance, $c((5, 1, 0, 0, 0, 0)) = (1, 1, 4)$. Finally, the distinct probabilities p_l for each level set l are given by

$$p_l = a_l / a^\top b = a_l (1/n)^n$$

and the total probability of all nodes contained in a level set (i.e., the probability of activating any node in the level set) by

$$\dot{p}_l = b_l \delta_l = (b_l a_l) / a^\top b = b_l a_l (1/n)^n.$$

Partition $\rho^{(l)}$	# Paths to node a_l	# Nodes per level set b_l	Probability p_l	Total probability \dot{p}_l
$\{6,0,0,0,0,0\}$	1	6	2.14×10^{-5}	1.29×10^{-4}
$\{5,1,0,0,0,0\}$	6	30	1.29×10^{-4}	3.86×10^{-3}
$\{4,2,0,0,0,0\}$	15	30	3.15×10^{-4}	9.65×10^{-3}
$\{4,1,1,0,0,0\}$	30	60	6.43×10^{-4}	3.86×10^{-2}
$\{3,3,0,0,0,0\}$	20	15	4.29×10^{-4}	6.43×10^{-3}
$\{3,2,1,0,0,0\}$	60	120	1.29×10^{-3}	1.54×10^{-1}
$\{3,1,1,1,0,0\}$	120	60	2.57×10^{-3}	1.54×10^{-1}
$\{2,2,2,0,0,0\}$	90	20	1.93×10^{-3}	3.86×10^{-2}
$\{2,2,1,1,0,0\}$	180	90	3.86×10^{-3}	3.47×10^{-1}
$\{2,1,1,1,1,0\}$	360	30	7.72×10^{-3}	2.31×10^{-1}
$\{1,1,1,1,1,1\}$	720	1	1.54×10^{-2}	1.54×10^{-2}

Table 4.1: Partitions for $n = 6$. There are $K = \binom{2n-1}{n} = \frac{(2n-1)!}{(n-1)!n!} = \sum_{l=1}^L b_l = 462$ grid nodes clustered in $L = 11$ probability level sets.

In this low-dimensional example, it is straightforward to compute the exact bootstrap distribution, as given in equation (4.1). One just needs to evaluate all K nodes by enumerating all b_l distinct permutations for each partition (e.g., there are $b_l = 6$ distinct permutations of the partition $\{6,0,0,0,0,0\}$).

Described combinatorial derivation readily extends to alternative bootstrap sampling procedures like the m -out-of- n bootstrap (Bickel et al., 1997). Let's undertake the partitioning exercise again for the scenario where $n = 6$ and $m = 4$, implying that we sample fewer than n observations. The partitions are to be made over m , but each partition vector $\rho^{(l)}$ must be extended with zeros to contain n elements. Calculations of node activation probabilities via the probability tree routine remains almost unchanged. Noticing that the tree has now m levels instead of n , we generalize the formula for the number of paths as

$$a_l = \left(\sum_{i=1}^n \rho_i^{(l)} \right)! / \prod_{i=1}^n \rho_i^{(l)}!$$

and observe that $m = \sum_{i=1}^n \rho_i^{(l)}$.

The distinctive and cumulative level set probabilities read:

$$\begin{aligned} p_l &= a_l / a^\top b = a_l (1/n)^m \\ \dot{p}_l &= b_l p_l = (b_l a_l) / a^\top b = b_l a_l (1/n)^m. \end{aligned}$$

Table 4.2 contains the partitions, probability tree elements and distinctive node activation probabilities induced by the bootstrap which samples $m = 4$ out of $n = 6$ observations.

The geometric interpretation of the bootstrap which samples fewer than n observations is that the induced grid, call it $G_{n,m}$, contains fewer grid nodes which are closer to the boundary of the simplex. In the limiting case $m = 1$, $G_{n,1}$ contains only the vertices of the simplex. Hence, the distribution of the 1 out of n bootstrap equals the empirical distribution of the data $\hat{F}(z)$. We will revisit the m -out-of- n bootstrap method in Chapter 4.4.2 when discussing bootstrap techniques with prior information.

Partition $\rho^{(l)}$	#Paths to node a_l	#Nodes per level set b_l	Probability p_l	Total probability \dot{p}_l
$\{4,0,0,0,0,0\}$	1	6	7.72×10^{-4}	4.63×10^{-3}
$\{3,1,0,0,0,0\}$	4	30	3.09×10^{-3}	9.25×10^{-2}
$\{2,2,0,0,0,0\}$	6	15	4.63×10^{-3}	6.94×10^{-2}
$\{2,1,1,0,0,0\}$	12	60	9.26×10^{-3}	5.56×10^{-1}
$\{1,1,1,1,0,0\}$	24	15	1.85×10^{-2}	2.78×10^{-1}

Table 4.2: Partitions for $n = 6$ and $m = 4$. There are $K = \sum_{l=1}^L b_l = 126$ grid nodes clustered in $L = 5$ probability level sets.

The learning from the above combinatorial exercise is that the total probability strongly concentrates in relatively few partitions. In our low dimensional example, one third of the total probability is attached to a single partition. With increasing n , only a small fraction of all partitions will have non-neglectable total probability. For instance, when $n = 81$, as in the example in Section 4.2.2 borrowed from [Bertsimas and Sturt \(2020\)](#), there are roughly $K = 1.83e + 47$ grid nodes and $L = 18'004'327$ partitions (i.e., distinct probability level sets), out of which just 1863 make up for 99% of the total probability (or, 8298 make up for 99.99% of the total probability which consists of roughly 41% of all nodes). When running the bootstrap, it becomes overwhelmingly likely that the randomly activated nodes belong to the group of partitions with high total probability.

Nodes within any partition (i.e., permutations) share some important geometric properties. First, they have identical Lp -norms for all $p > 0$.⁷ Second, the nodes have identical entropy levels as well as relative rentropy w.r.t. to the centroid of the simplex (measured in terms of Kullback-Leibler divergence).

Because of the concentration phenomenon of the partitions and because nodes within any partition live on the surface of a ball and the fact that high probability partitions have similar or identical distance measures, the overall distribution of nodes is strongly concentrated within a thin shell. If that shell also accommodates most weights from the BB (which it does), then it is clear that the two procedures produce inferentially similar results.

If we randomly, i.e., with uniform probability, draw a point, from either the grid G_n by sampling with replacement, or from the simplex by sampling from the flat Dirichlet distribution, the length of that vector will not entirely random. With high probability, the

⁷Note that for $p < 1$ the function $\|x\|_p = (|x_1|^p + \dots + |x_n|^p)^{1/p}$ does not define a norm since it violates the subadditivity property.

length is going to be close to radius $\mathbb{E}_{D_1}(\|\omega\|_2)$ or $\mathbb{E}_{M_{1/n}}(\|\omega\|_2)$. A simple upper bound on $|\hat{\theta}|$ is given by the Cauchy-Scharz inequality $|\hat{\theta}| \leq \|z\|_2^2 \cdot \|\omega\|_2^2$ where $\|z\|_2^2$ is known and $\|\omega\|_2^2$ can be bounded with high probability because of discussed concentration phenomena.

4.3 Exact moments

One of the primary applications of the bootstrap lies in the estimation of standard error (SE) by the second central moment of the bootstrap distribution. Let us denote by a_r and μ_r the r -th initial and central moments of a random variable X respectively as $a_r = \mathbb{E}(X^r)$ and $\mu_r = \mathbb{E}(X - \mathbb{E}(X))^r$, $r \in \mathbb{N}$. For the sample mean, denoted \bar{x} , recall the well known result $\mu_2(\bar{x}) = \frac{\mu_2}{n}$ which holds since $\mathbb{E}(\bar{x}) = \mathbb{E}(\frac{1}{n} \sum_{i=1}^n x_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(x_i) = \mathbb{E}(x_i) = a_1$ and $\mathbb{E}((\bar{x} - a_1)^2) = \frac{1}{n^2} \mathbb{E}\left(\left(\sum_{i=1}^n (x_i - a_1)\right)^2\right) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}(x_i - a_1)^2 = \frac{\mu_2}{n}$.

The classical bootstrap of the sample mean (in the limit $R \rightarrow \infty$) recovers the moments of the sampling distribution under \hat{F} . This is equivalent to replacing the unknown population moments a_r and μ_r with their sample counterparts $\hat{a}_r(x)$ and $\hat{\mu}_r(x)$.

If we consider as the statistic of interest a weighted average $\bar{x}_v := \sum_{i=1}^n v_i x_i$ for some fixed weights vector $v \in \mathcal{S}^{n-1}$, then the above expressions for the central moments are easily adapted. Considering that $\mathbb{E}(\bar{x}_v) = \mathbb{E}(\sum_{i=1}^n v_i x_i) = \sum_{i=1}^n v_i \mathbb{E}(x_i) = \mathbb{E}(x_i) = a_1$ we have

$$\mu_r(\bar{x}_v) = \mathbb{E}((\bar{x}_v - \mathbb{E}(\bar{x}_v))^r) = \mathbb{E}\left(\left(\sum_{i=1}^n v_i (x_i - a_1)\right)^r\right).$$

For instance, for $r = 2$ we get

$$\begin{aligned} \mu_2(\bar{x}_v) &= \mathbb{E}\left(\sum_{i=1}^n v_i^2 (x_i - a_1)^2 + 2 \sum_{i=1}^n \sum_{j>i}^n v_i v_j (x_i - a_1)(x_j - a_1)\right) \\ &= \sum_{i=1}^n v_i^2 \mathbb{E}(x_i - a_1)^2 = \mu_2 \sum_{i=1}^n v_i^2 \end{aligned}$$

which is equal to $\frac{\mu_2}{n}$ when $\forall i, v_i = \frac{1}{n}$. The double sum drops because $\forall i, \mathbb{E}(x_i - a_1) = 0$.

The above evaluation considers the weights as given and vector x as random. Returning to the formulation of $\hat{\theta}$ as a conditional estimate, the expectation is taken over the distribution of the weights and the double sum no longer drops. Writing $A = (z_i - \bar{z})(z_i - \bar{z})^\top$,

the expectation under $M_{1/n}$ is computed by a summation over all K grid node activation probabilities p_k as

$$\mathbb{E}_{M_{1/n}} \left((\hat{\theta} - \mathbb{E}_{M_{1/n}}(\hat{\theta}))^2 \right) = \mathbb{E}_{M_{1/n}} \left(\left(\sum_{i=1}^n \omega_i (z_i - \bar{z}) \right)^2 \right) = \mathbb{E}_{M_{1/n}} (\omega^\top A \omega) = \sum_{k=1}^K p_k [(\omega^{(k)})^\top A \omega^{(k)}]$$

which can be simplified to a summation over L as

$$\sum_{k=1}^K p_k [(\omega^{(k)})^\top A \omega^{(k)}] = \sum_{l=1}^L \dot{p}_l \sigma_l.$$

σ_l can be seen as a variance within a level set (i.e., within a set of permutations of a partition), i.e.,

$$\sigma_l = \frac{1}{|Q^{(l)}|} \sum_{j \in Q^{(l)}} (\omega^{(j)})^\top A \omega^{(j)} = \frac{1}{|Q^{(l)}|} \sum_{j \in Q^{(l)}} \left(\sum_{i=1}^n q_i^{(j,l)} z_i - \frac{1}{|Q^{(l)}|} \sum_{j \in Q^{(l)}} \sum_{i=1}^n q_i^{(j,l)} z_i \right)^2$$

where $q \in Q^{(l)}$, $Q^{(l)}$ denoting the set of all distinct permutations of partition $\rho^{(l)}$, as defined in Section 4.2. Notice that $\frac{1}{|Q^{(l)}|} \sum_{j \in Q^{(l)}} \sum_{i=1}^n q_i^{(j,l)} z_i = \frac{1}{n} \sum_{i=1}^n z_i = \bar{z}$.

Denoting the unbiased variance of z by $\sigma(z)$, the above further simplifies in terms of the (scaled) partition's norm as

$$\sigma_l = \sigma(z) \left(\|\rho^{(l)}/n\|_2^2 - \frac{1}{n} \right).$$

Therefore,

$$\mathbb{V}(\hat{\theta}) = \sigma(z) \sum_{l=1}^L \dot{p}_l \left(\|\rho^{(l)}/n\|_2^2 - \frac{1}{n} \right).$$

4.3.1 Exact standard error

Let $\bar{\omega} := \mathbb{E}(\omega)$. In either case, $\omega \sim D_1$ or $\omega \sim M_{1/n}$, the centre of mass corresponds to the centroid of the simplex $\bar{\omega} = (1/n, \dots, 1/n)$ with squared 2-norm $\|\bar{\omega}\|_2^2 = \frac{1}{n}$. Note that the sum of the r -th power of the non-negative weights corresponds to $\sum_{i=1}^n \omega_i^r = \|\omega\|_r^r$.

Further, let $m_2(z)$ denote the unbiased sample variance of z , defined in terms of the

biased estimator $\tilde{m}_2(z) = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2$, $\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$, times a bias correction term such that $m_2(z) = \tilde{m}_2(z) \frac{n}{n-1}$. Bear in mind that we treat z to be a fixed vector and that therefore, the variance is not to be understood in a probabilistic sense but as a functional such that $m_2(z)$ is a scalar.

Proposition 9. *The standard error of the exact bootstrap distribution, classical and Bayesian, of statistic $\hat{\theta}$ is given by*

$$SE(\hat{\theta}) = \sqrt{\mu_2(\hat{\theta})} = \sqrt{m_2(z) [\mathbb{E}(\|\omega\|_2^2) - \|\bar{\omega}\|_2^2]}.$$

Proof of proposition 9 under $\omega \sim M_{1/n}$.

First, let us establish that $\mathbb{E}(\omega_i^2) = \frac{2n-1}{n^3}$, $\mathbb{E}(\omega_i \omega_j) = \frac{n-1}{n^3}$ and $\mathbb{E}(\|\omega\|_2^2) = \frac{2n-1}{n^2}$, which

follows directly from the properties of the Multinomial distribution. Then,

$$\begin{aligned}
\mathbb{E} \left((\hat{\theta} - \mathbb{E}(\hat{\theta}))^2 \right) &= \mathbb{E} \left(\left(\sum_{i=1}^n \omega_i z_i - \sum_{i=1}^n \bar{\omega}_i z_i \right)^2 \right) \\
&= \mathbb{E} \left(\left(\sum_{i=1}^n \omega_i (z_i - \bar{z}) \right)^2 \right) \\
&= \sum_{i=1}^n \mathbb{E}(\omega_i^2) (z_i - \bar{z})^2 + 2 \sum_{i=1}^{n-1} \sum_{j>i}^n \mathbb{E}(\omega_i \omega_j) (z_i - \bar{z})(z_j - \bar{z}) \\
&= \frac{1}{n^3} \left((2n-1) \sum_{i=1}^n (z_i - \bar{z})^2 + (n-1) 2 \sum_{i=1}^{n-1} \sum_{j>i}^n (z_i - \bar{z})(z_j - \bar{z}) \right) \\
&= \frac{1}{n^3} \left((2n-1) \sum_{i=1}^n (z_i - \bar{z})^2 + (n-1) \left[\left(\sum_{i=1}^n (z_i - \bar{z}) \right)^2 - \sum_{i=1}^n (z_i - \bar{z})^2 \right] \right) \\
&= \frac{1}{n^3} \left((2n-1) \sum_{i=1}^n (z_i - \bar{z})^2 + (n+1) \sum_{i=1}^n (z_i - \bar{z})^2 \right) \\
&= \frac{1}{n^2} \sum_{i=1}^n (z_i - \bar{z})^2
\end{aligned}$$

and

$$\begin{aligned}
m(z) \left(\mathbb{E}(\|\omega\|_2^2) - \|\bar{\omega}\|_2^2 \right) &= \tilde{m}_2(z) \frac{n}{n-1} \left(\frac{2n-1}{n^2} - \frac{1}{n} \right) \\
&= \left(\frac{\tilde{m}_2(z)}{n-1} - \frac{1}{n^2-n} \right) \\
&= \frac{1}{n^2} \sum_{i=1}^n (z_i - \bar{z})^2
\end{aligned}$$

thus recovering the well known standard error of the sample mean $SE(\bar{z}) = \sqrt{\frac{\tilde{m}_2(z)}{n}}$. □

For the BB, the expectation can be derived from the property of the Dirichlet distribution that the marginals have a Beta(α, β) distribution with moments $\mathbb{E}(X^r) = \prod_{v=0}^{r-1} \frac{\alpha+v}{\alpha+\beta+v}$, i.e.,

$$\mathbb{E}_{D_\alpha}(\|\omega\|_r^r) = \mathbb{E}_{D_\alpha}\left(\sum_{i=1}^n \omega_i^r\right) = \sum_{i=1}^n \mathbb{E}_{D_\alpha}(\omega_i^r) = \sum_{i=1}^n \prod_{v=0}^{r-1} \frac{\alpha_i + v}{\alpha_0 + v}. \quad (4.4)$$

Hence, the standard error of the BB distribution for $\hat{\theta}$ can be expressed in terms of the parameters of the posterior Dirichlet distribution as

$$\mu_2^{BB}(\hat{\theta}) = m_2(z) \left(\mathbb{E}_{D_\alpha}(\|\omega\|_2^2) - \sum_{i=1}^n \bar{\omega}^2 \right) = m_2(z) \left(\sum_{i=1}^n \prod_{v=0}^1 \frac{\alpha_i + v}{\alpha_0 + v} - \sum_{i=1}^n \frac{\alpha_i^2}{n^2} \right).$$

In the symmetric case, $\alpha = \mathbf{1}$ (implying a Beta(1, $n - 1$) for the marginals), and

$$\mathbb{E}_{D_1}(\|\omega\|_2^2) = n \prod_{v=0}^1 \frac{1 + v}{n + v} = \frac{2}{n + 1}.$$

Therefore,

$$\begin{aligned} \mu_2^{BB}(\hat{\theta}) &= m_2(z) \left(\mathbb{E}_{D_\alpha}(\|\omega\|_2^2) - \|\bar{\omega}\|_2^2 \right) \\ &= \tilde{m}_2(z) \frac{n}{n-1} \left(\frac{2}{n+1} - \frac{1}{n} \right) \\ &= \frac{1}{n(n+1)} \sum_{i=1}^n (z_i - \bar{z})^2. \end{aligned}$$

Proposition 9 is not limited to the bootstraps of [Efron \(1979\)](#) and [Rubin \(1981\)](#) but extends to any parametrization of the respective models, i.e., $\omega \sim M_\pi$ or $\omega \sim D_\alpha$. A proof is given in [Appendix 4C](#). It depends on an adjustment of the bias correction term which will be discussed in [Section 4.4](#) when we extend the respective bootstrap models to reflect prior beliefs.

Corollary 10. *The difference between the standard errors of the BB and the classical bootstrap results from the difference in the expected squared 2-norm of the corresponding weights:*

$$\mu_2^B(\hat{\theta}) - \mu_2^{BB}(\hat{\theta}) = m_2(z) \left(\mathbb{E}_{M_{1/n}}(\|\omega\|_2^2) - \mathbb{E}_{D_1}(\|\omega\|_2^2) \right).$$

Given that $\mathbb{E}_{M_{1/n}}(\|\omega\|_2^2) \geq \mathbb{E}_{D_1}(\|\omega\|_2^2)$ because $\frac{2n-1}{n^2} \geq \frac{2}{n+1}$ it follows that $\mu_2^B(\hat{\theta}) \geq \mu_2^{BB}(\hat{\theta})$ for all n . The rate of convergence (or divergence) depends on the data z , respectively on its biased second sample moment $\tilde{m}_2(z)$ since $\mu_2^B(\hat{\theta}) - \mu_2^{BB}(\hat{\theta}) = \tilde{m}_2(z)(\frac{1}{n} - \frac{1}{n+1}) = \tilde{m}_2(z)\frac{1}{n^2+n}$.

Corollary 11. *The mapping $\lambda(n) = \frac{(n-1)^2}{n(n+1)}$ ensures that $\mu_2^B(\hat{\theta}) = \mu_{2,\lambda}^{BB}(\hat{\theta})$ for some version of the Bayesian bootstrap which uses a symmetric Dirichlet distribution parametrized by $\alpha = \mathbf{1}\lambda(n)$.*

Proof.

$$\begin{aligned} \mu_2^B(\hat{\theta}) &= \mu_{2,\mathbf{1}\lambda(n)}^{BB}(\hat{\theta}) && \Leftrightarrow \\ \mathbb{E}_{M_{1/n}}(\|\omega\|_2^2) &= \mathbb{E}_{D_{\mathbf{1}\lambda(n)}}(\|\omega\|_2^2) && \Leftrightarrow \\ \frac{2n-1}{n^2} &= \frac{1+\lambda(n)}{1+\lambda(n)n} \\ \Rightarrow \lambda(n) &= \frac{(n-1)^2}{n(n+1)}. \end{aligned}$$

□

4.3.2 Exact higher central moments

Let $m_3(z) = \tilde{m}_3(z)\frac{n^2}{(n-1)(n-2)}$, i.e., the third biased central moment $\tilde{m}_3(z) = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^3$ times the bias correction term for the third central moment.

