Randomized geometric tools for anomaly detection in stock markets

Cyril Bachelard
Faculty of Business and Economics, Dept of Operations
Univ of Lausanne, Switzerland

Apostolos Chalkis
Quantagonia, and GeomScale Org

Vissarion Fisikopoulos
Department of Informatics & Telecommunications
National & Kapodistrian University of Athens, and GeomScale Org

Elias Tsigaridas
Inria Paris and IMJ-PRG
Sorbonne Université
GeomScale Org

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.

1 INTRODUCTION

We consider two fundamental problems from Computational Geometry and Financial Economics through the lens of efficient sampling from high dimensional (truncated) distributions. 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 multiple connected components make the problem very challenging from an algorithmic and implementation point of view. Both sampling and volume estimation have a wide range of applications both in (computational) statistics, e.g., Friel and Wyse (2012); Fong and Holmes (2020); Gelman and Meng (1998) but also across the whole range of science and engineering, e.g., systems biology Chalkis et al. (2021) and robotics Althoff et al. (2011), to mention few of them.

The geometric viewpoint 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 (of independent interest) geometrically hard problem emanates from questions on the feasible space of investable portfolios obeying certain regulatory and risk related constraints. The fundamental (and long-standing) economic question we address is on 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?

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 wrt the distribution of sampling-based statistics, we conclude that the classical sorting-based portfolios form bad proxies and one should

Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS) 2023, Valencia, Spain. PMLR: Volume 206.

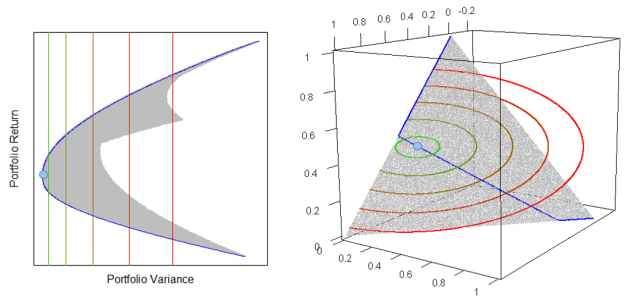


Figure 1: Left: Feasible set and iso-volatility lines in risk-return space. The grey area represents the set of feasible portfolios. Iso-volatility level sets form vertical lines and are color-coded by increasing variance from green to red. The blue curve highlights the set of Pareto-efficient portfolios in the trade-off between risk and return. The part above the minimum variance portfolio (light blue dot) is called the efficient frontier. Right (our settings): 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. 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.

be careful to base inference on them.

We introduce a geometric modeling of the financial problem and develop efficient randomized geometric tools to compute with. 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, geodesically, non-convex, and non-connected 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 3 for details).

1.1 Geometry, methodology, challenges

It is common in finance and economic literature to represent financial terms geometrically Markowitz (1992). In our setting, we estimate the covariance matrix from historical data and set several volatility levels that define a sequence of concentric ellipsoids intersecting the standard simplex that represents the set of all feasible portfolios (see Figure 1). By sampling independently and uniformly from each intersection, we obtain sets of volatility-