Proposition 12. *The third central moment of the exact bootstrap distribution, classical and Bayesian, of statistic $\hat{\theta}$ is given by*

$$\mu_3(\hat{\theta}) = m_3(z) \left[\mathbb{E}(\|\omega\|_3^3) - 3\mathbb{E}(\|\omega\|_2^2)\|\bar{\omega}\|_2^2 + 2\|\bar{\omega}\|_3^3 \right].$$

For the classical bootstrap we have $\mathbb{E}_{M_{1/n}}(\|\omega\|_3^3) = \frac{5n^2-6n+2}{n^4}$ and hence,

$$\begin{aligned}
\mu_3^B(\hat{\theta}) &= m_3(z) \left[\mathbb{E}_{M_{1/n}}(\|\omega\|_3^3) - 3\mathbb{E}_{M_{1/n}}(\|\omega\|_2^2)\|\bar{\omega}\|_2^2 + 2\|\bar{\omega}\|_3^3 \right] \\
&= \tilde{m}_3(z) \frac{n^2}{(n-1)(n-2)} \left[\frac{5n^2 - 6n + 2}{n^4} - 3\frac{2n-1}{n^2} \frac{1}{n} + 2\frac{1}{n^2} \right] \\
&= \frac{1}{n^3} \sum_{i=1}^n (z_i - \bar{z})^3 = \frac{\tilde{m}_3(z)}{n^2}.
\end{aligned}$$

Plugging in the previously obtained value for the expected squared 2-norm when $\omega \sim D_1$ and further recognizing that $\mathbb{E}(\|\omega\|_3^3) = n \prod_{v=0}^2 \frac{1+v}{n+v} = \frac{6}{(n+1)(n+2)}$, one finds the third central moment of the BB distribution of $\hat{\theta}$ to be equal to

$$\begin{aligned}
\mu_3^{BB}(\hat{\theta}) &= m_3(z) \left(\mathbb{E}_{D_1}(\|\omega\|_3^3) - 3\mathbb{E}_{D_1}(\|\omega\|_2^2)\|\bar{\omega}\|_2^2 + 2\|\bar{\omega}\|_3^3 \right) \\
&= \tilde{m}_3(z) \frac{n^2}{(n-1)(n-2)} \left(\frac{6}{(n+1)(n+2)} - 3\frac{2}{n+1} \frac{1}{n} + 2\frac{1}{n^2} \right) \\
&= \frac{2}{n(n+1)(n+2)} \sum_{i=1}^n (z_i - \bar{z})^3.
\end{aligned}$$

As for the standard error, proposition 17 holds for any parametrization of the Multinomial or the Dirichlet models, as long as the bias correction term in $m_3(z)$ is correctly adapted (see Section 4.4.1), implying the identity:

$$\begin{aligned}
\mathbb{E} \left((\hat{\theta} - \mathbb{E}(\hat{\theta}))^3 \right) &= \sum_{i=1}^n \mathbb{E}(\omega_i^3) b_i^3 + 3 \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathbb{E}(\omega_i^2 \omega_j) b_i^2 b_j + 6 \sum_{i=1}^{n-2} \sum_{j>i}^{n-1} \sum_{k>j}^n \mathbb{E}(\omega_i \omega_j \omega_k) b_i b_j b_k \\
&= m_3(z) \left(\mathbb{E}(\|\omega\|_3^3) - 3 \sum_{i=1}^n \bar{\omega}_i \mathbb{E}(\omega_i^2) + 2\|\bar{\omega}\|_3^3 \right)
\end{aligned}$$

where $b_i = (z_i - \bar{z})$, $i = 1, \dots, n$.⁸

⁸Under $\omega \sim D_\alpha$ one has:

$$\mathbb{E}(\omega_i^3) = \frac{(\alpha_i + 2)(\alpha_i + 1)\alpha_i}{(\alpha_0 + 2)(\alpha_0 + 1)\alpha_0}; \quad \mathbb{E}(\omega_i^2 \omega_j) = \frac{(\alpha_i + 1)\alpha_i \alpha_j}{(\alpha_0 + 2)(\alpha_0 + 1)\alpha_0}; \quad \mathbb{E}(\omega_i \omega_j \omega_k) = \frac{\alpha_i \alpha_j \alpha_k}{(\alpha_0 + 2)(\alpha_0 + 1)\alpha_0}.$$

Corollary 13. *The difference in third central moments of the classical bootstrap and the BB depends on $\mathbb{E}_{M_{1/n}}(\|\omega\|_3^3) - \mathbb{E}_{D_1}(\|\omega\|_3^3) + (\mathbb{E}_{M_{1/n}}(\|\omega\|_2^2) - \mathbb{E}_{D_1}(\|\omega\|_2^2))$.*

From the closed-form expressions above one finds that, for any finite n ,

$$\mu_3^B(\hat{\theta}) - \mu_3^{BB}(\hat{\theta}) = \tilde{m}_3(z) \left(\frac{1}{n^2} - \frac{2}{(n+1)(n+2)} \right) > 0.$$

Unlike the second and third central moments, where the distinction between biased and unbiased moment estimators is a constant factor unaffected by the data, higher-order moments require the computation of combinations of lower moments which results in systems of equations that need to be solved. However, any higher order central moment of the exact bootstrap distributions of linear statistics, Bayesian and classical, under flat parametrization of the Dirichlet or the scaled Multinomial model, can in principle be computed using the recursive schemes presented by [Cales et al. \(2021\)](#) and [Angelova \(2012\)](#) respectively⁹. [Cales et al. \(2021\)](#) report average computation time for moments up to $r = 20$ and sample size $n = 10'000$ to be roughly 0.05 seconds. We replicate their algorithm for central moment computation in [Appendix 4B](#) adapted to our notation and correcting a small error.

4.3.3 The moment problem

Since the moments of the exact bootstrap distributions can be computed in several cases, we may attempt to address the inverse problem of reconstructing the bootstrap distributions from a finite collection of moments. This poses a probabilistic instance of the (truncated) moment problem. Engaging in a rigorous discussion on the moment problem is beyond the scope of the current study. Nonetheless, we aim to provide a perspective for future research in this direction. There are several approaches to compute approximants of density functions from moments (or cumulants). Examples include the use of Pearson or Johnson curves, series expansions of the type of Gram-Charlier, Edgeworth or Cornish-Fisher, saddlepoint approximations or the maximum entropy principle. We recommend consulting [John et al. \(2007\)](#) for more information on these methods and [Provost \(2005\)](#) for a practically oriented overview.

⁹Effectively, the computation is limited by physical storage capabilities since it involves enumerating all partitions of the integer moment degrees.

The upcoming practical application serves to motivate further exploration into moment-based bootstrap inference.

4.3.4 Numerical example revisited

Let us revise the numerical example given by [Bertsimas and Sturt \(2020\)](#) presented in Section 4.2.2 concerning the estimation of the 0.025-quantile of the distribution of $\hat{\theta}$. Re-running the BB procedure 100 times with $R = 10^6$ produced quantile estimates ranging from 33.42 to 43.43 with mean 38.55942, highlighting the significance of the randomization error. Recall that the true quantile, which we found using Varsi’s algorithm, is 38.30214. We now approximate the quantile based on the exact central moments using Algorithm 11 (see Appendix 4B) and making use of the package `DPQutils` from the open source statistical programming language R which implements the Cornish-Fisher expansion to directly approximate the quantile of interest from moments. The function first requires a conversion of the central moments to cumulants which is straightforward. The function returns a quantile value of 38.15213 based on the first 4 central moments and of 38.30071 based on the first 20 central moments.

# of moments included	2	3	4	5	...	20
0.025-quantile estimate	27.58902	37.40548	38.15213	38.33664	...	38.30071

Table 4.3: Quantile estimates based on the Cornish-Fisher expansion using the first 2 to 20 exact central moments of the BB distribution.

4.4 Informative priors

This Section expands upon the earlier analysis of the bootstrap methods to incorporate prior information. For instance, we can contemplate sampling probabilities that deviate from uniformity to express the belief that certain data elements hold more significance than others in computing the statistic of interest. Instead of constraining the subjective information to a model of z , we adhere to the viewpoint that the data are fixed and, instead, manipulate the distribution of the weights.

We explore four types of modifications to the weight distribution, enabling the integration of prior beliefs: Asymmetry, concentration, constrained domain, and mixtures.

4.4.1 Asymmetry

Asymmetry is encoded by non-uniform probabilities in the scaled Multinomial distribution for the classical bootstrap and by a non-flat parameter vector of the Dirichlet distribution in the case of the BB. By deviating from uniformity in the initial parameter elements imparts prior information to either procedures which is why we use the term informative prior for the Bayesian as well as the frequentist bootstrap alike. The non-uniformity in the parameter elements then breaks the symmetry in the distribution of the weights.

In the classical bootstrap, unequal probabilities in the sampling procedure alter the initially determined grid node activation probabilities. In the extreme scenario where the probability vector of the scaled Multinomial distribution contains n distinct elements, each grid node possesses a unique activation probability, causing the probability level sets established in Section 4.2.3 to completely disappear.

For the BB, the distribution becomes skewed, with more mass concentrated towards the vertices corresponding to the lowest α_i . In both procedures, the center of mass deviates from the centroid of the simplex (in the classical bootstrap, generally $\bar{\omega} \notin G_n$).

Propositions 9 and 17 can be extended to accommodate asymmetry by substituting the estimates of the sample central moments of z , denoted as m_r , with weighted central moments and bias correction terms that account for the effective sample size.¹⁰

Recall that $\bar{\omega} := \mathbb{E}(\omega)$ which is known. The biased sample weighted central moments are defined in terms of the weighted mean $\bar{z}_{\bar{\omega}} = \frac{1}{V_1} \sum_{i=1}^n \bar{\omega}_i z_i$ and the sum of the weights $V_1 = \sum_{i=1}^n \bar{\omega}_i$, as follows:

$$\tilde{m}_r(z, \bar{\omega}) = \frac{1}{V_1} \sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z}_{\bar{\omega}})^r; \quad r = 2, 3, \dots$$

The sample-size bias corrected weighted central moments are derived assuming independent measurements and weights. They are defined in terms of sample estimators as follows (Rimoldini, 2014):

¹⁰as the weightings become more uneven, the biased weighted variance tends to decrease. However, increased unevenness in weightings makes the weighted mean more unstable, as it is increasingly influenced by one or a few points with high weights. To address the statistical uncertainty in the mean, a bias correction larger than the standard $\frac{n}{n-1}$ factor is necessary to compensate for the reduced degrees of freedom. This correction involves replacing n with the effective sample size $n_{eff} = \frac{\sum_i (v_i)^2}{\sum_i v_i^2}$. Consequently, $\frac{n_{eff}}{n_{eff}-1}$ is expressed as $\frac{(\sum_i v_i)^2}{(\sum_i v_i)^2 - \sum_i v_i^2}$ or equivalently $\frac{1}{1 - \sum_i v_i^2}$ for $v \in \mathcal{S}^{n-1}$.

$$\begin{aligned}
V_r &= \sum_{i=1}^n \bar{\omega}_i^r \\
m_2(z, \bar{\omega}) &= \frac{V_1^2}{V_1^2 - V_2} \tilde{m}_2(z, \bar{\omega}) \\
m_3(z, \bar{\omega}) &= \frac{V_1^3}{V_1^3 - 3V_1V_2 + 2V_3} \tilde{m}_3(z, \bar{\omega}).
\end{aligned}$$

When dealing with the fourth and higher moments, the bias correction terms become excessively intricate¹¹, making it challenging to provide a straightforward representation in terms of (expected) norms. However, for the second and third central moments of the bootstrap distribution, where weights are consistently non-negative, denoting $V_r = \sum_{i=1}^n \bar{\omega}_i^r = \|\bar{\omega}\|_r^r$, we can derive the following generic representations.

Proposition 14.

$$\mu_2(\hat{\theta}) = \tilde{m}_2(z, \bar{\omega}) \frac{\mathbb{E}(\|\omega\|_2^2) - \|\bar{\omega}\|_2^2}{1 - \|\bar{\omega}\|_2^2}.$$

Proposition 15.

$$\mu_3(\hat{\theta}) = \tilde{m}_3(z, \bar{\omega}) \frac{\mathbb{E}(\|\omega\|_3^3) - 3 \sum_{i=1}^n \bar{\omega}_i \mathbb{E}(\omega_i^2) + 2\|\bar{\omega}\|_3^3}{(\|\bar{\omega}\|_1^3 - 3\|\bar{\omega}\|_1 \|\bar{\omega}\|_2^2 + 2\|\bar{\omega}\|_3^3)}.$$

The computation of the expectation of the r -th power of the r -norm works as detailed in Section 4.3. Notice that $\bar{\omega}$ no longer constitutes the centroid of the simplex; rather, it may assume any position on the simplex. Consequently, its r -th power r -norm is no longer equal to $n^{-(r-1)}$ for a given n .

An important aspect is whether $\sum_{i=1}^n \alpha_i = n$. If this condition holds, the impact of prior belief on the deviation of the BB distribution under prior information compared to the BB distribution under a non-informative prior depends solely on the difference between the weighted and unweighted biased sample central moments of the data vector. For the second and third central moments the differences satisfy:

¹¹For the fourth weighted central moment one has

$$m_4(z, \bar{\omega}) = \frac{V_1^2(V_1^4 - 3V_1^2V_2 + 2V_1V_3 + 3V_2^2 - 3V_4)}{(V_1^2 - V_2)(V_1^4 - 6V_1^2V_2 + 8V_1V_3 + 3V_2^2 - 6V_4)} \tilde{m}_4(z, \bar{\omega}) - \frac{3V_1^2(2V_1^2V_2 - 2V_1V_3 - 3V_2^2 + 3V_4)}{(V_1^2 - V_2)(V_1^4 - 6V_1^2V_2 + 8V_1V_3 + 3V_2^2 - 6V_4)} \tilde{m}_2^2(z, \bar{\omega}).$$

$$\begin{aligned}\mu_{2,\bar{\omega}}(\hat{\theta}) - \mu_2(\hat{\theta}) &\propto \tilde{m}_2(z, \bar{\omega}) - \tilde{m}_2(z) \\ \mu_{3,\bar{\omega}}(\hat{\theta}) - \mu_3(\hat{\theta}) &\propto \tilde{m}_3(z, \bar{\omega}) - \tilde{m}_3(z).\end{aligned}$$

Or equivalently,

$$\begin{aligned}\frac{\mu_{2,\bar{\omega}}(\hat{\theta})}{\mu_2(\hat{\theta})} &= \frac{\tilde{m}_2(z, \bar{\omega})}{\tilde{m}_2(z)} \\ \frac{\mu_{3,\bar{\omega}}(\hat{\theta})}{\mu_3(\hat{\theta})} &= \frac{\tilde{m}_3(z, \bar{\omega})}{\tilde{m}_3(z)}.\end{aligned}$$

Introducing prior information by altering the symmetry of the distribution of weights is operationally straightforward, involving a simple adjustment of the hyperparameter. When $\sum_{i=1}^n \alpha_i = n$, the influence on central moments is intuitive, as it is proportional to the disparity between weighted and unweighted sample central moments. However, when $\sum_{i=1}^n \alpha_i \neq n$, an additional effect comes into play, related to the concentration of weights around the mode. This represents the second approach for integrating prior information into the bootstrap methodology, which we proceed to investigate subsequently.

4.4.2 Concentration

The central moments of the bootstrap distribution are influenced by the expected r -th power of the distance of the weights to their mean. Modifying the distribution of weights around the mean, either by concentrating or dispersing mass, has a direct impact on the bootstrap distribution.

Consider the case of the BB distribution with a symmetric but non-flat Dirichlet parameter $\alpha = \mathbf{1}\lambda$. Here, $\lambda \geq 0$ is a scaling parameter. For $\lambda > 1$, the distribution becomes more concentrated around the centroid of the simplex, while $\lambda < 0$ pushes weights outward towards the faces, edges, and vertices of the simplex, relative to the uniform density implied by the flat base case ($\lambda = 1$).

As the parameter λ approaches infinity (and consequently α_0 tends to infinity), the mass converges to a singular point. This results in a degenerate distribution characterized

by the Dirac delta function centered at the centroid. Conversely, as λ tends to zero (leading to α_0 approaching zero), the weight distribution becomes sharply focused around the $n - 1$ possible Dirac delta distributions, each centered on one of the corners of the simplex.

The computation of $\mathbb{E}_{D_{1\lambda}}(\|\omega\|^{r^r})$ follows the expression in equation 4.4. In the cases of $r = 2$ and $r = 3$, this results to

$$\mathbb{E}_{D_{1\lambda}}(\|\omega\|_2^2) = \frac{1 + \lambda}{1 + n\lambda} \quad \text{and} \quad \mathbb{E}_{D_{1\lambda}}(\|\omega\|_3^3) = \frac{\lambda^2 + 3\lambda + 2}{n^2\lambda^2 + 3n\lambda + 2}.$$

Expressed in terms of $\log \lambda$, the expectation exhibits the characteristics of a reverse asymmetric sigmoidal function.

The parameter λ can be interpreted as a measure of prior confidence influencing the uncertainty in the estimator $\hat{\theta}(z)$. When $\lambda > 1$, uncertainty is reduced, whereas for $\lambda < 1$, uncertainty is increased. In the extreme cases where $\lambda \rightarrow \infty$ and $\lambda \rightarrow 0$, all bootstrap central moments converge to zero in the former and to the sample central moments of z in the latter case.

In the classical bootstrap, a similar impact is achieved by modifying the number of draws from the data vector, resulting in the m -out-of- n bootstrap (Bickel et al., 1997). As detailed in Section 4.2.3, adjusting m gives rise to a new grid structure $G_{n,m}$ with either fewer nodes ($m < n$) compared to the base grid G_n induced by the classical n out of n bootstrap. In such cases, nodes are positioned closer to the faces, edges, and vertices of the simplex, indicating that $\mathbb{E}_{M_{1/n,m}}(\|\omega\|_2^2) > \mathbb{E}_{M_{1/n}}(\|\omega\|_2^2)$. Conversely, when $G_{n,m}$ has more nodes ($m > n$) than G_n , nodes concentrate nearer to the centroid of the simplex, resulting in $\mathbb{E}_{M_{1/n,m}}(\|\omega\|_2^2) < \mathbb{E}_{M_{1/n}}(\|\omega\|_2^2)$.

The exact expected r -th r -norm can be computed as in Section 4.2.3 from the node activation probabilities of grid $G_{n,m}$ and partitions ρ , given by $\mathbb{E}_{M_{1/n,m}}(\|\omega\|_r^r) = \sum_{i=1}^n \dot{p}^{(i)} \|\rho^{(i)}/n\|_r^r$. For instance, when $m \leq n$, the calculation for $\mathbb{E}_{M_{1/n,m}}(\|\omega\|_2^2) = \frac{m+n-1}{mn}$.

In either case, BB with $\lambda \neq 1$ or classical bootstrap with $m \neq n$, the impact of the prior view on the posterior central moments is proportional to the difference in the expected r -th power of the r -norm. Let $\omega \sim D_{\mathbf{1}}$ and $\omega' \sim D_{1\lambda}$ (or $\omega \sim M_{1/n}$ and $\omega' \sim M_{1/n,m}$). Then,

$$\begin{aligned}\mu'_2(\hat{\theta}) - \mu_2(\hat{\theta}) &\propto \mathbb{E}(\|\omega'\|_2^2) - \mathbb{E}(\|\omega\|_2^2) \\ \mu'_3(\hat{\theta}) - \mu_3(\hat{\theta}) &\propto \mathbb{E}(\|\omega'\|_3^3) - \mathbb{E}(\|\omega\|_3^3) + \mathbb{E}(\|\omega'\|_2^2) - \mathbb{E}(\|\omega\|_2^2)\end{aligned}$$

Further notice that setting $\lambda = \frac{1-m}{1-n} + \frac{m-1}{n(1-n)} = \frac{m-1}{n}$ ensures that $\mu_{2,m}^B(\hat{\theta}) = \mu_{2,\lambda}^{BB}(\hat{\theta})$.

An alternative, non-analytical approach to introducing prior uncertainty into the bootstrap procedure (both classical and Bayesian) is to initially sample weights uniformly and then (non-linearly) push weights toward the vertices by applying $\omega'_i = \omega_i^q / \sum_{i=1}^n \omega_i^q$. This is based on the observation that as $q \rightarrow \infty$, $\|\omega\|_q \rightarrow \max_i \omega_i$, meaning that weights converge to a vector of all zeros except for a one at the position of the initially largest weight. The expected distance measure becomes $\mathbb{E}(\|\omega'\| r^r) = \mathbb{E}(\|\frac{\omega^q}{\|\omega\|_q^q}\| r^r) = \mathbb{E}\left(\sum_{i=1}^n \left(\frac{\omega_i^q}{\sum_{i=1}^n \omega_i^q}\right)^r\right) = \sum_{i=1}^n \mathbb{E}(\omega_i^{qr} (\sum_{i=1}^n \omega_i^q)^{-r})$.

4.4.3 Constrained domain

An additional method for integrating prior information entails imposing restrictions on the distribution support of the weights. For the classical bootstrap, this would mean to discard certain nodes of grid G_n . For the BB, the restriction we consider here involves subsetting the simplex through linear constraints to form a polytope (i.e., slicing the simplex with hyperplanes).

Constraining the simplex means that we can no longer make use of the Dirichlet distribution as it is a distribution over the entire simplex. However, in specific instances, it is feasible to manipulate the domain to create a variation of the Dirichlet model that facilitates a straightforward extension of propositions 14 and 15. One such model is the shadow Dirichlet distribution, proposed by Frigiyk et al. (2010). If $\omega \sim D_\alpha$ and \mathcal{T} an $n \times n$ left-stochastic matrix (i.e., each column sums to one) of full rank, then $R\omega =: \tilde{\omega} \sim SD(\mathcal{T}, \alpha)$, where SD denotes the shadow Dirichlet model with pdf

$$f(\tilde{\omega}) = \frac{1}{|\det(\mathcal{T})|B(\alpha)} \prod_{i=1}^n (\mathcal{T}^{-1}\tilde{\omega})_i^{\alpha_i-1}.$$

Frigiyk et al. (2010) describe how to construct a shadow Dirichlet for three types of

restricted domains: regularized pmf's, bounded variation between pmf components, and monotonic pmfs.

Figure 4.2 visualizes the *shadow* of a flat Dirichlet distribution restricted by a monotonic \mathcal{T} which has k -th column $[0 \dots 0 \ 1/(n-k+1) \dots 1/(n-k+1)]^\top$. The resulting constrained weights are uniformly distributed over the restricted domain given by the orange surface. Within this domain it holds that $\omega_1 > \omega_2 > \dots > \omega_n$.

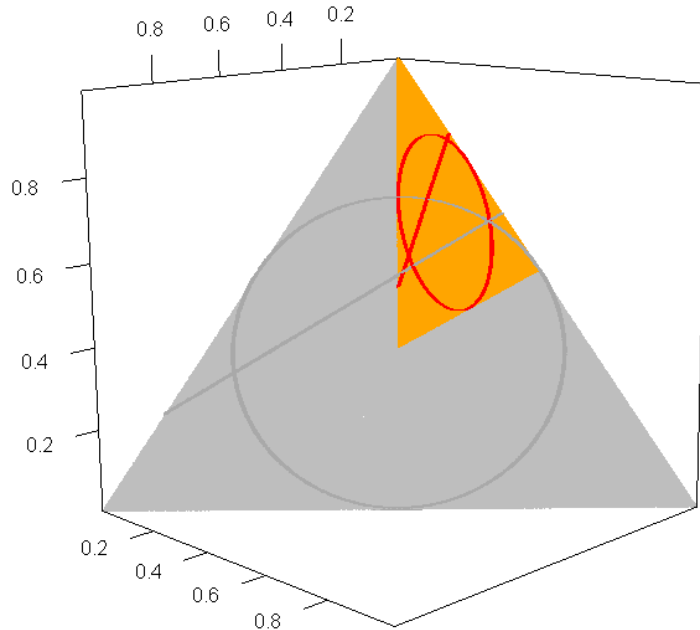


Figure 4.2: Grey surface: $\omega \sim D_\alpha$. Orange surface: $\omega \sim SD(\mathcal{T}, \mathbf{1})$. Grey circle: $\Omega := \{\omega \in \mathcal{S}^{n-1} : \|\omega\|_2^2 = \mathbb{E}(\|\omega\|_2^2)\}$. Red ellipsoid: $\mathcal{T}\Omega$. Grey line: Hyperplane with normal z . Red line: Hyperplane with normal $\mathcal{T}^\top z$.

If we were to solve for the standard error under the shadow Dirichlet model, i.e., $\mathbb{E}_{SD} \left(\left(\hat{\theta} - \mathbb{E}_{SD}(\hat{\theta}) \right)^2 \right)$, as we did in Section 4.3, we could proceed with the expressions $\mathbb{E}_{SD}(\tilde{\omega}_i \tilde{\omega}_j) = (\mathcal{T} Cov(\omega) \mathcal{T}^\top)_{ij} + \left(\mathcal{T} \frac{\alpha}{\alpha_0} \right)_i \left(\mathcal{T} \frac{\alpha}{\alpha_0} \right)_j$ and $\mathbb{E}_{SD}(\tilde{\omega}_i^2) = (\mathcal{T} Cov(\omega) \mathcal{T}^\top)_{ii} + \left(\mathcal{T} \frac{\alpha}{\alpha_0} \right)_i^2$. However, rather than going through the algebra it suffices to replace the variance of the data in Proposition 9 with the weighted variance of transformed data, i.e., replacing i.e., $m_2(z, \omega)$ with

$$m_2(M, z, \omega) = \sum_{j=1}^m \bar{\omega}_j \left((\mathcal{T}^\top z)_j - \mathcal{T}^\top \bar{z} \right)^2 \left(\frac{1}{1 - \sum_{j=1}^m \bar{\omega}_j^2} \right).$$

Notice that the columns in \mathcal{T} represent the vertices of the constrained simplex which we denote by \mathcal{S}_T .

Proposition 16.

$$\begin{aligned}\mu_2(\hat{\theta}) &= m_2(\mathcal{T}, z, \omega) \left(\mathbb{E}(\|\omega\|_2^2) - \|\mathbb{E}(\omega)\|_2^2 \right) \\ &= \left(\sum_{j=1}^m \bar{\omega}_j \left((\mathcal{T}^\top z)_j - \mathcal{T}^\top \bar{z} \right)^2 \right) \frac{\mathbb{E}(\|\omega\|_2^2) - \|\bar{\omega}\|_2^2}{1 - \|\bar{\omega}\|_1^2}.\end{aligned}$$

Proposition 17.

$$\begin{aligned}\mu_3(\hat{\theta}) &= m_3(\mathcal{T}, z, \omega) \left(\mathbb{E}(\|\omega\|_3^3) - 3\mathbb{E}(\|\omega\|_2^2)\|\bar{\omega}\|_2^2 + 2\|\bar{\omega}\|_3^3 \right) \\ &= \left(\sum_{j=1}^m \bar{\omega}_j \left((\mathcal{T}^\top z)_j - \mathcal{T}^\top \bar{z} \right)^3 \right) \frac{\mathbb{E}(\|\omega\|_3^3) - \|\bar{\omega}\|_3^3 - 3 \sum_{i=1}^n \bar{\omega}_i \mathbb{E}(\omega_i^2) + 3\|\bar{\omega}\|_3^3}{(\|\bar{\omega}\|_1^3 - 3\|\bar{\omega}\|_1\|\bar{\omega}\|_2^2 + 2\|\bar{\omega}\|_3^3)}.\end{aligned}$$

When $\tilde{\omega} \sim \text{SD}(\mathcal{T}, \mathbf{1})$, i.e., $\mathcal{T}^{-1}\tilde{\omega} = \omega$ follows a flat Dirichlet distribution, then Varsi's algorithm (10) can be used to compute the full bootstrap distribution exactly.

Proposition 18. Let $H(\mathcal{T}, z, \gamma) := \{h \in \mathbb{R}^n | z^\top \mathcal{T}h \leq \gamma\}$ denote a half-space with normal $\mathcal{T}^\top z$ and scalar γ . Then $C_{\mathcal{T}}(\gamma) = \mathbb{P}(\hat{\theta}(\mathcal{T}, z, \omega) \leq \gamma) = \frac{V(\mathcal{S}^{n-1} \cap H(\mathcal{T}, z, \gamma))}{V(\mathcal{S}^{n-1})}$.

Proof.

$$\begin{aligned}C_{\mathcal{T}}(\gamma) &= \mathbb{P}(\hat{\theta} \leq \gamma) = \int_{\mathcal{S}_{\mathcal{T}}} f(\omega) \mathbb{1}\{z^\top \omega \leq \gamma\} d\omega \\ &= \int_{\mathcal{S}_{\mathcal{T}}} \frac{1}{|\det(\mathcal{T})| \mathbb{B}(\alpha)} \prod_{i=1}^n \omega_i^{\alpha_i-1} \mathbb{1}\{z^\top \omega \leq \gamma\} d\omega \\ &= \frac{1}{|\det(\mathcal{T})|} \frac{1}{\int_{\mathcal{S}^{n-1}} \prod_{i=1}^n \omega_i^{\alpha_i-1} d\omega} \int_{\mathcal{S}_{\mathcal{T}} \cap H(z, \gamma)} \prod_{i=1}^n \omega_i^{\alpha_i-1} d\omega \\ &= \frac{1}{|\det(\mathcal{T})|} \frac{V(\mathcal{S}_{\mathcal{T}} \cap H(z, \gamma))}{V(\mathcal{S}^{n-1})} \\ &= \frac{V(\mathcal{S}_{\mathcal{T}} \cap H(z, \gamma))}{V(\mathcal{S}_{\mathcal{T}})} \frac{|\det(\mathcal{T})|^{-1}}{|\det(\mathcal{T})|^{-1}} \\ &= \frac{V(\mathcal{S}^{n-1} \cap H(\mathcal{T}, z, \gamma))}{V(\mathcal{S}^{n-1})}.\end{aligned}$$

□

4.4.4 Mixtures

Our findings so far can be extended by examining weights that adhere to a mixture of previously discussed distributions. In the case of finite mixtures, let us consider a vector v

with a finite integer J elements, where $v_j > 0$ and $\sum_{j=1}^J v_j = 1$. The central moments of the distribution of the statistic $\hat{\theta}$ based on weights $\omega \sim \text{Mix}(v)$ can be expressed as follows:

$$\mu_r^{\text{Mix}(v)}(\hat{\theta}) = \sum_{j=1}^J v_j \mu_r^{(j)}(\hat{\theta}).$$

For instance, for $r = 2$, one has

$$\mu_2^{\text{Mix}(v)}(\hat{\theta}) = \sum_{j=1}^J v_j m_2(z, \omega^{(j)}) \left(\mathbb{E}(\|\omega^{(j)}\|_2^2) - \|\mathbb{E}(\omega^{(j)})\|_2^2 \right)$$

where $\omega^{(j)}$ follows a different distribution for each $j = 1, \dots, J$.

Taking $v(a)$ to be a probability density and $a \in \mathcal{A}$ it is thinkable to extend the idea to compound distributions

$$\mu_r^{\text{Mix}(v(a))}(\hat{\theta}) = \int_{\mathcal{A}} v(a) \mu_r^{(a)}(\hat{\theta}) da.$$

4.5 Applications

We next explore three statistical procedures commonly subjected to bootstrap applications: regression, eigenvalue estimation, and nonparametric kernel density estimation. The reader will certainly find more examples. In the first application we want to learn about the distribution of the regression coefficients. In the second case we are interested in the distributon of eigenvalues and in the third case we show how to use our results on the bootstrap for density approximation based on kernel methods. The examples are presented from a procedural standpoint without delving into a detailed analysis of their properties, as such an analysis would exceed the scope of this article.

By utilizing either Varsi's algorithm for obtaining the exact bootstrap distribution or the exact central moments based on the norms of the weights to approximate the bootstrap distribution via the Cornish-Fisher expansion, we significantly reduce computational

complexity. Moreover, these approaches eliminate the need for sampling and fully eradicate randomization error.

4.5.1 Regression

The bootstrap has been proposed as a data-driven alternative for conducting inference in regression, offering a departure from some of the conventional assumptions and asymptotic arguments.

Consider the linear regression model

$$y = X\beta + \varepsilon \tag{4.5}$$

with response variable $y = \{y_i\}_{i=1}^n$, covariates $X = \{x_i\}_{i=1}^n$ where $x_i = (x_{i,1}, \dots, x_{i,d})$, error term $\varepsilon = \{\varepsilon_i\}_{i=1}^n$ and $\beta = (\beta_1, \dots, \beta_d)^\top$ the unknown parameter of interest. There are two main classes of bootstrap regressions distinguished by their resampling plans: those acting on the data and those acting on the residuals.

In the former case, where manipulation is made on the data, one resamples the rows of the observed data matrix (y, X) , i.e., the dependent variable and the independent variables are always selected in pairs, hence the name *pairs* bootstrap (Freedman, 1981). Ordinary least squares (OLS) estimates are then computed on the resampled data as $\hat{\beta}^* = (X^{*\top} X^*)^{-1} X^{*\top} y^*$. Alternatively, rather than by resampling we may introduce random weights ω stored in a diagonal matrix W and write $\tilde{\beta} = (X^\top W X)^{-1} X^\top W y = \beta + (X^\top W X)^{-1} X^\top W \varepsilon$.

This formulation is equivalent to the weighted least squares (WLS) estimator used for addressing heteroskedasticity, except that in WLS, the weights are not random but typically correspond to the reciprocal of the skedastic function $v(x) = \mathbb{E}(\varepsilon^2|x)$ determining the conditional heteroskedasticity (see e.g., Romano and Wolf (2017)). Applied to the classical pairs bootstrap, weights follow a Multinomial law with $\mathbb{E}(\omega_i) = 1, \forall i$. However, by constructing random weights which satisfy $\mathbb{E}(\omega) = 1/\hat{v}(x)$, for instance via $\omega \sim D_\alpha$, with $\alpha \propto 1/\hat{v}(x)$, one gets a pairs bootstrap which is also an instance of WLS regression with incorporated uncertainty about the skedastic function. This addresses the limitation of the WLS theory, which assumes that the weights are known exactly, a rare occurrence. Unfortunately, with this added complexity, there is no longer a closed form for the

covariance of the regression coefficients. From this perspective, the second bootstrap variant, acting on the residuals, is more attractive.

In the residuals bootstrap, coefficients are initially estimated on the full sample using OLS: $\hat{\beta} = (X^\top X)^{-1} X^\top y$. The residuals are then resampled (with replacement) and incorporated into model (4.5), leading to altered response variables $y^* = X\hat{\beta} + \hat{\varepsilon}^*$. This, in turn, allows for the re-estimation of the coefficients: $\tilde{\beta}^* = (X^\top X)^{-1} X^\top y^* = (X^\top X)^{-1} X^\top (X\hat{\beta} + \hat{\varepsilon}^*)$. As before, we can reformulate the equation in terms of (scaled) random weights $W = \text{diag}(\omega)n$ as $\tilde{\beta} = (X^\top X)^{-1} X^\top (X\hat{\beta} + W\hat{\varepsilon}) = \hat{\beta} + (X^\top X)^{-1} X^\top W\hat{\varepsilon}$.

The second-step estimations are conditional on X and $\hat{\beta}$, and we can express them as

$$\mathbb{E}(\tilde{\beta}|X, \hat{\beta}) = \hat{\beta} + (X^\top X)^{-1} X^\top \mathbb{E}(W)\hat{\varepsilon}$$

and

$$\begin{aligned} \text{Cov}(\tilde{\beta}|X, \hat{\beta}) &= \mathbb{E} \left(\left(\hat{\beta} + (X^\top X)^{-1} X^\top - \hat{\beta} - (X^\top X)^{-1} X^\top \mathbb{E}(W)\hat{\varepsilon} \right)^2 \right) \\ &= \mathbb{E}(AW\Sigma WA) - b^2 \end{aligned}$$

where $A = (X^\top X)^{-1} X^\top$, $\Sigma = \hat{\varepsilon}\hat{\varepsilon}^\top$, $b = (X^\top X)^{-1} X^\top \mathbb{E}(W)\hat{\varepsilon}$ and $\mathbb{E}(W) = \text{diag}(\bar{\omega})$ with $\bar{\omega} = \mathbb{E}(\omega)$.

Let a_j denote the j -th row of matrix A , then

$$\tilde{\beta}_j = \hat{\beta}_j + \omega^\top \underbrace{(a_j^\top \odot \hat{\varepsilon})n}_{=:z}$$

where \odot denotes the Hadamard product (i.e., elementwise multiplication). By proposition 14 and acknowledging that $\hat{\beta}_j|z$ is a constant we have

$$SE(\tilde{\beta}_j) = \sqrt{m_2(z, \bar{\omega}) (\mathbb{E}(\|\omega\|_2^2) - \|\bar{\omega}\|_2^2)}.$$

Taking $\omega \sim M_{\mathbf{1}/n}$, this reduces to $SE(\tilde{\beta}) = (\sum_{i=1}^n ((z - \bar{z})^2)/n^2)^{1/2}$. In contrast, the variance of the classical residual bootstrap estimator is $SE(\hat{\beta}^*) = (\sigma_{\hat{\varepsilon}^*}^2 (X^\top X)^{-1})^{1/2}$ where $\sigma_{\hat{\varepsilon}^*}^2$ is the average variance of the bootstrap error terms. The formulation in terms of

random weights offers an advantage as it allows for easy modification of the bootstrap resampling plan by changing the distribution of the weights. Additionally, it is worth noting that if we choose $\omega \sim D_1$, then the complete distribution of $\tilde{\beta}_j$ becomes available.

In our approach, conditional heteroskedasticity can be addressed in two ways. The first solution involves scaling the OLS residuals by a factor dependent on the leverage value, i.e., the diagonal elements of the *hat* matrix $h_{ii} = [X(X^\top X)^{-1}X^\top]_{ii}$. This scaling is given by $\frac{\hat{\varepsilon}_i}{1-h_{ii}}$, as originally suggested in Wu (1986), underlying the wild bootstrap. This approach is used in the Heteroskedasticity Consistent Covariance Estimators (HCCE) introduced in MacKinnon and White (1985). With this formulation, it is reasonable to take weights from a flat Dirichlet distribution.

The second possibility is to choose a Dirichlet parametrization $\alpha \propto \frac{1}{1-h_{ii}}$ for the weights without adding any scaling to the OLS residuals. Unlike the first variant, selecting an asymmetric distribution for the random weights vector adjusts the uncertainty on each residual corresponding to the i -th leverage term. While the full conditional distribution of the coefficient is no longer available, exact central moments still are.

We note that all the results discussed also apply to regressions under Ridge regularisation, i.e., with L_2 penalty, since in this case: $\hat{\beta} = (X^\top X + \lambda I)^{-1}X^\top y \Rightarrow A = (X^\top X + \lambda I)^{-1}X^\top$ and $b = (X^\top X + \lambda I)^{-1}X^\top \mathbb{E}(W)\hat{\varepsilon}$.

4.5.2 Eigenvalue Estimation

The distribution of eigenvalues for a class of rotation equivariant (i.e., with constant eigenvectors) covariance matrices can be estimated in the following manner.

Let X denote an elliptically distributed random vector having mean zero and covariance $\Sigma = Cov(X)$ with eigenvector matrix Γ and diagonal eigenvalue matrix Λ allowing for the factorization $\Sigma = \Gamma\Lambda\Gamma^\top$. Let $S = X\Gamma$ be principal components such that $Cov(S) = \Lambda$. Taking s_j to be the j -th column of S one has $\mathbb{E}(s_j) = \lambda_j$, where λ_j is the j -th eigenvalue. We aim to learn about the distribution of the average of an estimate $\hat{\lambda}_j$ obtained from the sample covariance matrix $\hat{\Sigma} = Cov(\{X_t\}_{t=1}^T)$. Let us write $\tilde{\lambda}_j = \sum_{i=1}^n \omega_i z_i$, where ω is a random probability vector and z is the squared j -th sample principal component $z = \tilde{s}_j^2$ obtained from the rotated data sample $\{X_t\}_{t=1}^T$. Then, the standard error of the bootstrap distribution of the mean of the j -th eigenvalue can be derived analytically

using our previous results. Under $\omega \sim D_1$, the entire exact bootstrap distribution of $\tilde{\lambda}_j$ is computable using Varsi's algorithm.

4.5.3 Nonparametric Density Estimation

Let $y = (y_1, y_2, \dots, y_n)$ be an iid sample drawn from some univariate distribution with an unknown density f . A common way to estimate f from data is to use kernel smoothers. The kernel density estimator (KDE) accredited to [Rosenblatt \(1956\)](#) and [Parzen \(1962\)](#) is

$$\hat{f}_h(y) = \frac{1}{n} \sum_{i=1}^n K_h(y - y_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{y - y_i}{h}\right),$$

where K is the kernel and $h > 0$ is a smoothing parameter called the bandwidth.

Bootstrap applications to KDE are mainly twofold: i) bandwidth selection and ii) confidence interval estimation. In the first case, the bootstrap analog to the Parzen-Rosenblatt estimator, which is

$$\hat{f}_h^*(y) = \frac{1}{n} \sum_{i=1}^n K_h(y - y_i^*),$$

is used to pick h so as to minimize the mean integrated squared error (MISE)

$$\mathbb{E}^* \int \left(\hat{f}_h^*(y) - \hat{f}_{h'}(y) \right)^2 dy,$$

where h' is a pilot bandwidth and \mathbb{E}^* denotes the expected value w.r.t. the bootstrap sample.

In the second case, confidence intervals about the estimated \hat{f}_h at a given value x are drawn from the bootstrap distribution of $\hat{f}_h^*(x)$.

To leverage on our previous results, define a bootstrap estimator as a randomly weighted sum of kernel function values $z_i = K_h(y - y_i)$, i.e., as

$$\hat{f}_{h,\omega}(y) = \sum_{i=1}^n \omega_i z_i. \tag{4.6}$$

Under $\omega \sim D_1$ Varsi's algorithm returns the exact quantile values of the distribution

of $\hat{f}_{h,\omega}(x)$. Beside the elimination of randomization error and the improvements in computational speed, another advantage of this approach is that confidence bounds are smooth (just like the KDE).

Further, the ideas from Section 4.4 to include prior information draw a link to *conditional* kernel density estimation. Let $x = (x_1, \dots, x_n)$ be an iid data sample from a second random variable $X \in \mathbb{R}^p$ having continuous distribution and density $f(x)$. Consider the Nadaraya-Watson kernel smoother first proposed in Rosenblatt (1969):¹²

$$\hat{f}_h(y|x) = \sum_{i=1}^n v_i(x) K_{h_y}(y_i - y)$$

with weight function

$$v_i(x) = \frac{K_{h_x}(x_i - x)}{\sum_{i=1}^n K_{h_x}(x_i - x)}$$

and smoothing parameter $h = (h_y, h_x)$.

Let us define random weights such that $\mathbb{E}(\omega_i) = v_i(x)$. For instance, we may choose a Dirichlet distribution with parameter $\alpha = v(x)n$. Doing so, we get

$$\hat{f}_{h,\omega}(y|x) = \sum_{i=1}^n \omega_i z_i$$

which comes with standard error and third central moment for free as we can apply propositions 14 and 15.

Additionally, one could think of adding a data-driven (i.e., automatically tuned) concentration parameter $\lambda(x) > 0$ and use a Dirichlet model with $\alpha = v(x)n\lambda(x)$ creating a wider spread for the distribution of $\hat{f}_{h,\omega(x)}(y|x)$ in sparsely populated neighbourhoods of x , e.g., using $\lambda(x) = e^{-\sum_{i=1}^n v(x) \log(v(x))} \in [1/n, 1]$.

4.6 Conclusion and further discussion

In our examination of nonparametric bootstrap techniques, we approached the subject from a geometric perspective. We identified that the various methods, which include both classical and Bayesian frameworks, can be effectively represented by random weights. We

¹²See also Nadaraya (1964) and Watson (1964).

found that there is a one-to-one correspondence between the resampling plan and the model of the random weights.

The geometric properties of the distribution of the weights were then mapped to exact results on the bootstrap distributions. For example, we demonstrated that the Bayesian bootstrap distribution proposed by [Rubin \(1981\)](#) for linear statistics can be expressed in terms of the ratio of the volume of the intersection of a standard unit simplex with a half-space over the volume of the simplex. This ratio can be computed exactly using a closed-form expression provided in [Lasserre \(2015\)](#) or an exact iterative formula proposed by [Varsi \(1973\)](#). As a result, Rubin’s bootstrap distribution and inferential results can be obtained directly without resorting to sampling procedures, regardless of the distribution of the data x .

We further established a link between the exact central moments of the bootstrap distribution and the (expected) norm of the random weights which holds for all commonly used bootstrap procedures. Additionally, we discussed methods for incorporating prior information into bootstrap procedures by integrating the information into the parametrization of the distribution of the weights, inherently resulting in an adjusted resampling plan.

Finally, applications were discussed where it is common to use bootstrapped standard errors, which by our solution, are given without a need for resampling.

The following considerations present avenues for further exploration.

Dependent data. Discussed bootstrap methods assume that data are iid. Extensions to dependent data include the moving block bootstrap ([Künsch, 1989](#)), the non-overlapping block bootstrap ([Carlstein, 1986](#)), the circular block bootstrap ([Politis and Romano, 1992](#)) and the stationary bootstrap ([Politis and Romano, 1994](#)). See [Lahiri \(1999\)](#)¹³ for a comparison of the of the different methods.

Instead of sampling blocks, let us define $g(x_t) = \sum_{(t-b):t}(x_t - \bar{x})^2$, i.e., we apply a rolling average over the time series $\{x_t\}_{t=1}^T$ where b defines the block length. Then, run a bootstrap where at each step, sampling with replacement is done $n/b = m < n$ times.

¹³See also [Nordman \(2009\)](#) which correct an error in Lahiri’s calculation of the variance associated with the stationary bootstrap of [Politis and Romano \(1994\)](#).

The resulting m -out-of- n bootstrap distribution is equivalent to the overlapping block bootstrap.

Information theoretic considerations. The concept of typicality from information theory allows for a characterization of the sequence of sampled partitions discussed in Section 4.2.3. When running the bootstrap R times, the average log probability of the sequence of sampled partitions will be close to the entropy of the distribution of the partitions. Formally, $-\frac{1}{B} \log_2 p(\rho_1, \dots, \rho_B) \approx H(\mathcal{P}(n))$ where $H(\mathcal{P}(n)) = \sum_{l=1}^L p(\rho_l) \log_2 \frac{1}{p(\rho_l)}$. Note that the most likely sequence, the one where the most likely partition is sampled R times, does not belong to the typical set since $(\max p)^B \not\approx 2^{-BH(\mathcal{P}(n))}$. Because typical sequences assume a probability about $2^{-BH(\mathcal{P}(n))}$ and their total probability is almost 1, the size of the typical set has to be approximately $2^{BH(\mathcal{P}(n))}$.

When running the classical bootstrap, the sequence of partitions induced by the activated grid nodes likely belongs to the typical set of such sequences when R is chosen large enough. Hence, the information-theoretic perspective provides a means to choose R .

References

- Aggarwal, C. C., Hinneburg, A., and Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional space.
- Ali, M. M. (1973). Content of the frustum of a simplex. *Pacific Journal of Mathematics*, 49:313 – 322.
- Andrews, G. E. (1998). The theory of partitions (vol. 2 encyclopedia of mathematics and its applications).
- Angelova, J. A. (2012). On moments of sample mean and variance. *International Journal of Pure and Applied Mathematics*, 79:67 – 85.
- Bertsimas, D. and Sturt, B. (2020). Computation of exact bootstrap confidence intervals: Complexity and deterministic algorithms. *Operations Research*, 68:655 – 964.
- Bickel, P. J., Götze, F., and van Zwet, W. R. (1997). Resampling fewer than n observations. *Statistica Sinica*, 7:1 – 31.
- Boor, C. D. (1972). On calculating with b-splines. *Journal of Approximation Theory*, 6:50 – 62.
- Boos, D. D. and Monahan, J. F. (1986). Bootstrap methods using prior information. *Biometrika*, 73:77 – 83.
- Cales, L., Chalkis, A., and Emiris, I. (2021). On the cross-sectional distribution of portfolios’ returns. *arXiv:2105.06573*.
- Cales, L., Chalkis, A., Emiris, I., and Fisikopoulos, V. (2018). Practical volume computation of structured convex bodies, and an application to modeling portfolio dependencies and financial crises.
- Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics*, 14:1171 – 1179.
- Cho, Y. and Cho, E. (2001). The volume of simplices clipped by a half space. *Applied Mathematics Letters*, 14:731 – 735.

- Cox, M. (1972). The numerical evaluation of b-splines. *IMA Journal of Applied Mathematics*, 10:134 – 149.
- Curry, H. B. and Schoenberg, I. J. (1966). On pólya frequency functions iv: the fundamental spline functions and their limits. *Journal d'Analyse Mathématique*, 7:7 – 107.
- Dirksen, H. (2017). Sections of the regular simplex volume formulas and estimates. *Mathematische Nachrichten*, 290:2567 – 2584.
- Efron, B. (1979). The bootstrap: Another look at the jackknife. *The Annals of Statistics*, 7:1 – 26.
- Freedman, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics*, 9:1218 – 1228.
- Frigyik, B., Gupta, M., and Chen, Y. (2010). Shadow dirichlet for restricted probability modeling. *Advances in Neural Information Processing Systems 23*, pages 613 – 621.
- John, V., Angelov, I., Oncül, A. A., and Thévenin, D. (2007). Techniques for the reconstruction of a distribution from a finite number of its moments. *Chemical Engineering Science*, 62:2890 – 2904.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17:1217 – 1241.
- Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27:386 – 404.
- Lasserre, J. B. (2015). Volume of slices and sections of the simplex in closed form. *Optimization Letters*, 9:1263 – 1269.
- Ledoux, M. (2001). The concentration of measure phenomenon.
- Lo, A. Y. (1987). A large sample study of the bayesian bootstrap. *The Annals of Statistics*, 15:360 – 375.
- MacKinnon, J. G. and White, H. L. (1985). Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 21:53 – 70.

- Mason, D. M. and Newton, M. A. (1992). A rank statistics approach to the consistency of a general bootstrap. *The Annals of Statistics*, 20:1611 – 1624.
- Milman, V. D. and Schechtman, G. (1986). Asymptotic theory of finite dimensional normed spaces. lecture notes in mathematics.
- Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, 9:141 – 142.
- Newton, M. A. and Raftery, E. (1994). Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56:3 – 26.
- Nordman, D. J. (2009). A note on the stationary bootstraps variance. *The Annals of Statistics*, 37:359 – 370.
- Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33:1065 – 1076.
- Politis, D. N. and Romano, J. P. (1992). A circular block resampling procedure for stationary data. In Lepage, R. and Billard, L., editors, *Exploring the Limits of Bootstrap*. Wiley, New York.
- Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89:1303 – 1313.
- Provost, S. B. (2005). Moment-based density approximants. *The Mathematica Journal*, 9:727 – 756.
- Quenouille, M. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society. Series B*, 11:68 – 84.
- Rimoldini, L. (2014). Weighted skewness and kurtosis unbiased by sample size and gaussian uncertainties. *Astronomy and Computing*, 5:1 – 8.
- Romano, J. P. and Wolf, M. (2017). Resurrecting weighted least squares. *Journal of Econometrics*, 197:1 – 19.

- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27:832 – 835.
- Rosenblatt, M. (1969). Conditional probability density and regression estimators. pages 25 – 31.
- Rubin, R. (1981). The bayesian bootstrap. *The Annals of Statistics*, 9:130 – 134.
- Singh, K. (1981). On the asymptotic accuracy of efron's bootstrap. *The Annals of Statistics*, 9:1186 – 1195.
- Tukey, J. (1958). Bias and confidence in not-quite large samples. *The Annals of Mathematical Statistics*, 29:614.
- Varsi, G. (1973). The multidimensional content of the frustum of the simplex. *Pacific Journal of Mathematics*, 46:303 – 314.
- Watson, G. S. (1964). Smooth regression analysis. *The Indian Journal of Statistics*, 26:359 – 372.
- Weng, C.-S. (1989). On a second-order asymptotic property of the bayesian bootstrap mean. *The Annals of Statistics*, 17:705 – 710.
- Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, 14:1261 – 1295.
- Zheng, Z.-G. (1987). Random weighting method. *Acta Mathematicae Applicatae Sinica*, 2:247 – 253.

Appendix

4A Algorithm of Varsi (1973)

ALGORITHM 10: Varsi's algorithm

Compute $u_i = z_i - \gamma$, $i = 1, \dots, n$.

Label non-negative u_i as u_1^+, \dots, u_F^+ and negatives as u_1^-, \dots, u_H^- .

Initialize $A_0 = 1, A_1 = A_2 = \dots = A_F = 0$.

For $h = 1, \dots, H$ repeat $A_f = \frac{u_f^+ A_f - u_h^- A_{f-1}}{u_f^+ - u_h^-}$, $f = 1, \dots, F$.

Then, for $h = H$, $A_f = \frac{V(S^{n-1} \cap H(z, \gamma))}{V(S^{n-1})}$.

4B Algorithm of Cales et al. (2021)

ALGORITHM 11: Exact central moments under $\omega \sim D_1$

Input : set of all partitions of integer r : $\mathcal{P}(r)$

Output : r -th central moment $\mu_r(\hat{\theta})$

Set $S = 0$;

foreach $\rho \in \mathcal{P}(r)$ **do**

decompose ρ in its d unique non-zero elements $(l_j)_{j=1}^d$ with multiplicities

$$\begin{aligned} & (p_j)_{j=1}^d; \\ a &= \prod_{j=1}^d \left(\sum_{i=1}^n (z_i - \bar{z})^{l_j} \right)^{p_j}; \\ b &= \prod_{j=1}^d p_j! l_j^{p_j}; \\ S &= S + a/b; \end{aligned}$$

end

return $\mu_r(\hat{\theta}) = S / \binom{n-1+r}{r}$;

4C Proof of proposition 14 under $\omega \sim D_\alpha$

$$\begin{aligned}
LHS &= \mathbb{E} \left((\hat{\theta} - \mathbb{E}(\hat{\theta}))^2 \right) \\
&= \mathbb{E} \left(\left(\sum_{i=1}^n \omega_i (z_i - \bar{z}) \right)^2 \right) \\
&= \sum_{i=1}^n \mathbb{E}(\omega_i^2) (z_i - \bar{z})^2 + 2 \sum_{i=1}^{n-1} \sum_{j>i}^n \mathbb{E}(\omega_i \omega_j) (z_i - \bar{z})(z_j - \bar{z}) \\
&= \sum_{i=1}^n \frac{\alpha_i(\alpha_i + 1)}{\alpha_0(\alpha_0 + 1)} (z_i - \bar{z})^2 + 2 \sum_{i=1}^{n-1} \sum_{j>i}^n \frac{\alpha_i \alpha_j}{\alpha_0(\alpha_0 + 1)} (z_i - \bar{z})(z_j - \bar{z}) \\
&= \frac{1}{\alpha_0(\alpha_0 + 1)} \left(\sum_{i=1}^n \alpha_i (z_i - \bar{z})^2 + \sum_{i=1}^n \alpha_i^2 (z_i - \bar{z})^2 + 2 \sum_{i=1}^{n-1} \sum_{j>i}^n \alpha_i \alpha_j (z_i - \bar{z})(z_j - \bar{z}) \right) \\
&= \frac{1}{\alpha_0(\alpha_0 + 1)} \left(\sum_{i=1}^n \alpha_i (z_i - \bar{z})^2 + \underbrace{\left(\sum_{i=1}^n \alpha_i (z_i - \bar{z}) \right)^2}_{=0 = \sum_{i=1}^n \alpha_i (z_i - \frac{1}{\alpha_0} \sum_{i=1}^n \alpha_i z_i) = \sum_{i=1}^n \alpha_i z_i - \frac{\sum_{i=1}^n \alpha_i}{\alpha_0} \sum_{i=1}^n \alpha_i z_i} \right) \\
&= \frac{1}{\alpha_0(\alpha_0 + 1)} \sum_{i=1}^n \alpha_i (z_i - \bar{z})^2. \\
RHS &= \sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z})^2 \frac{1}{1 - \sum_{i=1}^n \bar{\omega}_i^2} (\mathbb{E}(\|\omega\|_2^2) - \|\bar{\omega}\|_2^2) \\
&= \sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z})^2 \frac{1}{1 - \sum_{i=1}^n \bar{\omega}_i^2} \left(\sum_{i=1}^n \frac{\alpha_i(\alpha_i + 1)}{\alpha_0(\alpha_0 + 1)} - \sum_{i=1}^n \frac{\alpha_i^2}{\alpha_0^2} \right) \\
&= \sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z})^2 \frac{1}{1 - \sum_{i=1}^n \bar{\omega}_i^2} \left(\sum_{i=1}^n \frac{\alpha_0(\alpha_i^2 + \alpha_i) - (\alpha_0 + 1)\alpha_i^2}{\alpha_0^2(\alpha_0 + 1)} \right) \\
&= \sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z})^2 \frac{1}{1 - \sum_{i=1}^n \bar{\omega}_i^2} \left(\frac{1}{\alpha_0^2(\alpha_0 + 1)} \sum_{i=1}^n (\alpha_0 \alpha_i - \alpha_i) \right) \\
&= \sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z})^2 \frac{1}{1 - \sum_{i=1}^n \bar{\omega}_i^2} \left(\frac{\alpha_0^2 - \sum_{i=1}^n \alpha_i^2}{\alpha_0^2(\alpha_0 + 1)} \right) \\
&= \sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z})^2 \frac{1}{1 - \sum_{i=1}^n \bar{\omega}_i^2} \left(\frac{1 - \sum_{i=1}^n (\frac{\alpha_i}{\alpha_0})^2}{\alpha_0 + 1} \right) \\
&= \sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z})^2 \frac{1}{1 - \sum_{i=1}^n \bar{\omega}_i^2} \left(\frac{1 - \sum_{i=1}^n \bar{\omega}_i^2}{\alpha_0 + 1} \right) \\
&= \sum_{i=1}^n \left(\frac{\alpha_i}{\alpha_0} \right) (z_i - \bar{z})^2 \frac{1}{\alpha_0 + 1} \\
&= \frac{1}{\alpha_0(\alpha_0 + 1)} \sum_{i=1}^n \alpha_i (z_i - \bar{z})^2.
\end{aligned}$$

4D Closed-form expressions

Closed-form expressions of the first four scaled central moments (i.e., mean, variance, skewness and kurtosis) of the exact classical and Bayesian bootstrap distribution.

Let us define sample skewness and sample kurtosis as $sk(z) = \frac{\frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^3}{\left[\frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2\right]^{3/2}}$ and $k(z) = \frac{\frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^4}{\left[\frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2\right]^{4/2}}$. Then:

$$\begin{aligned} M_2^B &= \frac{1}{n^2} \sum_{i=1}^n (z_i - \bar{z})^2 \\ M_3^B &= \frac{1}{M_2^{3/2}} \frac{1}{n^3} \sum_{i=1}^n (z_i - \bar{z})^3 \\ &= sk(z) \frac{1}{n^{1/2}} \\ M_4^B &= \frac{1}{(M_2^B)^2} \frac{1}{n^4} \sum_{i=1}^n (z_i - \bar{z})^4 + 3 - \frac{3}{n} \\ &= k(z) \frac{1}{n} + 3 - \frac{3}{n} \end{aligned}$$

and

$$\begin{aligned} M_2^{BB} &= \frac{1}{n(n+1)} \sum_{i=1}^n (z_i - \bar{z})^2 \\ M_3^{BB} &= \frac{1}{M_2^{3/2}} \frac{2}{n(n+1)(n+2)} \sum_{i=1}^n (z_i - \bar{z})^3 \\ M_4^{BB} &= \frac{1}{M_2^2} \frac{1}{n(n+1)(n+2)(n+3)} \left(6 \sum_{i=1}^n (z_i - \bar{z})^4 + 3 \left(\sum_{i=1}^n (z_i - \bar{z})^2 \right)^2 \right). \end{aligned}$$

Under the Multinomial model with non-uniform probabilities we have:

$$\begin{aligned}
M_2^B(\bar{\omega}) &= \frac{1}{n} \sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z}_{\bar{\omega}})^2 \\
&= \frac{\tilde{s}_{\bar{\omega}}}{n} \\
M_3^B(\bar{\omega}) &= \frac{1}{(M_2^{\bar{\omega}})^2} \frac{1}{n^2} \sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z}_{\bar{\omega}})^4 \\
&= sk(z, \bar{\omega}) \frac{n^{1/2}}{n} \\
M_4^B(\bar{\omega}) &= \frac{1}{(M_2^{\bar{\omega}})^2} \frac{1}{n^3} \sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z}_{\bar{\omega}})^4 + 3 - \frac{3}{n} \\
&= k(z, \bar{\omega}) \frac{1}{n} + 3 - \frac{3}{n}.
\end{aligned}$$

Under the flat Dirichlet Model one has:

$$\begin{aligned}
M_2^{BB}(\bar{\omega}) &= \frac{1}{n+1} \sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z}_{\bar{\omega}})^2 \\
&= \frac{\tilde{s}_{\bar{\omega}}}{n+1} \\
M_3^{BB}(\bar{\omega}) &= \frac{1}{(M_2^{\bar{\omega}})^{3/2}} \frac{2}{(n+1)(n+2)} \sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z}_{\bar{\omega}})^3 \\
&= sk(z, \bar{\omega}) \frac{2(n+1)^{1/2}}{n+2} \\
M_4^{BB}(\bar{\omega}) &= \frac{1}{(M_2^{\bar{\omega}})^2} \frac{1}{(n+1)(n+2)(n+3)} \left(6 \sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z}_{\bar{\omega}})^4 + 3n \left(\sum_{i=1}^n \bar{\omega}_i (z_i - \bar{z}_{\bar{\omega}})^2 \right)^2 \right) \\
&= \frac{3(n+1)}{(n+2)(n+3)} (2k(z, \bar{\omega}) + n).
\end{aligned}$$

Chapter 5

Good and Bad Turbulence

Cyril Bachelard

A version of the article has been submitted to *Journal of Applied Econometrics*.

Abstract

Kritzman and Li (2010) quantified financial turbulence by the squared Mahalanobis distance applied to the cross-section of asset returns and advocated its adoption as an effective indicator of market turmoil, aiding decision making in portfolio selection and risk management. In this study, we empirically and theoretically assess the predictive power of their indicator. We establish a set of jointly sufficient conditions on the data generating process of asset returns to validate its applicability for dynamic hedging. Additionally, we explore potential pitfalls in practical applications and propose an enhanced indicator based on the discrepancy between two smoothed distance measures. This improved indicator effectively distinguishes good from bad turbulence in the sense of turbulent but favorable versus turbulent and unfavorable market developments. Our approach extends discriminant analysis to a nonparametric framework specifically designed for financial time series that exhibit non-normal distribution and temporal dependence.

JEL classification: C11, C13, C14, C32, C38, C58.

Keywords: Mahalanobis Distance, Dynamic Hedging, Discriminant Analysis, Stochastic Volatility, GARCH.

5.1 Introduction

Turbulence qualitatively describes a state of confusion and disorganized change. In a financial context, the term is used to describe episodes of turmoil characterized by large market swings. Like the physical phenomenon, the economic interpretation comes with a negative connotation and intuition dictates that investors should steer clear of turbulent events. Typical ingredients used to cook up indicators of financial turbulence include (implied) volatility, correlations, spreads and liquidity as well as macroeconomic factors. [Chow et al. \(1999\)](#) and [Kritzman and Li \(2010\)](#) related financial turbulence to the mathematical concept of distance and its use to detect unusual patterns in stock market price fluctuations. The authors define instances as unusual when a vector of asset returns, $x_t \in \mathbb{R}^n$, representing the cross-section of price changes for a set of n assets from time $t - 1$ to t , deviates significantly from the center of the multivariate asset return distribution. In other words, the irregularity of a data point x_t , relative to other data points, is measured by its distance from the distribution's center.

[Kritzman and Li \(2010\)](#) state that the dates where one observes such distant or outlying observations tend to cluster over time and that those episodes which they call *turbulent* are characterized by substantially lower risk adjusted returns than those observed during non-turbulent, i.e., *quiet* periods. Their definition aligns with the negative connotation of turbulence, signifying undesirable and persistent circumstances. The authors further highlight the potential for mitigating losses by acting on early warning signals derived from a distance-based monitoring of market conditions. Although it may not be possible to anticipate the initial shock, quickly adjusting investment positions through hedging or reallocating to safe havens after a spike in turbulence is detected can be advisable to minimize exposure to subsequent aftershocks, which tend to occur because of the persistence property of the turbulence measure.

The methodology presented by [Chow et al. \(1999\)](#) and [Kritzman and Li \(2010\)](#) utilizes the Mahalanobis distance as the underlying metric for measuring turbulence. This distance is commonly used in multivariate outlier detection and is closely related to Hotelling's T-square distribution, which is used for multivariate statistical testing. From this perspective, monitoring market turbulence boils down to multivariate outlier detection. But why should multivariate outliers in return series cluster over time and more often than not reside in

the negative orthant? And are these two features sufficient for a turbulence-based risk management approach to provide the desired protection?

In the following section we will demonstrate that the empirical observations regarding the persistence and negative correlation of turbulence with market returns find support in the academic literature. These findings can be partly explained by two well-established stylized facts¹ of asset returns (Cont, 2001): volatility clustering and the leverage effect. However, the two features alone are not enough to create a meaningful turbulence-based investment strategy. To make such a strategy viable, a further condition on the data generating process of asset returns is required. A simple simulation will make this point clear.

While being straightforward and accessible, the practical effectiveness of a Mahalanobis-based monitoring of market developments is limited. This limitation arises from two primary factors: i) trading costs and ii) missed profits resulting from false positive signals. While the issue of trading costs can be mitigated through signal filtering techniques, the missing out problem is more inherent. Although turbulence values spike during market crashes, they also exhibit such spikes during strong market rallies. This symmetry poses a problem for a turbulence-based risk management approach as it is prone to overlooking subsequent rebounds that typically follow significant negative market corrections. Our primary contribution is the introduction of an alternative Mahalanobis-based measure that effectively distinguishes between good and bad turbulence, corresponding to unusual positive and negative returns, respectively.

We promote a change of perspective in the use of the Mahalanobis distance for turbulence identification. While the approach based on the work of Chow et al. (1999) and Kritzman and Li (2010) relates to the statistical application of multivariate outlier detection, our proposal aligns with another common usage of the Mahalanobis distance in cluster analysis and supervised classification. Our idea involves adopting a binary mixture model to describe the evolution of asset returns and classifying a test point based on the distribution with the smallest Mahalanobis distance. Two Mahalanobis distances are computed using two sets of parameters, i.e., centers and shape matrices describing the asset return distribution during either upward or downward trending regimes (bear and

¹Stylized facts are statistical properties that appear to be present in empirical asset returns across time and markets.

bull markets). The difference between these two distances provides an improved indicator allowing for a distinction of good and bad turbulence with considerable predictive power.

Our proposed measure exhibits a close relationship with the objective function employed in quadratic discriminant analysis (QDA), a supervised classification method. However, unlike in QDA, our objective is not to learn posterior class probabilities but rather to develop a statistical indicator for portfolio and risk management purposes. As a result, our work offers a finance-specific perspective on supervised learning, encompassing QDA as a Bayesian classifier. This perspective contributes to two additional aspects: adapting discriminant analysis to financial time series that do not adhere to normal distribution or independence over time, and addressing the labeling problem.

When working with financial time series, supervised learning encounters a challenge uncommon in typical classification exercises involving distinct categories like cats and dogs: labeling the training data. Determining a sample of well-classified elements is not straightforward when working with financial data. This issue has received limited attention in the literature. In our case, the labels need to identify the market as either upward or downward trending. While multiple approaches can achieve this, each produces a distinct set of labels, introducing ambiguity in the learning target. We propose employing a robust pattern recognition technique to objectify this choice.

In a final step, we present a method that results in a smoothly time-varying discriminant, effectively balancing statistical significance and economic significance in terms of predictive power and financial rewards after accounting for transaction costs.

The remainder of this article is structured as follows: Section 5.2 presents the definition of the turbulence measure together with methods to discriminate turbulent from quiet events and discusses the indicators properties. Section 5.3 explores the connection between turbulence and discriminant analysis and introduces a novel indicator to distinguish between good and bad turbulence based on a pattern recognition technique to distinguish between bear and bull phases. A data-driven approach to smooth the turbulence measure based on an orthogonalized stochastic volatility model is proposed to reduce trading and related costs. Section 5.4 conducts a statistical and economic analysis of hedging rules derived from various turbulence measures using both real and synthetic data. Section 5.5 concludes.

5.2 Financial turbulence: Measures, properties and hedging rules

Let X be a real-valued random vector having probability distribution $F_X \in \mathcal{F}$, \mathcal{F} denoting the family of elliptically symmetric distributions defined on \mathbb{R}^n . In our context, X represents the price variation of a collection of assets. Throughout this work, we are going to use a set of $n = 24$ time series of daily discrete country index returns of the index provider MSCI² to exemplify and illustrate theoretical concepts. Details about the data are given in section 5.4.1. Whenever we refer to computations on empirical returns we use the notation $\{x_t\}_{t=1}^T$ to emphasize the time series nature of the data.

The Mahalanobis distance is defined as

$$D = \sqrt{(X - \mu)^\top \Sigma^{-1} (X - \mu)} \quad (5.1)$$

with $\Sigma = \text{Cov}(X)$ a positive definite covariance matrix and $\mu = \mathbb{E}(X)$ the center of the distribution of X .

The Mahalanobis metric measures the distance of a realization x of a random vector X to its expected value³. The scaling by the covariance matrix Σ differentiates the Mahalanobis from the Euclidean distance (note that the two are equivalent for $\Sigma = I$). Because of the scaling, points with equal distance to the common center do not reside on a sphere but on the shell of an ellipsoid. The Mahalanobis distance thus accounts for the variances of and correlations among the marginals of X . Geometrically we can think of the Euclidean ball being stretched and compressed by the eigenvalues of the covariance matrix and in direction of the matrix's eigenvectors to form an ellipsoid.

For the family \mathcal{F} the ellipsoid forms an iso-density curve allowing for an evaluation of how likely it is that a point lies at a particular distance (or closer) from the center of the distribution⁴. This is why the Mahalanobis distance is ubiquitously used for multivariate outlier detection and statistical testing. Note that a point can be a multivariate outlier even if it is not a univariate outlier for any variable. The Mahalanobis distance essentially

²See <https://www.msci.com>

³Note that nothing prevents us from replacing μ with any other vector $y \in X$ so that the Mahalanobis distance in this case would not measure distance to the center but among two realizations.

⁴For the special case of the multivariate Normal distribution the probability density of an observation is uniquely determined by the Mahalanobis distance.

serves as a multivariate analog to a z-score, i.e., of the idea of measuring how many standard deviations away a realization of a random variable is from its mean. In practical applications where the population distribution is often unknown, point estimates like the sample moments or robust statistics like the median for the centroid or shrinkage type estimators for the covariance are plugged into the distance function (Hardin and Rocke, 2005).

In Kritzman and Li (2010), the authors compute turbulence as the squared Mahalanobis metric for all realization of a set of n financial time series $\{x_t\}_{t=1}^T$ based on the set's sample mean $\hat{\mu}$ and covariance $\hat{\Sigma}$. Their turbulence indicator reads

$$d_t = (x_t - \hat{\mu})^\top \hat{\Sigma}^{-1} (x_t - \hat{\mu}). \quad (5.2)$$

Values of d_t above a certain threshold γ are considered as turbulent events and values below as quiet. The obvious question is how to choose γ . If asset returns would be independent and identically distributed (iid) Gaussian random variables with known population parameters, then D would follow a $\chi^2(n)$ distribution and the sample counterpart, $d_t^{1/2}$, a scaled Beta distribution (Hardin and Rocke, 2005), and one could set γ such that the probability of exceeding the critical value is fixed, say for instance, at 5%. But, if that assumption was indeed true, then there would be no point to do the exercise because, by the symmetry of the Normal distribution, distant observations would not be associable to negative market developments. When the distribution of the data at hand departs from a Normal (or any elliptical for that matter), the quadratic form in equation 5.2 still yields an ellipsoid when set to a constant. However, the interpretation of its shell as a probability density contour (and its volume as representing probabilities following a χ^2 distribution under the Gaussian assumption) no longer holds and it is not clear how to select and interpret γ .

Figure 5.1 compares the empirical distribution of observed turbulence values computed on the MSCI World country index series with a theoretical counterpart under a Normal assumption. From the chart it is evident that the χ^2 approximation does not fit the data. Most empirical distances lie below theoretical values while some range way above creating a heavy right tail (which might best be described by a power law). In the following section we investigate the question as to why this is. As we will see, it is not primarily the

departure from a Gaussian or elliptical shape that causes this phenomenon.

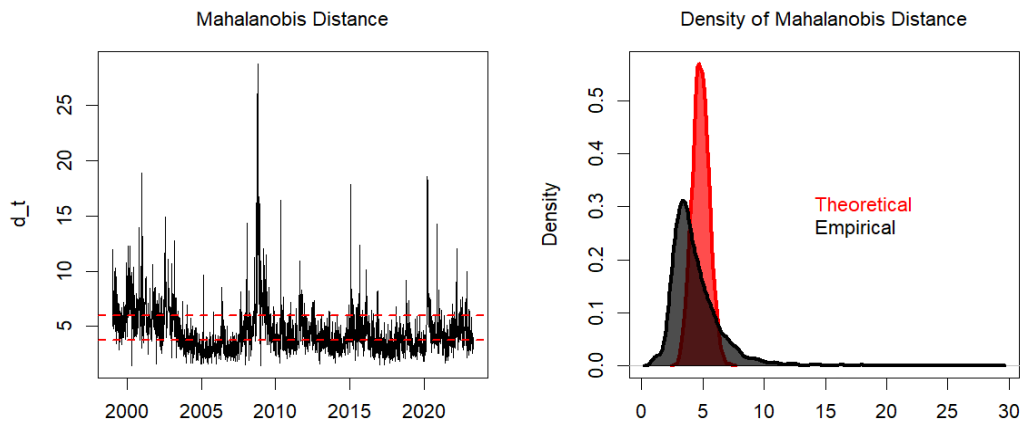


Figure 5.1: The left panel shows the time evolution of a turbulence indicator, defined by the Mahalanobis distance, computed on 24 MSCI country indices. All substantial corrections in the MSCI World index are characterized by a spike in the turbulence index. To compare, the red dashed line indicate the 95% confidence interval based on a Gaussian model fitted to the empirical data. The right panel compares the corresponding distributions of the two indicators. The departure of the empirical distribution from the theoretical χ^2 model is evident.

[Kritzman and Li \(2010\)](#) make no statement about the distributional family asset returns belong to. They provide illustrative examples using a Gaussian assumption describing the ellipsoid as the boundary discriminating turbulent from quiet dates but ultimately select γ according to an empirical quantile. I.e., turbulent events are those with the highest distance values within the dataset, without considering a model for X . This makes sense for descriptive in-sample analysis but less so for out-of-sample backtests as the method will reclassify previously turbulent dates as non-turbulent with the advent of new data.

An alternative approach is to compare empirical distance values with cutoff values derived from a theoretical model. In this approach, dates are identified as turbulent when empirical observations deviate from the predictions of the model (implying that turbulence coincides with a breakdown of the validity of the equations used to describe the system during normal times). By adopting this method, turbulent events are not necessarily considered unusual in terms of the empirical distribution of asset returns, but rather in relation to an idealized elliptical model. In other words, turbulence is deemed *excessive* when it exceeds the expectations of the model. Consequently, turbulent observations may not be empirically rare.

In section 5.4 we investigate the benefit of hedging a long exposure in the MSCI World

index during periods of elevated turbulence and compare it to other Mahalanobis based hedging rules which will be presented in the remainder.

The first generic hedging rule is:

$$R_1 : \quad h(x) = \begin{cases} 1 & \text{if } x \in \mathcal{T} := \{d(x) \in \mathbb{R} \mid d(x) \geq P^{-1}(p)\} \\ 0 & \text{else} \end{cases}$$

According to rule R_1 , a hedge is constructed whenever the turbulence indicator $d(x)$ exceeds the p -quantile of a distribution function P . This can be the indicators' empirical distribution, i.e., $P = \hat{F}_n$, as suggested in [Kritzman and Li \(2010\)](#), or the χ^2 distribution implied by the Gaussian hypothesis, i.e., $P = \chi_n^2$.

5.2.1 Properties

The aforementioned setup allows us to be unspecific about the distribution of asset returns. Hedging rule R_1 heuristically acts to the presence of excess turbulence, either vis-à-vis a theoretical model or relative to past observations. However, avoiding turbulent episodes only makes sense if i) turbulent events tend to coincide with negative market movements more frequently than with positive ones and if ii) turbulence exhibits persistence over a certain period, making it probable that once a spike occurs, more spikes will follow which in turn are more often than not connected to negative market developments.

Empirical evidence supporting these two conditions was presented by [Kritzman and Li \(2010\)](#). However, their analysis was based on an in-sample framework utilizing empirical percentiles, and they did not provide an explanation as to why this behavior occurred in the past or whether it should be expected to continue in the future. We therefore ask the question: what are the conditions regarding the data generating process of asset returns under which one can expect the observed characteristics of the turbulence measure d_t to generally manifest?

The first thing to note is that under an iid symmetrically elliptical model, neither feature can occur. Because of the symmetry of the distribution, distant points are no more likely to reside in the negative orthant than anywhere else and, because of independence, persistence is prevented.

Second, recall that by the scaling with the precision matrix (i.e., the inverse covariance matrix) in the Mahalanabis metric, observations can be distant for two reasons. It might be that elements in x_t take extreme values or that correlations among the elements of x_t differ from historical patterns. I.e., observations can deviate from their usual behavior in terms of variance or co-variance.

To see this more clearly, let $z_t = x_t - \mu$ and reformulate equation 5.2 as

$$d_t = \sum_{i=1}^n z_{t,i}^2 \tilde{\sigma}_{i,i} + 2 \sum_{i=1}^n \sum_{j>i}^n z_{t,i} z_{t,j} \tilde{\sigma}_{i,j} \quad ; \tilde{\sigma}_{i,j} = [\hat{\Sigma}^{-1}]_{i,j}. \quad (5.3)$$

Both, variances and co-variances of asset returns are known to be time-varying and persistent. This property is considered a stylized fact of asset returns and usually subsumed under the term volatility clustering. A wealth of econometric models have been designed to account for these effects, either deterministically or probabilistically, building on the seminal work of Engle (1982) and Bollerssev (1986) on GARCH processes in the former case, and Taylor (1982) on stochastic volatility (SV) models in the latter case. In our setup, this means that both terms, $z_{t,i}^2$ and $z_{t,i} z_{t,j}$ display autocorrelation. Therefore, the dynamics of the turbulence indicator d_t is driven by the dynamics of second (co-) moments of asset returns and thus, the empirical persistence of turbulence may be explained by the dependence over time in second (co-) moments⁵.

Within the vast literature on volatility dynamics, a spezialized strain has focused on explaining and modeling another stylized fact stating that there exists a negative correlation between returns and volatility. High volatility is somehow connected to negative returns. However, the causality underlying this phenomenon remains controversial. Black (1976) first documented the inverse relationship and attributed it to the effects of financial leverage. Further empirical support followed by Christie (1982), Cheung and Ng (1992) and Duffee (1995) which strengthened the leverage hypothesis supporting the perspective that return shocks lead to changes in conditional volatility. On the other hand, time-varying risk premium theory

The GJR-GARCH model of Glosten et al. (1993) may be the most prominent adaptation of the standard GARCH explicitly modeling an asymmetric response of volatility to past

⁵Note that although we assume a time-varying structure for the second (co-) moments of asset returns, the standardization used to compute turbulence is done with a static covariance matrix.

positive or negative returns. Other implementations include [Omori et al. \(2007\)](#) or [Harvey and Shepard \(1996\)](#). Irrespective of the causal nature of the asymmetry, the negative correlation between volatility and returns affects the correlation between turbulence and returns since, according to equation (5.3), turbulence is largely volatility driven. The effect may explain why we observe more negative returns than positive ones during high turbulence phases.

The presence of both stylized facts, volatility clustering and the asymmetric response of volatility to returns (or vice versa), provides a theoretical basis for utilizing the Mahalanobis-based turbulence indicator in risk management. In order to assess whether these facts alone are sufficient, a simulation study is conducted in Section 5.4.4. In this study, turbulence-based hedging is backtested on synthetic data generated through multivariate asymmetric GARCH processes, ensuring the incorporation of both stylized properties.

Figure 5.2 illustrates the density of the turbulence indicator computed from a sample of such simulated data. It is observed that the distribution falls somewhere between the theoretical and empirical models. In the synthetic environment, turbulence clearly surpasses theoretical thresholds and exhibits clustering over time. However, it does not reach the extremes observed empirically, and the clustering over time is not as pronounced.

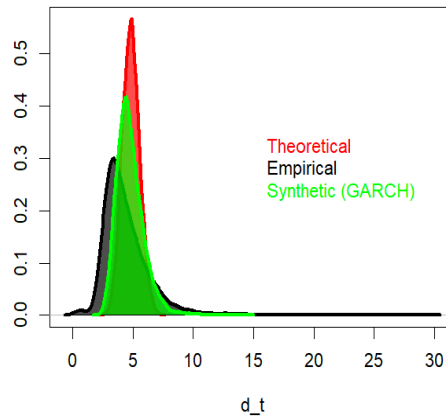


Figure 5.2: Comparison of the probability distributions of the Mahalanobis-based turbulence indicators under different assumptions regarding the data generating process.

The outcomes of the simulation study reveal that turbulence-based hedging enhances efficiency, although it comes with reduced returns when accounting for transaction costs. This diverges from results on real data, which show no performance drag, hinting at the presence of an overlooked aspect.

Our conjecture is that the absence of simultaneous shocks in the simulated data

might be a crucial missing element. In real data, major shocks tend to propagate globally, impacting all series roughly simultaneously and leading to a synchronization of the volatility dynamics.

Synthetically generating multivariate datasets with this characteristic is challenging. An econometric model proposed by [Paolella and Polak \(2015\)](#) captures stochastic co-jumps in volatility series and across assets. More recent developments explore the use of generative adversarial networks (GANs) as a deep learning model for generating artificial financial time series, but their multidimensional applications remain elusive.

As a simple and thus easily replicable alternative to the more involved GARCH or GAN-based simulations, we suggest to employ a nonparametric technique to resample real data in [Section 5.4.4](#). This approach allows for the modification of historical asset price paths while preserving the mentioned stylized properties. The simulation study concludes that synthetic turbulence, exhibiting volatility clustering, asymmetric response, and volatility synchronization, behaves similarly to observed turbulence in real data and can be exploited to time risk management decisions.

5.3 Discriminating good from bad turbulence

As previously discussed, the term turbulence is associated with negative market developments and we have given some theoretical justification for its use as a risk signal. However, empirical observations show that among the largest spikes in turbulence happen when markets close far in the positive. This is not surprising, given the symmetry of definition [\(5.2\)](#), by which unusually positive market developments are just as turbulent as unusually negative ones. Although empirical observations confirm that avoiding turbulent phases boosts risk-adjusted returns (winning by losing less), a price has to be paid in terms of missed rallies which also cause excessive levels of turbulence. We now introduce a simple technique to discriminate up- from downturn induced turbulence.

Let $\{x_t\}_{t=1}^T$ represent a multivariate time series of asset returns, as before. We posit that the observation x_t is part of one of two populations: class \mathcal{C}_1 , characterized by a mean vector μ_1 and dispersion matrix Σ_1 , or class \mathcal{C}_2 with parameters μ_2 and Σ_2 . Our training set comprises data up to a specific point $\tau < T$, which includes a bivariate series of state variables or class labels $\{y_t\}_{t=1}^\tau$. Each $y_t \in \{1, 2\}$ denotes the class membership of a data

point. Consequently, our training set can be expressed as $\{(x_t, y_t)\}_{t=1}^T$.

The state variable partitions the historical data into either upward- or downward-trending market regimes. Consequently, classes \mathcal{C}_1 and \mathcal{C}_2 represent falling and rising market environments (we interchangeably refer to classes as regimes or states). When $y_t = 1$, it signifies a downward-trending market at time t , while $y_t = 2$ indicates an upward-trending phase. Notice that $y_t = 2$ doesn't imply a positive return on day t but rather that it belongs to an upward-trending phase.

Details on obtaining the state series are provided in section 5.3.2. Using the estimated parameters $\hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}_1, \hat{\Sigma}_2$ characterizing the regimes, we can compare the (squared) distance of a data point x_t to each class center as follows:

$$\begin{aligned} d_{t,1} &= (x_t - \hat{\mu}_1)^\top \hat{\Sigma}_1^{-1} (x_t - \hat{\mu}_1) \\ d_{t,2} &= (x_t - \hat{\mu}_2)^\top \hat{\Sigma}_2^{-1} (x_t - \hat{\mu}_2). \end{aligned}$$

Elevated levels in $d_{t,1}$ or $d_{t,2}$ indicate turbulence under \mathcal{C}_1 or \mathcal{C}_2 , respectively. However, these turbulence indicators lead to different interpretations. Large values in $d_{t,1}$ suggest that observed events are unusual within the context of \mathcal{C}_1 , i.e., they are atypical during falling markets. This type of turbulence is referred to as *good turbulence*. Conversely, large values in $d_{t,2}$ suggest a departure from typical positive market conditions, termed *bad turbulence*.

In the subsequent analysis, apart from distinguishing turbulent from quiet episodes, we further compare $d_{t,1}$ and $d_{t,2}$ to assess whether x_t is more likely to belong to \mathcal{C}_2 than to \mathcal{C}_1 .

To interpret the difference $d_{t,1} - d_{t,2}$ meaningfully, it is essential to appropriately scale the two turbulence values. This consideration leads us to delve into the theory of discriminant analysis.

5.3.1 Turbulence and (quadratic) discriminant analysis

Discriminant analysis is a supervised classification method making use of Bayes' theorem. The objective is to find the discriminant, i.e., the function which separates a dataset into two groups, by equating the posterior probabilities of x belonging to either class \mathcal{C}_1 or \mathcal{C}_2 as:

$$\mathbb{P}(x \in \mathcal{C}_1|X = x) \stackrel{\text{set}}{=} \mathbb{P}(x \in \mathcal{C}_2|X = x). \quad (5.4)$$

Note that for this excursion on discriminant analysis we drop the time index t since described methods assume data to be iid. We will later on propose a method which can be seen as an adaptation of classic QDA for financial time series data.

By Bayes theorem, (5.4) can be decomposed into

$$\frac{f_1(x)\pi_1}{\sum_{i=k}^2 \mathbb{P}(x \in \mathcal{C}_k|X = x)\pi_k} \stackrel{\text{set}}{=} \frac{f_2(x)\pi_2}{\sum_{i=k}^2 \mathbb{P}(x \in \mathcal{C}_k|X = x)\pi_k} \quad (5.5)$$

$$\implies f_1(x)\pi_1 = f_2(x)\pi_2.$$

$f_k(x) = \mathbb{P}(X = x|x \in \mathcal{C}_k)$ and $\pi_k = \mathbb{P}(x \in \mathcal{C}_k)$ are the class conditional likelihood and the prior probability, respectively, and the denominator is the marginal probability, which cancels out.

In the case where f represents the probability density function (pdf) of the multivariate Normal distribution, we get

$$\frac{1}{(\sqrt{2\pi})^d|\Sigma_1|} \exp\left(-\frac{(x - \mu_1)^\top \Sigma_1^{-1}(x - \mu_1)}{2}\right) \pi_1 \stackrel{\text{set}}{=} \frac{1}{(\sqrt{2\pi})^d|\Sigma_2|} \exp\left(-\frac{(x - \mu_2)^\top \Sigma_2^{-1}(x - \mu_2)}{2}\right) \pi_2.$$

Taking the natural logarithm, multiplying with 2 and bringing everything to the right side we obtain

$$\begin{aligned}
0 &= [-\ln(|\Sigma_2|) - (x - \mu_2)^\top \Sigma_2^{-1} (x - \mu_2) + 2 \ln(\pi_2)] \\
&\quad - [-\ln(|\Sigma_1|) - (x - \mu_1)^\top \Sigma_1^{-1} (x - \mu_1) + 2 \ln(\pi_1)] \\
&= \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + 2 \ln\left(\frac{\pi_2}{\pi_1}\right) + x^\top (\Sigma_1^{-1} - \Sigma_2^{-1})x + 2(\Sigma_2^{-1}\mu_2 - \Sigma_1^{-1}\mu_1)^\top x + \mu_1^\top \Sigma_1^{-1}\mu_1 - \mu_2^\top \Sigma_2^{-1}\mu_2 \\
&= \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + 2 \ln\left(\frac{\pi_2}{\pi_1}\right) + \Delta_d.
\end{aligned}$$

The symbol Δ_d is used to denote the difference in the squared Mahalanobis distances which we have written out in a quadratic form $x^\top Ax + b^\top x + c$.⁶

The discriminant function in QDA reads

$$\delta(x) := \Delta_d + \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + 2 \ln\left(\frac{\pi_2}{\pi_1}\right). \quad (5.6)$$

The decision rule in QDA follows as

$$x \in \begin{cases} \mathcal{C}_1 & \text{if } \Delta_d \leq \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + 2 \ln\left(\frac{\pi_2}{\pi_1}\right) \\ \mathcal{C}_2 & \text{else.} \end{cases} \quad (5.7)$$

Bose et al. (2015) have generalized QDA to the family of symmetrically elliptical distributions having a pdf of the form

$$f(x) = \frac{1}{|\Sigma|^{\frac{1}{2}}} g((x - \mu)^\top \Sigma^{-1} (x - \mu))$$

resulting in decision rule

$$x \in \begin{cases} \mathcal{C}_1 & \text{if } \Delta_d \leq C \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) \\ \mathcal{C}_2 & \text{else} \end{cases} \quad (5.8)$$

for some scaling constant C which may depend on x and assuming $\pi_1 = \pi_2$.

When $C = 1$, the decision rule aligns with ordinary QDA. In the case of $C = 0$, the class with the minimum Mahalanobis distance is chosen. Bose et al. (2015) propose a

⁶Where $A = \Sigma_1^{-1} - \Sigma_2^{-1}$, $b = \Sigma_2^{-1}\mu_2 - \Sigma_1^{-1}\mu_1$ and $c = \mu_1^\top \Sigma_1^{-1}\mu_1 - \mu_2^\top \Sigma_2^{-1}\mu_2$.

procedure for determining an optimal value for C through cross-validation.

Returning to our application, we needed a method to scale $d_{t,1}$ and $d_{t,2}$ before assessing their difference. We suggest a nonparametric approach, scaling the turbulence measures $d_{t,1}$ and $d_{t,2}$ by their respective sample variances, denoted as $\sigma_{t,1}$ and $\sigma_{t,2}$. We then define the difference between good and bad turbulence using the quadratic form:

$$\begin{aligned}
\Delta_t &= \sigma_{t,1}^{-1}d_{t,1} - \sigma_{t,2}^{-1}d_{t,2} \\
&= x_t^\top \left[(\sigma_{t,1}\hat{\Sigma}_1)^{-1} - (\sigma_{t,2}\hat{\Sigma}_2)^{-1} \right] x_t \\
&\quad + \left[2(\sigma_{t,2}\hat{\Sigma}_2)^{-1}\mu_2 - 2(\sigma_{t,1}\hat{\Sigma}_1)^{-1}\mu_1 \right]^\top x_t \\
&\quad + \mu_1^\top (\sigma_{t,1}\hat{\Sigma}_1)^{-1}\mu_1 - \mu_2^\top (\sigma_{t,2}\hat{\Sigma}_2)^{-1}\mu_2 \\
&= x_t^\top \tilde{A}x_t + \tilde{b}^\top x_t + \tilde{c}.
\end{aligned} \tag{5.9}$$

Applied to hedging, we define the decision rule

$$x_t \in \begin{cases} \mathcal{C}_1 & \text{if } \Delta_t \leq 0 \\ \mathcal{C}_2 & \text{else.} \end{cases} \tag{5.10}$$

Correspondingly, we define the following hedging rule for the upcoming empirical analysis:

$$R_2 : \quad h(x_t) = \begin{cases} 1 & \text{if } x_t \in \mathcal{T} \cap \mathcal{C}_1 \\ 0 & \text{else.} \end{cases}$$

We term our solution nonparametric because the scaling factor is solely data-driven and not derived from a distributional assumption about \mathcal{C}_1 and \mathcal{C}_2 . However, akin to QDA and generalized QDA, the training of our relative turbulence indicator/classifier involves parameter estimation within each class.

In practice, the parameters in the quadratic form of the classifier (see Eq. 5.10) are unknown and must be estimated from the training data. This necessitates a mechanism for labeling the training set into bear and bull states. Unfortunately, the state variable is latent and cannot be identified without ambiguity. Consequently, we discuss potential procedures to unveil historical states in the following section.

The rationale for distinguishing between bear and bull phases, as opposed to a simplistic classification based solely on the sign of market returns, which is commonly done in the

literature, rests on three considerations. Firstly, relying on the sign of market returns introduces a dependence on the chosen data frequency (e.g., daily, weekly, or monthly returns), potentially limiting system functionality to a specific temporal resolution. Secondly, categorizing returns as either positive or negative based on their sign fails to discriminate between returns close to zero and those significantly deviating from it in absolute terms. Thirdly, the overarching objective is to derive a discerning signal for identifying opportune and adverse investment periods. The desired indicator should demonstrate persistence and parsimony in issuing trading signals, selectively activating trades in the presence of genuine shifts in market dynamics. The inherent volatility in the sign of market returns renders it inadequate for this purpose.⁷

5.3.2 Bears and bulls

Wall street jargon uses bears and bulls as symbols to differentiate between positive and negative market developments respectively. However, these terms lack a precise definition regarding the duration or extent of price changes that categorize a phase as bearish or bullish. Despite their significance, there is surprisingly limited literature addressing this issue. [Kole and van Dijk \(2017\)](#) distinguish between parametric and semi-parametric models. Parametric procedures find their foundation in the seminal work of [Hamilton \(1989\)](#) who applied a Markov switching model to the quarterly growth rates of the American gross domestic product and derived an optimal dating rule from inferred smoothed probabilities. Such tools rely on a specification of a statistical model for the return distributions in either states and use the estimated parameters of the models to obtain the chronology of turning points. The latent state variable follows a Markov chain. Building on this, [Maheu and McCurdy \(2000\)](#) applied a Markov switching approach to internally identify a high mean, low variance bull state and a low mean, high variance bear state.

Semi-parametric methods are rule based procedures operating directly on the non-stationary price series, rather than on the invariants, seeking local peaks and troughs. [Lunde and Timmermann \(2004\)](#) developed an algorithm that incorporates a magnitude parameter for both bear and bull phases. According to their approach, for a phase to be classified as bearish or bullish, stock market prices must demonstrate significant changes

⁷Even if we can predict the sign with a probability exceeding 50%, the real-world imposition of transaction costs would likely erode any positive paper returns.

from their previous peak or trough.

Pagan and Sossounov (2003) defined bear and bull markets by adapting a pattern recognition technique initially formulated by Bry and Boschan (1971) for detecting turning points in business cycles. This method underwent further refinement by Harding and Pagan (2002). Their algorithm segments a time series into upward and downward trending phases, relying on two duration parameters that specify the minimum phase and cycle lengths.

In the forthcoming sections, we amalgamate the concepts proposed by Harding and Pagan (2002) and Lunde and Timmermann (2004). This integration involves the inclusion of a magnitude parameter into the pattern recognition technique. Additionally, subtle adjustments are made to enhance its adaptability to data frequencies beyond the initial monthly or quarterly requirements.

The methodology uses as input the prices of an equity series, denoted as $\{p_t\}_{t=1}^T$, along with two user-defined control variables, namely B and b . These control variables determine the minimum length of a cycle and the minimum length of a phase, respectively. The algorithm's objective is to identify turning points, characterized as local maxima and minima (peaks and troughs), based on the following criterion: a peak occurs at time t if $p_{t-\tau}, \dots, p_{t-\tau+1} < p_t > p_{t+1}, \dots, p_{t+\tau}$ (similarly, a trough is identified when $p_{t-\tau}, \dots, p_{t-\tau+1} > p_t < p_{t+1}, \dots, p_{t+\tau}$).

Employing a recursive scheme, the identified turning points undergo further processing to satisfy two conditions. Initially, peaks and troughs are adjusted to alternate; thus, if two consecutive peaks or troughs are detected, the algorithm eliminates the peak or trough with the lower or higher value of p_t , respectively. Subsequently, the algorithm ensures that a cycle phase (from peak to trough or from trough to peak) lasts for a minimum of b dates, while the complete cycle (from peak to peak or from trough to trough) has a minimum duration of B dates. However, an exception may arise if the price change since the last local extremum exceeds a certain magnitude (in absolute terms), determined by an additional input parameter denoted as θ . As a result, the algorithm produces a state variable $y_t \in \{1, 2\}$, where the value 1 signifies bear markets, and the value 2 corresponds to bull markets.

The categorization into bear and bull phases using the described technique is inherently influenced by the chosen parameter values. To enhance the robustness of the procedure, we adopt an ensemble approach, deriving a state variable by averaging (and rounding to

the next integer) over a grid of parameter values, where b varies across $\{10, 15, 21, 31, 42\}$ and B across $\{30, 45, 63, 93, 126\}$. For the magnitude parameters, we maintain a consistent value of $\theta = 0.2$.

5.3.3 Smoothing

The methodology we devised to distinguish positive from negative turbulence demonstrates strong statistical predictive power for short-term forecasts of the latent state. However, this statistical accuracy does not translate into economic rewards when realistic transaction costs are considered. Figure 5.3 illustrates the impressive performance difference of a hedging strategy based on decision rule (5.10), derived from our index of relative turbulence Δ_t . The strategy involves taking a long position in the MSCI World Index along with a short position in the index as a full hedge whenever the indicator is negative. Despite the impressive performance before considering costs, the strategy's net returns are adversely impacted by transaction costs, assuming costs (fees and spreads) ranging from 10 to 20 basis points (Bp).

It is evident that a reduction in trading activity is necessary to achieve a satisfactory balance between trading rewards and costs. The simplest approach involves smoothing the indicator on a rolling basis, such as through filtering with a simple or exponentially weighted moving average. Our tests indicate that an appropriate balance can be achieved with smoothing. However, the challenge then lies in setting the smoothing parameter, which poses a risk of overfitting. Despite running the backtest out-of-sample for a given smoothing rule, the smoothing parameter is chosen in-sample, making it a potentially dangerous game.

We propose leveraging upon the wealth of econometric tools designed to address the stylized facts introduced in Section 5.2.1 and to model a smoothly⁸ varying turbulence measure in a fully data-driven manner using a series of univariate GARCH or SV processes.

Recall from Equation (5.3) that turbulence is driven by the sum of variance and covariance effects, and their dynamics can be effectively captured by volatility models. To avoid the need to model dependence in the covariance terms, we first apply a change of

⁸The term *smoothly* is used informally to convey the idea of a less rough or fluctuating indicator, rather than in a strict mathematical sense.

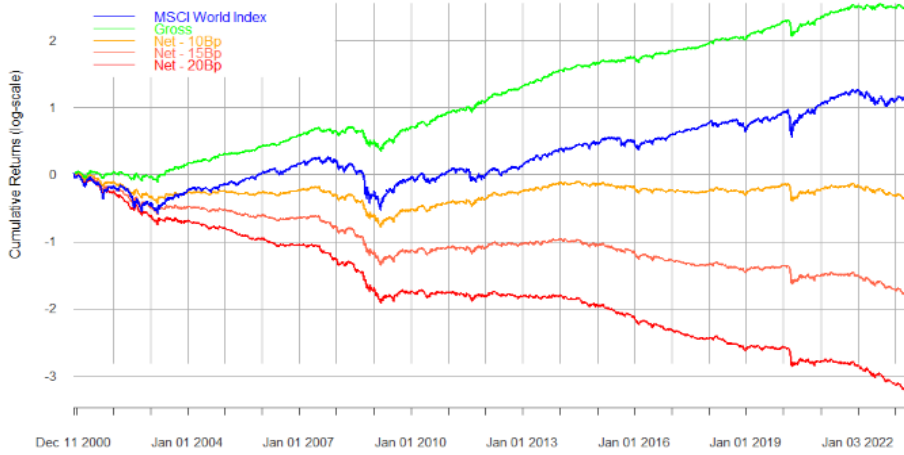


Figure 5.3: Impact of transaction costs on the cumulative returns attained by hedging rule R_2 .

base by rotating demeaned return series with the orthonormal⁹ matrix of eigenvectors Γ of the static sample covariance matrix to obtain uncorrelated principal components using the map $X \mapsto S = (X - \mu)\Gamma$. Recall that we can factorize the covariance matrix into canonical form $\Sigma = \Gamma\Lambda\Gamma^{-1}$, where Λ is a diagonal matrix of ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ and is equivalent to the covariance matrix of the components. Because Λ is a diagonal matrix, its inverse is also diagonal with $[\Lambda^{-1}]_{ii} = \frac{1}{\lambda_i}$. Hence, in transformed space, the squared Mahalanobis distance reduces to the squared Euclidean distance:

$$d_t = (x_t - \hat{\mu})^\top \hat{\Sigma}^{-1} (x_t - \hat{\mu}) = z_t^\top \hat{\Sigma}^{-1} z_t = z_t^\top (\Gamma\Lambda\Gamma^{-1})^{-1} z_t = z_t^\top \Gamma\Lambda^{-1}\Gamma^\top z_t \quad (5.11)$$

where $z_t = x_t - \hat{\mu}$. Rewriting $s_t = \Gamma^\top z_t$,

$$d_t = s_t^\top \Lambda^{-1} s_t = \sum_{i=1}^n s_{t,i}^2 / \lambda_i = \|\tilde{s}_t\|_2^2 \quad ; \tilde{s}_t = \lambda^{-\frac{1}{2}} s_t. \quad (5.12)$$

Turbulence can thus be formulated as the sum of standardized squared components¹⁰. This formulation shows that the dynamics of turbulence depends of the sum of squared components. If, analogous to squared returns, squared components exhibit similar auto-correlation, it makes sense to capture this dynamic with a time series model. We suggest

⁹Each column vector has length one and is orthogonal to all the other column vectors. For square orthonormal matrices, the inverse is simply the transpose.

¹⁰Note that under $x \sim N(\mu, \Sigma) \rightarrow \lambda^{-1/2} s \equiv \tilde{s} \sim N(0, 1)$ and $\sum_{i=1}^n \tilde{s}_i^2 \equiv d \sim \chi^2(n)$.

fitting a GARCH or SV model. From a modelling aspect it may seem odd to consider GARCH and SV as exchangeable tools since they fundamentally differ in the construction and assumptions on the volatility states. Nevertheless, we do so because inferentially, both processes lead to similar results in our context. For the following analytical presentation, we use a standard SV model.

Given observations $s_t = \Gamma^\top z_t$ in the orthogonal space, the SV model assumes the following structure for the i -th component $s_{t,i}$:

$$s_{t,i} = \exp(h_{t,i}/2)\epsilon_{t,i} \quad \epsilon_{t,i} \in N(0, 1) \quad (5.13)$$

$$h_{t,i} = \mu_i + \varphi_i(h_{t-1,i} - \mu_i) + \nu_i\eta_{t,i} \quad \eta_{t,i} \in N(0, 1) \quad (5.14)$$

$h_{t,i}$ denotes the value of the log-variance process at day t for component i which is assumed to follow a stationary AR(1) process with persistence $|\varphi_i| < 1$. The standard SV model employs three parameters $\vartheta_i = (\mu_i, \varphi_i, \nu_i)$ which are to be estimated. μ_i refers to the level, φ_i to the persistence, and ν_i to the standard deviation of the log-variance.

The log-variance process is initialized by $h_0 | (\mu_i, \nu_i, \varphi_i) \sim N(\mu_i, \frac{\nu_i^2}{1-\varphi_i^2})$. Then, it holds that $h_{t,i} | (h_{t-1,i}, \mu_i, \nu_i, \varphi_i) \sim N(\mu_i + \varphi_i(h_{t-1,i} - \mu_i), \nu_i^2)$.

Our approach is to plugin the squared components $s_{t,i}^2 = \exp(h_{t,i})\epsilon_{t,i}^2$ into equation (5.12) and to replace the noise term $\epsilon_{t,i}^2 \sim \chi^2(1)$ with its expected value $\mathbb{E}(\epsilon_{t,i}^2) = 1$.

Doing so, we retrieve the following persistent measure of turbulence

$$\mathcal{D}_t = \sum_{i=1}^n \lambda_i^{-1} e^{\mu_i + \varphi_i(h_{t-1,i} - \mu_i) + \nu_i\eta_{t,i}}. \quad (5.15)$$

Repeating steps (5.11) - (5.15) for both regime specific mean and covariance estimators, i.e., $\hat{\mu}_1, \hat{\Sigma}_1$ for regime \mathcal{C}_1 and $\hat{\mu}_2, \hat{\Sigma}_2$ for regime \mathcal{C}_2 , we further obtain persistent measures of good and bad turbulence separately, denoted $\mathcal{D}_{t,1}$ and $\mathcal{D}_{t,2}$. As for the computation of Δ_t , we scale both terms by their empirical variances. The scaled turbulence measures, denoted $\tilde{\mathcal{D}}_{t,1}$ and $\tilde{\mathcal{D}}_{t,2}$, define a discriminant as

$$\psi_t = \tilde{\mathcal{D}}_{t,1} - \tilde{\mathcal{D}}_{t,2} \quad (5.16)$$

giving raise to decision rule

$$x_t \in \begin{cases} \mathcal{C}_1 & \text{if } \psi_t < 0 \\ \mathcal{C}_2 & \text{else} \end{cases} \quad (5.17)$$

and adjusted hedging rule

$$R_2 : \quad h(x_t) = \begin{cases} 1 & \text{if } x_t \in \mathcal{T} \cap \mathcal{C}_1 \\ 0 & \text{else.} \end{cases} \quad \mathcal{T} := \{\mathcal{D}(x) \in \mathbb{R} \mid \mathcal{D}(x) \geq P^{-1}(p)\}$$

5.3.4 Refinements

In consideration of practical application, we propose several refinements to the previously introduced theoretical concepts, aiming to enhance the overall performance of the system.

Minimum torsion. When employing an eigenvalue transformation, there is a possibility that the last components may not exhibit autocorrelation in the squared values, making the fitting of a GARCH or SV model inappropriate. One approach is to simply ignore the last components, which is equivalent to dimensionality reduction. The advantage of such an approach lies in the saving of computation time and the filtering out of noise. However, the challenge is to determine at which component to set the cutoff.

An alternative is to utilize a linear transformation that minimally disrupts the original return series. This concept of minimum (linear) torsion (MT) was introduced by [Meucci et al. \(2013\)](#). The authors developed a numerical procedure to determine the transformation that minimizes the tracking error concerning the return series while maintaining zero correlation among the transformed components. Their algorithm efficiently solves the following quadratically constrained quadratic program:

$$\Gamma_{MT} = \arg \min_{Cor(Z\Gamma)=I} \sqrt{\frac{1}{n} \sum_{i=1}^n \mathbb{V} \left(\frac{(Z\Gamma)_i - Z_i}{\mathbb{V}(Z_i)} \right)} \quad (5.18)$$

where $Z = X - \mu$. We refer to [Meucci et al. \(2013\)](#) for details on the solution procedure.

Robust estimators. Robust estimators of multivariate location and dispersion may be preferable for the computation of the Mahalanobis metric. Various options are available (see e.g., Olive (2004)). In our empirical study we employ the quadratic shrinkage approach of Ledoit and Wolf (2022) for the covariance matrix and the sample median for the location.

Weighted Mahalanobis distance. Given the upcoming empirical analysis, where we seek to improve upon a buy-and-hold investment in the capitalization-weighted MSCI World index, it is reasonable to incorporate a weighting of the country indices proportional to their constituents' weights in the parent index. Therefore, we compute turbulence by

$$d_t = (W(x_t - \hat{\mu}))^\top \hat{\Sigma}^{-1} (W(x_t - \hat{\mu})) = (x_t - \hat{\mu})^\top W^\top \hat{\Sigma}^{-1} W (x_t - \hat{\mu}) \quad (5.19)$$

where $W = \text{diag}(w)$ is a diagonal matrix of weights. Notice that weighting the data does not change the correlation structure but effects the variance-covariance matrix.

Setting $z_t = x_t - \hat{\mu}$ and using the spectral decomposition $\hat{\Sigma} = \Gamma \Lambda \Gamma^{-1}$ we can further write:

$$d_t = z_t^\top W^\top \Gamma \Lambda^{-1} \Gamma^\top W z_t = s_t^w{}^\top \Lambda^{-1} s_t^w = \sum_{i=1}^n s_{t,i}^w{}^2 / \lambda_i = \|\tilde{s}_t^w\|_2^2$$

where $s_t^w = \Gamma^\top W z_t$ and $\tilde{s}_t^w = \lambda^{-\frac{1}{2}} s_t^w$.

5.4 Empirical study

This section aims to empirically test the theoretical concepts discussed earlier, with the goal of assessing the effectiveness of the introduced measures of turbulence in guiding investment decisions. Answering this question is not straightforward and is fraught with ambiguity. Two main avenues are available: statistical inference and backtesting. Statistical inference allows for conclusions regarding the explanatory and predictive power of the indicators, while backtesting serves as a reality check. However, statistical significance in the former case does not guarantee significant financial rewards in the latter due to real-world frictions like transaction costs, as illustrated in the example visualized in Figure 5.3.

The tension between statistical accuracy and economic profitability prompts us to

restrict the use of statistical analysis to descriptive comparisons of different turbulence measures. Ultimately, the practical utility of these measures hinges on economic considerations and needs to be addressed accordingly.

Our focus is, therefore, on a backtesting framework, which we will describe shortly. Before delving into that, however, we need to further clarify the objective of the hedging exercise. In the initial stage, the goal is to enhance efficiency in terms of risk-adjusted returns. However, solely achieving a high Sharpe ratio is not compelling if it results from excessive risk reduction. Ideally, we aim for an increased Sharpe ratio while matching the return of the buy-and-hold index. This is a challenging objective, as we do not permit any form of leverage, meaning that all performance must be gained through timing.¹¹

A second goal is to mitigate drawdowns. Consequently, the performance statistics of interest include average (geometric) return, volatility, Sharpe ratio, and the five largest drawdowns.

In the following we provide a description of used stock market data and proceed with a statistical analysis of introduced turbulence measures before backtesting turbulence based binary hedging strategies.

5.4.1 Data and backtesting framework

Data. Our empirical examination relies on index return data sourced from the index provider MSCI. The MSCI World Index is a globally recognized stock market index that monitors the performance of large and mid-cap companies spanning 23 developed countries (as of the time of writing). To conduct an out-of-sample backtest, we incorporate Greece, which was reclassified by MSCI as an emerging market in 2013, and exclude Israel before May 2009, when it attained the status of a developed market.

Our computations are based on daily total¹² return prices spanning the period from 1999-01-01 to 2023-04-18.

Backtesting framework. All turbulence measures are constructed out-of-sample using an expanding window so that each element in the resulting indicator series is guaranteed

¹¹In the industry, asset managers often use a static portfolio as a benchmark, holding 50% in the index and 50% in cash, or they choose a weighting such as to calibrate the volatility of the static portfolio to that of the hedged strategy (which is necessarily lower than that of the index since it is not invested at times). We, however, keep the (MSCI World) index as our base strategy that we seek to improve upon.

¹²I.e., dividends are reinvested.

to be only computed on data available up to that point in time. This involves a series of computational steps which are repeated daily. At every day $\tau \in \{k, k + 1, \dots, T\}$ in the history starting at some point k , define the training data $\{x_t\}_{t=1}^\tau$ to contain returns from the first available day $t = 1$ to the current day $t = \tau$ and proceed as follows:

1. Estimate $\hat{\mu}_\tau = \mathbb{E}(\{x_t\}_{t=1}^\tau)$ and $\hat{\Sigma}_\tau = \text{Cov}(\{x_t\}_{t=1}^\tau)$ to get $d_\tau = (x_\tau - \hat{\mu}_\tau)^\top \hat{\Sigma}_\tau^{-1} (x_\tau - \hat{\mu}_\tau)$.
1. Compute the binary bear/bull state sequence $\{y_t\}_{t=1}^\tau$ on the benchmark price series (i.e., the MSCI World Index in our application) using the pattern recognition technique described in section 5.3.2.
2. Estimate $\hat{\mu}_{\tau,1} = \mathbb{E}(\{x_t\}_{\mathbb{1}(\{y_t\}_1^\tau=1)})$, $\hat{\Sigma}_{\tau,1} = \text{Cov}(\{x_t\}_{\mathbb{1}(\{y_t\}_1^\tau=1)})$ and $\hat{\mu}_{\tau,2} = \mathbb{E}(\{x_t\}_{\mathbb{1}(\{y_t\}_1^\tau=2)})$, $\hat{\Sigma}_{\tau,2} = \text{Cov}(\{x_t\}_{\mathbb{1}(\{y_t\}_1^\tau=2)})$. and compute Δ_τ as in equation 5.9.
3. Transform the training data, i.e., $\{x_t\}_{t=1}^\tau$, by a MT decomposition (equation 5.18) of $\hat{\Sigma}_{\tau,1}$ and $\hat{\Sigma}_{\tau,2}$ to get uncorrelated components $\{s_{t,i,1}\}_{t=1}^\tau$ and $\{s_{t,i,2}\}_{t=1}^\tau$, $i = 1, \dots, n$.
4. Fit a SV model to the components obtained in the previous step.
5. Compute $\psi(x_\tau)$ by equations 5.15 and 5.16.

Notice that the re-fitting of the orthogonal SV model is a time-intensive task, requiring daily repetition as the MT-components undergo a complete historical change with each new data point in the return series. We suggest utilizing the state-of-the-art open-source platform STAN¹³ for this purpose.

To keep things simple and replicable, applied hedging rules neutralize the equity exposure during risky episodes by selling the long position in the MSCI World index (i.e., holding cash at a zero interest rate). In a real-world application it would be natural to short index futures. Also, it would be wiser to build a more granular hedging scheme rather than the binary full-hedge or no-hedge approach followed here for simplicity.

Tested strategies account for transaction cost of 0.1% which generates a strong penalty on excessive trading. Furthermore, a lag of one day between computation and trade execution is considered.

¹³See <https://mc-stan.org/>.

We begin our study with an evaluation of hedging rule R_1 from which we derive the following sub-rules:

$$\begin{aligned}
R_{1,a} : h(x_\tau) &= 1 \quad \text{if } d_\tau \geq \hat{F}_n^{-1}(p) \quad \text{else } 0. \\
R_{1,b} : h(x_\tau) &= 1 \quad \text{if } \mathcal{D}_\tau \geq \hat{F}_n^{-1}(p) \quad \text{else } 0. \\
R_{1,c} : h(x_\tau) &= 1 \quad \text{if } d_\tau \geq \chi_n^{-2}(p) \quad \text{else } 0. \\
R_{1,d} : h(x_\tau) &= 1 \quad \text{if } \mathcal{D}_\tau \geq \chi_n^{-2}(p) \quad \text{else } 0. \\
R_{1,VIX} : h(x_\tau) &= 1 \quad \text{if } VIX > 25 \quad \text{else } 0.
\end{aligned}$$

$R_{1,VIX}$ serves as a straightforward benchmark strategy, implementing a hedge whenever the CBOE Volatility Index¹⁴ (VIX) is elevated (considered as $VIX > 25$). While the VIX is specific to the U.S. market, it can act as a proxy for the risk of the MSCI World index, given the latter's substantial investment in U.S. stocks ($\sim 70\%$ as of the time of writing). The variations among rules $R_{1,a}$ to $R_{1,d}$ stem from differences in the turbulence measure's smoothing using the orthogonal SV filter and the choice of the inverse distribution defining the turbulent and quiet thresholds (empirical versus χ^2). In this context, we set the level of p to 0.9. It's noteworthy that the empirical percentile method's results are highly sensitive to the choice of p . While a value of $p \approx 0.8$ yields the best backtest results, outperforming the index in all performance measures, we believe the method's sensitivity to the percentile threshold's parameterization disqualifies it for further use. Conversely, the cutoff percentile based on the theoretical χ^2 model has minimal impact on results for a broad range of values, especially when SV-smoothing is additionally applied.

In a subsequent analysis, we examine strategies associated with hedging rule R_2 , which overrides $R_{1,c}$ and $R_{1,d}$ when our 'normative' discriminant $\psi(x_\tau)$ categorizes turbulences as *good*.

¹⁴See https://www.cboe.com/tradable_products/vix/.

$$\begin{aligned}
R_{2,c} : h(x_\tau) &= 1 \quad \text{if } (R_{1,c} = 1) \cap (\Delta_\tau < 0) \quad \text{else } 0 && \text{see eq. (5.9).} \\
R_{2,d} : h(x_\tau) &= 1 \quad \text{if } (R_{1,d} = 1) \cap (\psi_\tau < 0) \quad \text{else } 0 && \text{see eq. (5.16).} \\
R_{2,QDA} : h(x_\tau) &= 1 \quad \text{if } x_\tau \in \mathcal{C}_1 \quad \text{else } 0 && \text{see eq. (5.7).}
\end{aligned}$$

$R_{2,QDA}$ denotes the benchmark strategy for this second application which employs ordinary QDA to exit the equity exposure whenever the model classifies the cross-section of returns to come from a bear regime (as defined in section 5.3.2).

5.4.2 Descriptive analysis

Given that our proposed hedging rules are binary in nature, we evaluate their performance using standard classification model metrics. Two scenarios are considered for true class labels: firstly, utilizing the bear-bull labels, as outlined in section 5.3.2, derived in-sample from the complete return series of the MSCI World Index; secondly, based on the sign of daily index returns. The predictors are computed out-of-sample and lagged by one day.

In our context, an effective predictor should strike a balance between minimizing both type I and type II errors. A type II error (false negative) occurs when we fail to initiate a hedge when warranted, leading to losses due to adverse market developments. Conversely, a type I error (false alarm) arises when we mistakenly implement a hedge when unnecessary, resulting in trading costs and missed positive returns.

Table 5.1 provides an overview of the performance of the classification rules. Caution is warranted when interpreting the relatively high accuracy values observed in the two percentile-based rules ($R_{1,a}$ and $R_{1,b}$) and the VIX-based benchmark model. These models excel in predicting non-events, as they are inherently conservative in triggering a hedge. However, they exhibit shortcomings in accurately identifying positive instances. This discrepancy is evident in their elevated specificity, representing the proportion of correctly identified actual negative instances, and concurrently low sensitivity¹⁵, reflecting the proportion of correctly identified actual positive instances. The subpar performance in

¹⁵In general, a sensitivity of 0.5 is considered relatively low, indicating that the model is not effective at capturing positive instances, and there is a significant risk of false negatives.

terms of F-scores, calculated as the harmonic mean of precision (the proportion of correctly predicted positive instances among all predicted positives) and recall (i.e., sensitivity), further underscores the models' failure to achieve a satisfactory balance between accurately identifying positive instances and avoiding false positives.

Strategies $R_{1,c}$ and $R_{1,d}$ demonstrate improvements with sensitivities of 0.47 and 0.50, respectively. Nevertheless, these models only identify approximately half of the actual positive instances, indicating a significant risk of false negatives. The same goes for $R_{2,c}$ which exhibits a sensitivity of 0.52. $R_{2,d}$ achieves the highest sensitivity at 0.62, indicating superior performance in identifying positive cases. $R_{1,VIX}$ lags behind with a sensitivity of 0.20, so does $R_{2,QDA}$ with a sensitivity at 0.41. In terms of an overall ranking, considering all performance metrics, we find that $R_{2,d}$ outperforms others, followed by $R_{2,c}$, $R_{1,c}$, $R_{1,d}$, $R_{1,b}$, $R_{1,a}$, $R_{2,QDA}$, and $R_{1,VIX}$.

Table 5.2 illustrates results when employing the sign of daily index returns as the labeling method. Overall, performance metrics decrease, while the relative ranking among different classifiers remains mostly consistent. Notably, $R_{2,c}$ stands out as an exception, demonstrating an improvement in overall performance, showcasing remarkable predictive capacity for the direction of next-day index returns. Regrettably, the statistical attractiveness of the model does not translate into economic success, as elaborated in the subsequent section.

	$R_{1,a}$	$R_{1,b}$	$R_{1,c}$	$R_{1,d}$	$R_{1,VIX}$	$R_{2,c}$	$R_{2,d}$	$R_{2,QDA}$
Sensitivity	0.20	0.23	0.47	0.50	0.20	0.52	0.62	0.41
Specificity	0.93	0.91	0.72	0.69	0.94	0.65	0.82	0.82
Accuracy	0.72	0.71	0.65	0.64	0.73	0.61	0.77	0.70
Balanced Accuracy	0.56	0.57	0.60	0.59	0.57	0.58	0.72	0.62
Precision	0.52	0.49	0.40	0.39	0.55	0.37	0.58	0.47
F-Score	0.29	0.31	0.43	0.44	0.29	0.43	0.60	0.44

Table 5.1: Performance metrics of the different hedging rule classifiers (lagged by one day). Bear and bull states computed in-sample on the entire available history of the MSCI World Index by the pattern recognition method introduced in section 5.3.2 serve as labels of the true class.

	$R_{1,a}$	$R_{1,b}$	$R_{1,c}$	$R_{1,d}$	$R_{1,VIX}$	$R_{2,c}$	$R_{2,d}$	$R_{2,QDA}$
Sensitivity	0.13	0.14	0.37	0.38	0.12	0.66	0.42	0.32
Specificity	0.91	0.88	0.69	0.65	0.91	0.81	0.79	0.81
Accuracy	0.57	0.55	0.55	0.53	0.57	0.75	0.63	0.59
Balanced Accuracy	0.52	0.51	0.53	0.52	0.52	0.74	0.61	0.56
Precision	0.53	0.46	0.48	0.46	0.53	0.73	0.61	0.57
F-Score	0.22	0.21	0.42	0.42	0.20	0.70	0.50	0.41

Table 5.2: Performance metrics of the different hedging rule classifiers (lagged by one day). As true class labels, the sign of the daily returns of the MSCI World Index is used.

5.4.3 Backtest results

Figure 5.4 presents the results of backtesting various iterations of hedging rule R_1 . Contrary to expectations from the statistical analysis, the straightforward benchmark strategy using the VIX effectively mitigates the largest drawdowns and improves efficiency before factoring in transaction costs. However, it falls short of aligning with the return of the buy-and-hold investment in the index.¹⁶ The difference between net and gross performance is relatively small because the VIX-based approach triggers relatively few trades, rarely surpassing the critical level of 25, and when it does, it tends to persist above that threshold for an extended period. The most challenging phase for the strategy unfolded in 2022, where the crossing of the critical value occurred multiple times consecutively but consistently later than optimal. This period posed challenges for many volatility-driven risk management strategies.

Rules $R_{1,a}$ and $R_{1,c}$ exhibit significant performance disparities between net and gross results due to their high trading activity. The χ^2 rule, triggering even more trades than the empirical percentile method, widens this gap further. However, the χ^2 rule proves to be more effective in protecting against drawdowns. Specifically, $R_{1,a}$ struggles to identify slowly developing crises, such as the extended bear phase following the dot-com bubble and the unfolding of the great financial crisis in 2007. In contrast, it reacts promptly to rapid and V-shaped shocks, as seen in the response to the Covid-crash in 2020.

Both hedging rules exhibit enhanced stability when turbulence is computed with the orthogonal SV model. Rule $R_{1,d}$, in particular, excels in mitigating drawdowns.

¹⁶It's worth noting that alternative threshold choices, such as $VIX > 30$, yield somewhat improved performance. To avoid overfitting, we adhere to the 25 threshold, selected based on economic intuition before conducting experiments.

Though, this comes at the cost of extended periods out of the market, resulting in missed opportunities during positive market developments.

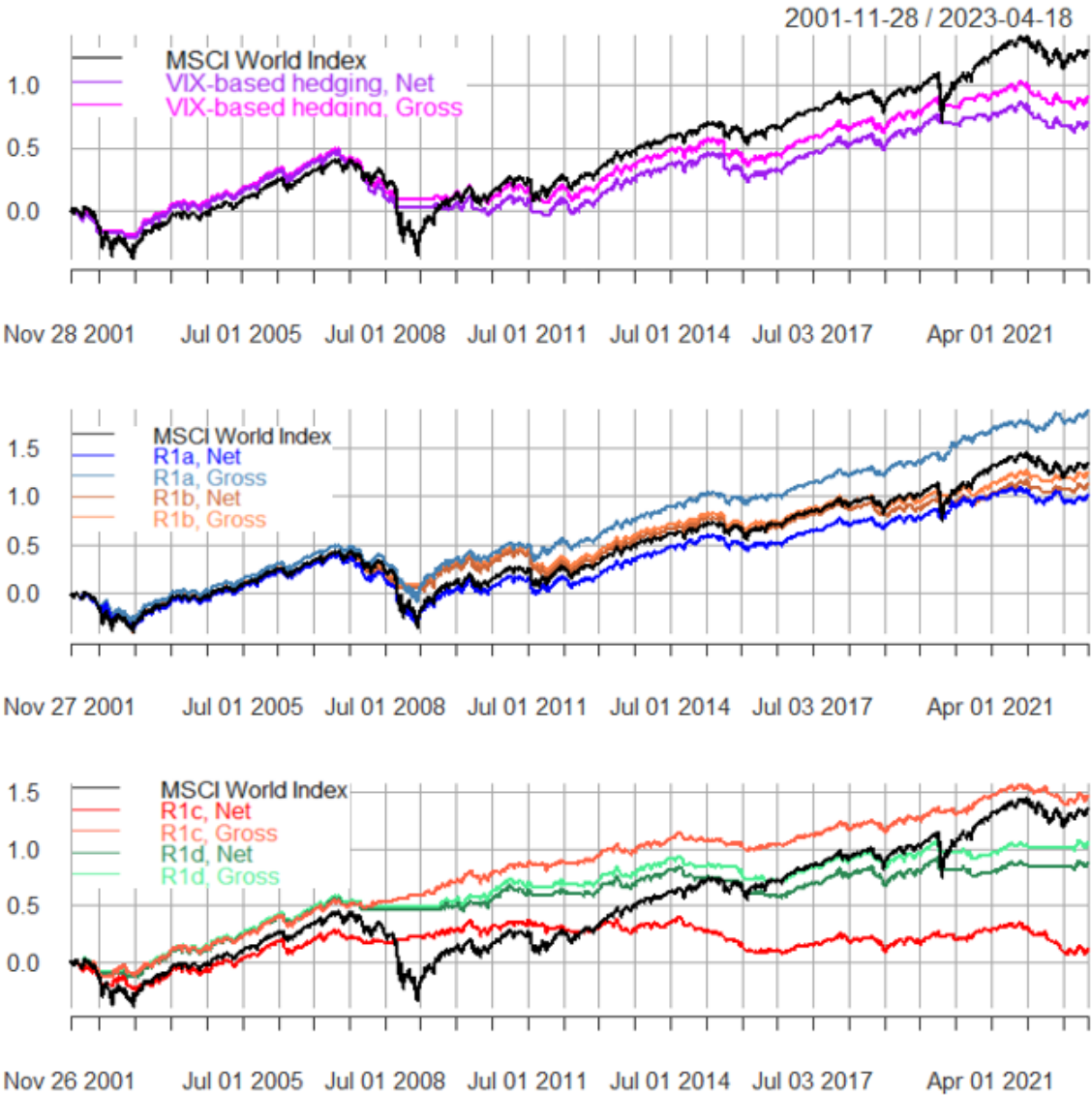


Figure 5.4: Cumulative returns (log-scale) of strategies related to hedging rule R_1 . Net series include transaction costs of 10Bp.

The decision to abstain from hedges during periods of elevated turbulence, as long as they are classified as favorable, leads to a substantial improvement in returns without compromising on protection. Similar to previous observations, the use of the SV filter plays a pivotal role in curbing trading costs. The benchmark strategy associated with the second hedging rule, ordinary QDA, exhibits promising performance on theoretical grounds. However, the excessive trading it generates proves to be prohibitively costly in practice. The cumulative returns of the backtests related to hedging rule R_2 are shown in

figure 5.5.

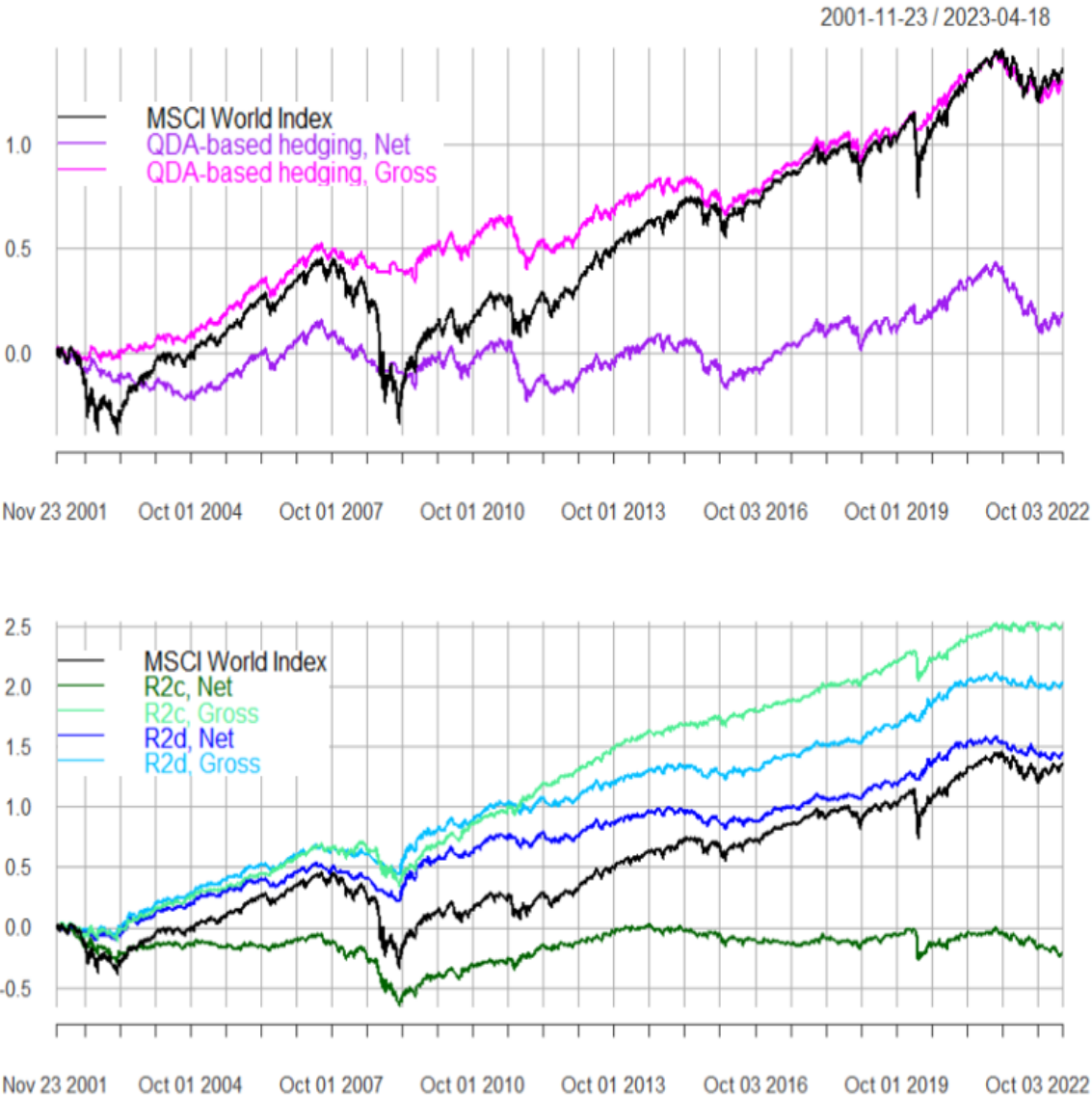


Figure 5.5: Cumulative returns (log-scale) of strategies related to hedging rule R_2 , Net series include transaction costs of 10Bp.

Performance statistics for all strategies are presented in tables 5.3 and 5.4. For a more in-depth drawdown analysis, we refer to Appendix 5A, where a comparison of the five largest drawdowns per strategy is provided. It is reasonable to assert that all strategies, with the exception of $R_{1,a}$, achieve their objective of reducing drawdowns, both in terms of magnitude and duration to recovery.

	Mean (geometric, p.a.)	Std.Dev. (p.a.)	Sharpe Ratio	MaxDD
MSCI World Index	0.063	0.157	0.401	-0.548
$R_{1,a}$, Net	0.047	0.125	0.377	-0.525
$R_{1,a}$, Gross	0.089	0.125	0.713	-0.447
$R_{1,b}$, Net	0.053	0.114	0.462	-0.344
$R_{1,b}$, Gross	0.058	0.114	0.513	-0.340
$R_{1,c}$, Net	0.005	0.088	0.057	-0.281
$R_{1,c}$, Gross	0.069	0.088	0.784	-0.168
$R_{1,d}$, Net	0.041	0.082	0.498	-0.248
$R_{1,d}$, Gross	0.049	0.082	0.600	-0.215
$R_{1,VIX}$, Net	0.034	0.100	0.340	-0.408
$R_{1,VIX}$, Gross	0.044	0.100	0.437	-0.359

Table 5.3: Summary statistics for the different hedging strategies related to hedging rule R_1 . The backtests assume trading costs of 10Bp (Net) or no costs (Gross).

	Mean (geometric, p.a.)	Std.Dev. (p.a.)	Sharpe Ratio	MaxDD
MSCI World Index	0.063	0.157	0.401	-0.548
$R_{2,c}$, Net	-0.010	0.105	-0.094	-0.495
$R_{2,c}$, Gross	0.119	0.105	1.142	-0.315
$R_{2,d}$, Net	0.067	0.096	0.697	-0.253
$R_{2,d}$, Gross	0.096	0.096	0.996	-0.196
$R_{2,QDA}$, Net	0.009	0.096	0.090	-0.322
$R_{2,QDA}$, Gross	0.061	0.096	0.637	-0.230

Table 5.4: Summary statistics for the different hedging strategies related to hedging rule R_2 . The backtests assume trading costs of 10Bp (Net) or no costs (Gross).

5.4.4 Simulation study

Our aim is to show that when simulated data display the characteristics of i) volatility clustering, ii) asymmetric volatility-return dependence and iii) synchronous volatility, then our suggested turbulence indicator proves useful in the management of investment risks. We test the following two data generative processes (dgp).

Multivariate asymmetric GARCH. We generate multivariate return series by employing the generalized orthogonal GARCH model (GO-GARCH) introduced by [van der Weide \(2002\)](#). This model can be formulated to incorporate skewed student-t residuals and

asymmetric modeling of positive and negative shocks on univariate conditional variances, following the methodology outlined by [Glosten et al. \(1993\)](#). The model parameters are estimated using the daily return series of the MSCI country indices, and synthetic series are generated based on these parameters. The resulting set of synthesized series exhibits similar covariance structure, clustering, and asymmetry in the volatility process, albeit with different price evolutions.

To ensure alignment with the original data, the simulated series undergo further scaling and shifting (univariately) to match the mean and variance of the original return series. Despite these adjustments, volatility synchronization is not achieved.¹⁷

Alternative weighting. We generate counterfactual return series by adjusting historical data using a synthetic weight vector. The real returns imply weights $\pi_t = 1$, for all $t = 1, \dots, T$ trading days. We generate alternative weights $\tilde{\pi}_t \geq 0$ for $t = 1, \dots, T$, by simulating from an ARMA(1,1) process with AR and MA coefficients set to 0.95 and -0.5, respectively, and ensuring that $\sum_t \tilde{\pi}_t = T$. Factoring the returns with those weights leads to blocks of consecutive daily returns experiencing either an increase or decrease in magnitude. Importantly, the modification is applied uniformly to all cross-sectional returns through the common weight factor $\tilde{\pi}_t$, thus preserving synchronous volatility.

This straightforward approach retains the sign of historical returns while altering their magnitudes. As a result, synthetic price patterns differ significantly compared to the original data while maintaining a likely high correlation. The procedure introduces autocorrelation in the synthetic returns, which can, however, be removed via another ARMA(1,1) filter. The obtained residuals form the final artificial return series after scaling them to match the mean and variance of the original data. These synthetic time series exhibit the three desired characteristics.

Results. Figures 5.6 and 5.7 illustrate the results obtained through backtesting of hedging rule R_2 on 100 artificially generated datasets. The graphs depict the percentage-points enhancements in the performance metrics of our hedging strategies compared to their synthetic buy-and-hold indices which is always an equally weighted portfolio of the

¹⁷This lack of synchronization becomes apparent when examining plots of rolling volatilities (or garch-implied volatilities) of the synthetic series, which are omitted here due to space constraints. The correlation between these synthetic series is about 0.2, in contrast to a correlation of roughly 0.7 observed in rolling volatilities of real returns.

simulated series. For instance, an improvement of .05 in standard deviation indicates a reduction by 5 percentage-points in the annualized value. For mean and Sharpe ratio the interpretation would be an increase in the annualized statistics by 5 percentage-points. For drawdown, the y-scale shows the percentage-point reduction of the sum of the five largest drawdowns.

Our focus is on strategy $R_{2,d}$ ($R_{2,c}$ is added to confirm the method’s theoretical strength observed on real data and the strong penalty it endures from transaction costs). As expected, risk metrics improve under both dgp. This is obvious for volatility since we introduce zero-returns during the days where the hedge is held. The improvement in terms of drawdown can be explained by the asymmetric volatility-return dependence and the persistence in the second moments under either dgp. Average returns, on the other hand, are slightly negatively affected by the hedging for the GARCH simulations. Nevertheless, the Sharpe ratio is substantially improved as a result of the large decrease in volatility (nearly 4.5 percentage-points on average after costs). On the re-weighted data, which keep the synchronous volatility property of the original data, the hedge takes out risk while exceeding the return of the index, on average and net of costs, leading to considerable increase in terms of efficiency (i.e., Sharpe ratio).

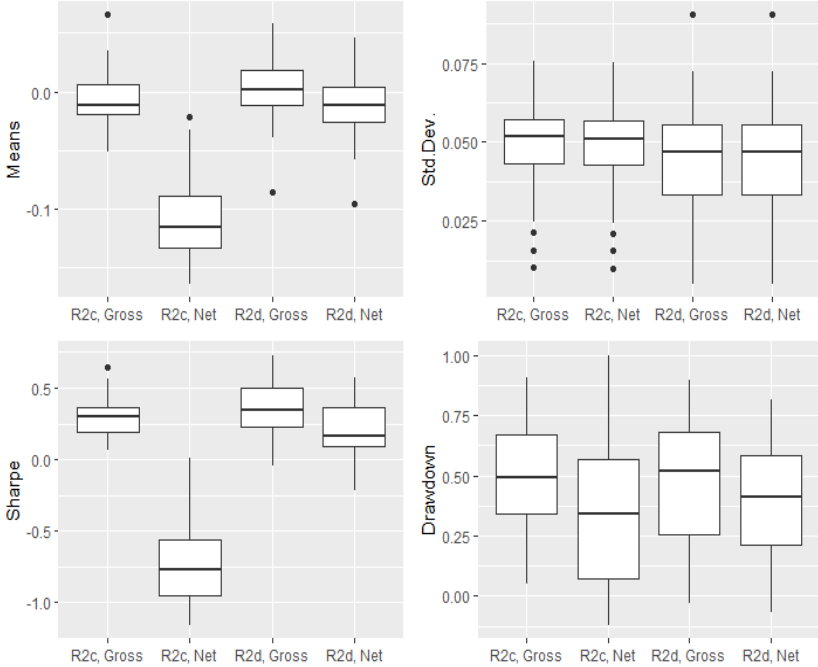


Figure 5.6: Performance differences obtained by applying hedging rule R_2 on 100 synthetically generated datasets using the GO-GARCH simulator.

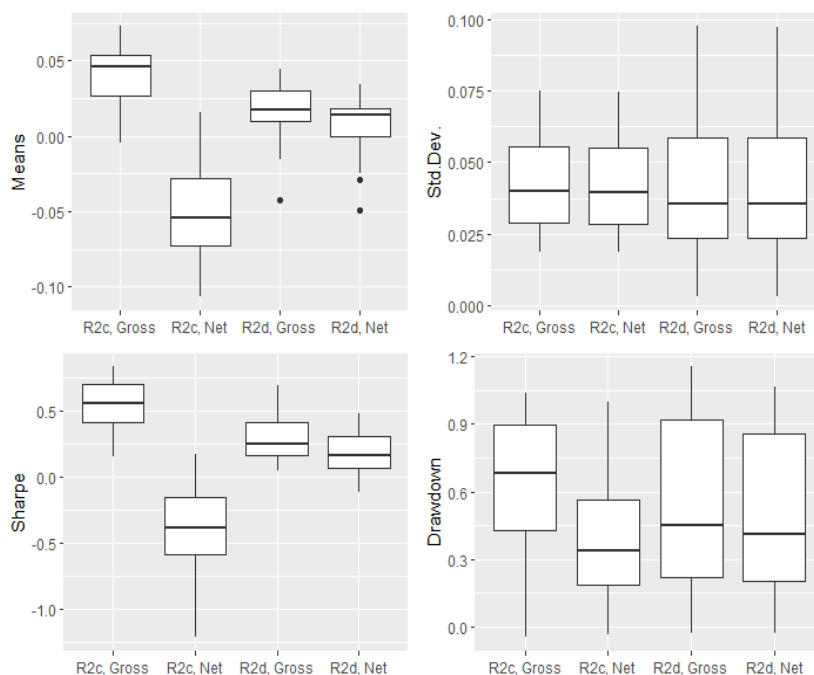


Figure 5.7: Performance differences obtained by applying hedging rule R_2 on 100 synthetically generated datasets using the method of reweighting historical dates.

5.5 Conclusion

We have explored the utilization of the Mahalanobis distance as a measure for gauging turbulence within global financial markets and its pertinence in the context of risk management. Initially introduced by [Kritzman and Li \(2010\)](#), the Mahalanobis-based indicator was explained as a tool for multivariate outlier detection and we have discussed methods for determining a threshold to distinguish periods of market turbulence from calmer times.

We have examined the empirically observed property of the turbulence measure to exhibit temporal clusters and to generally coincide with negative market developments when experiencing spikes. We were able to trace these properties back to two well-known stylized facts of asset returns, volatility clustering and asymmetric volatility-return dependence, which are expected to persist in the future, thus providing a theoretical justification for a Mahalanobis-based hedging approach. However, through the help of a simulation study, we found that these two properties alone were insufficient to support a Mahalanobis-based hedging strategy when the objective is not only to reduce risk (in terms of volatility and drawdowns) but also to maintain competitive returns. We identified a third criterion,

termed synchronous volatility, as crucial.

As the main contribution of this study, we introduced a method to distinguish turbulence into positive and negative market developments, labeled as good and bad, utilizing a basic generative model known as quadratic discriminant analysis. This approach calculates the difference between two Mahalanobis distances under different market conditions, marking another notable statistical application of the Mahalanobis distance, besides the detection of outliers, namely cluster analysis. The adaptation of a pattern recognition technique, initially designed for identifying economic cycles, facilitated the labeling of bear and bull markets, a prerequisite for learning class-specific parameters.

Various turbulence-based hedging rules were formulated and subjected to empirical examination. The assessment included both a statistical framework, utilizing classification metrics, and an economic framework, conducted through backtesting and the analysis of multiple performance measures. It became apparent that statistical significance does not always translate seamlessly into commensurate economic reward, a phenomenon attributed to transaction costs. In this context, a data-driven smoothing approach, based on an orthogonal stochastic volatility model, was proposed to reduce trading activity while achieving a balanced trade-off between errors of type I (missed signals) and type II (false alarms).

References

- Black, F. (1976). Studies of stock price volatility changes. pages 177 – 181.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307 – 327.
- Bose, S., Pal, A., SahaRay, R., and Nayak, J. (2015). Generalized quadratic discriminant analysis. *Pattern recognition*, 48:2676 – 2684.
- Bry, G. and Boschan, C. (1971). Cyclical analysis of time series: Selected procedures and computer programs.
- Cheung, Y.-W. and Ng, L. K. (1992). Stock price dynamics and firm size: An empirical investigation. *The Journal of Finance*, 47:1985 – 1997.
- Chow, G., Jacquier, E., Kritzman, M., and Lowry, K. (1999). Optimal portfolios in good times and bad. *Financial Analysts Journal*, 55:65 – 73.
- Christie, A. (1982). The stochastic behavior of common stock variances: Value, leverage, and interest rate effects. *Journal of Financial Economics*, 10:407 – 432.
- Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223 – 236.
- Duffee, G. (1995). Stock returns and volatility: A firm level analysis. *Journal of Financial Economics*, 37:399 – 420.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50:987 – 1007.
- Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48:1779 – 1801.
- Hamilton, W. P. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357 – 384.

- Hardin, J. and Rocke, D. M. (2005). The distribution of robust distances. *Journal of Computational and Graphical Statistics*, 14:928 – 946.
- Harding, D. and Pagan, A. (2002). Dissecting the cycle: a methodological investigation. *Journal of Monetary Economics*, 49:365 – 381.
- Harvey, A. C. and Shepard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business Economic Statistics*, 14:429 – 434.
- Kole, E. and van Dijk, D. (2017). How to identify and forecast bull and bear markets. *Journal of Applied Econometrics*, 32:120 – 139.
- Kritzman, M. and Li, Y. (2010). Skulls, financial turbulence, and risk management. *Financial Analysts Journal*, 66:30 – 41.
- Ledoit, O. and Wolf, M. (2022). Quadratic shrinkage for large covariance matrices. *Bernoulli*, 28:1519 – 1547.
- Lunde, A. and Timmermann, A. (2004). Duration dependence in stock prices: An analysis of bull and bear markets. *Journal of Business and Economic Statistics*, 22:253 – 273.
- Maheu, J. and McCurdy, T. (2000). Identifying bull and bear markets in stock returns. *Journal of Business and Economic Statistics*, 18:100 – 112.
- Meucci, A., Santangelo, A., and Deguest, R. (2013). Measuring portfolio diversification based on optimized uncorrelated factors. <http://ssrn.com/abstract=2276632>.
- Olive, D. J. (2004). A resistant estimator of multivariate location and dispersion. *Computational Statistics Data Analysis*, 46:99 – 101.
- Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140:425 – 449.
- Pagan, A. R. and Sossounov, K. A. (2003). A simple framework for analysing bull and bear markets. *Journal of applied econometrics*, 18:23 – 46.
- Paoletta, M. S. and Polak, P. (2015). Comfort: A common market factor non-gaussian returns model. *Journal of Econometrics*, 187:593 – 605.

- Taylor, S. J. (1982). Financial returns modeled by the product of two stochastic processes:
A study of daily sugar prices 1961 75. pages 203 – 226.
- van der Weide, R. (2002). Go-garch: A multivariate generalized orthogonal garch model.
Journal of Applied Econometrics, 17:549 – 564.

Appendix

5A Drawdown Statistics

Strategy	Gross				Net			
	From	Trough	To	Depth	From	Trough	To	Depth
MSCI World Index	2007-07-15	2009-03-09	2013-04-11	-0.538	2007-07-15	2009-03-09	2013-04-11	-0.538
	2002-03-20	2003-03-12	2004-12-22	-0.344	2002-03-20	2003-03-12	2004-12-22	-0.344
	2020-02-20	2020-03-23	2020-09-02	-0.331	2020-02-20	2020-03-23	2020-09-02	-0.331
	2022-01-05	2022-10-12	-	-0.228	2022-01-05	2022-10-12	-	-0.228
	2015-05-22	2016-02-11	2016-12-08	-0.174	2015-05-22	2016-02-11	2016-12-08	-0.174
$R_{1,a}$	2007-07-16	2009-03-02	2011-02-01	-0.447	2007-06-05	2009-03-02	2013-12-26	-0.525
	2002-01-07	2003-03-12	2004-01-16	-0.278	2002-01-07	2003-03-12	2004-12-22	-0.325
	2011-02-21	2011-08-19	2012-02-20	-0.180	2015-05-22	2016-02-11	2017-02-21	-0.163
	2015-05-22	2015-09-29	2016-12-08	-0.131	2021-11-17	2022-05-12	-	-0.161
	2021-11-17	2022-03-08	2022-06-24	-0.121	2018-01-29	2018-12-20	2019-04-05	-0.117
$R_{1,b}$	2002-03-20	2003-03-12	2004-11-12	-0.340	2002-03-20	2003-03-12	2004-11-15	-0.344
	2007-07-16	2009-04-28	2011-01-12	-0.319	2007-07-16	2009-04-28	2011-02-14	-0.336
	2011-02-21	2011-11-25	2013-03-05	-0.254	2011-02-21	2011-11-25	2013-03-08	-0.263
	2015-05-22	2016-02-11	2017-02-21	-0.207	2015-05-22	2016-02-11	2017-03-01	-0.212
	2018-01-29	2019-01-22	2019-11-27	-0.133	2018-01-29	2019-01-22	2019-12-27	-0.150
$R_{1,c}$	2021-11-17	2022-11-02	-	-0.168	2014-09-05	2023-03-15	-	-0.281
	2014-09-22	2016-02-03	2017-09-11	-0.148	2001-11-27	2003-03-31	2005-02-04	-0.230
	2002-01-07	2002-09-18	2003-09-02	-0.135	2010-04-16	2010-07-05	2013-05-17	-0.134
	2010-04-16	2010-07-05	2011-01-11	-0.118	2006-04-21	2006-07-18	2006-12-15	-0.122
	2018-09-24	2019-01-03	2019-04-12	-0.115	2013-05-22	2013-08-30	2014-08-19	-0.120
$R_{1,d}$	2014-09-22	2016-11-04	2018-01-05	-0.215	2014-09-22	2016-11-04	2018-09-21	-0.248
	2002-03-20	2003-01-27	2003-09-08	-0.157	2002-03-20	2003-01-27	2003-10-16	-0.163
	2020-02-20	2020-12-14	-	-0.150	2020-02-20	2020-12-14	-	-0.158
	2018-09-24	2018-12-25	2019-10-30	-0.148	2018-09-24	2018-12-25	2019-11-04	-0.155
	2013-05-22	2013-08-30	2013-12-23	-0.098	2007-07-16	2010-02-25	2010-10-13	-0.114
$R_{1,VIX}$	2007-07-16	2010-08-31	2014-08-26	-0.359	2007-07-16	2011-11-02	2017-09-12	-0.408
	2015-05-22	2016-02-24	2017-05-08	-0.201	2021-11-17	2022-12-28	-	-0.216
	2021-11-17	2022-12-28	-	-0.195	2002-01-07	2003-04-16	2003-12-29	-0.197
	2002-01-07	2003-01-16	2003-12-22	-0.188	2018-09-24	2018-12-20	2019-03-18	-0.116
	2018-09-24	2018-12-19	2019-03-15	-0.114	2006-05-10	2006-06-13	2006-10-05	-0.104

Table 5A.5: The five largest drawdowns per strategy related to hedging rule R_1 .

Strategy	Gross				Net			
	From	Trough	To	Depth	From	Trough	To	Depth
$R_{2,c}$	2008-05-20	2009-03-09	2010-02-18	-0.315	2002-01-07	2009-03-09	-	-0.496
	2020-02-13	2020-03-23	2020-11-09	-0.214	2001-12-07	2001-12-14	2001-12-27	-0.019
	2002-01-07	2003-03-11	2003-06-04	-0.149	2001-11-27	2001-12-03	2001-12-05	-0.018
	2007-07-16	2008-01-22	2008-04-28	-0.119	2001-12-31	2002-01-01	2002-01-03	-0.008
	2015-12-24	2016-02-11	2016-03-30	-0.087	-	-	-	-
$R_{2,d}$	2007-06-05	2009-02-10	2009-07-23	-0.197	2007-06-05	2009-02-11	2009-09-10	-0.253
	2021-11-17	2023-01-05	-	-0.131	2021-11-17	2023-01-05	-	-0.174
	2014-12-08	2016-02-08	2016-12-09	-0.121	2014-12-08	2016-02-08	2017-06-02	-0.165
	2002-03-20	2002-09-13	2003-05-02	-0.101	2002-03-20	2003-02-20	2003-05-30	-0.118
	2011-05-03	2011-12-19	2012-02-17	-0.094	2011-05-03	2011-12-19	2012-03-15	-0.117
$R_{2,QDA}$	2011-05-03	2011-11-25	2013-04-29	-0.230	2007-07-16	2011-11-25	2018-01-23	-0.322
	2021-11-17	2022-10-20	-	-0.216	2021-11-17	2022-11-09	-	-0.289
	2007-07-16	2009-07-10	2010-01-04	-0.173	2001-12-07	2004-08-12	2006-10-23	-0.221
	2015-04-28	2016-02-29	2017-02-10	-0.167	2018-09-24	2019-01-03	2019-12-13	-0.148
	2018-09-24	2018-12-25	2019-04-15	-0.129	2020-02-20	2020-04-13	2020-08-24	-0.088

Table 5A.6: The five largest drawdowns per strategy related to hedging rule R_2 .

Chapter 6

Conclusion

Bears and bulls are deeply ingrained animal symbols in financial culture used to describe market sentiment and trends. Meanwhile, monkeys have become emblematic of chance and randomness within both financial and statistical contexts. In finance, the metaphorical image of monkeys throwing darts reflects the stochastic process of sampling – a central theme of this dissertation.

The primary contribution that this work offered to the finance community lies in techniques for sampling from high-dimensional and truncated distributions, thereby enabling the analysis of systematic effects between portfolio characteristics and performance under real-world scenarios.

Chapters 2 and 3 introduced geometric random walks, a subset of Markov chain Monte Carlo methods tailored for sampling from distributions confined within geometric bodies. These methods were employed for sampling portfolios with targeted attributes, such as maintaining a predetermined variance or adhering to strict investment criteria like upper limits on individual investments or groups of securities, risk thresholds, or factor exposures.

Chapter 2 introduced two novel walks crafted to sample portfolios with a predetermined variance. From an algorithmic point of view, the sampling problem is very challenging as it involves the computation of volume of and sampling from non-convex and disconnected spherical patches arising by the intersection of a non-standard simplex with a sphere. Aside from solving the geometrically hard problem, the contribution with introduced tools lies in their application to explore the low-volatility anomaly – the empirical observation that assets with lower risk levels often produce higher returns, diverging from predictions made by classical economic theory. Through empirical analysis across various markets and

sub-markets, the chapter confirmed the presence of the low-volatility anomaly while also questioning conventional sorting-based methodologies.

Chapter 3 delved deeply into geometric random walks, offering a comprehensive survey of existing methods and presenting novel complexity results. These techniques were applied to analyze factor anomalies within a constrained setup, providing insights into the relationship between factor tilts and performance. Additionally, the chapter scrutinized the common practice of using random portfolios for performance analysis to test investment skill, exposing flaws in the techniques used to construct these portfolios through geometric reasoning.

Chapter 4 offered a geometric perspective on nonparametric bootstrap methods (identified in the previous chapter as a common approaches for generating random portfolios). It presented a unified representation of classical and Bayesian frameworks using random weights, establishing a link between resampling plans and the distribution of weights. A direct computational method for the exact Bayesian bootstrap distribution was introduced, which eliminates randomization error. Additionally, it was demonstrated that the exact moments of bootstrap distributions depend on the L_p norm of the random weights. Moreover, methods to integrate prior information into bootstrap procedures were proposed through a parametrization of the weight distribution. Furthermore, applications were discussed where it is common to employ bootstrap standard errors, which by the geometric perspective, can be determined without the need for resampling.

Chapter 5 explored the Mahalanobis distance as a measure of financial turbulence and its application in risk management. It linked empirical observations of turbulence coinciding with negative market events to known asset return patterns. A set of jointly sufficient conditions on the data generating process of asset returns were identified to validate its applicability for dynamic hedging and methods for setting thresholds to distinguish turbulent from calm periods were discussed. Additionally, the chapter introduced a methodology related to quadratic discriminant analysis to disentangle turbulent episodes into good and bad, in the sense of positive and negative market developments, by adapting a pattern recognition technique to identify up- and downward trending markets (bear and bull states) initially devised to date economic cycles. Various turbulence-based hedging rules were discussed and evaluated empirically, emphasizing the need to balance statistical significance with economic considerations, notably transaction costs. A data-driven

smoothing approach based on an orthogonal stochastic volatility model was proposed to improve the trade-off between signal accuracy and trading activity.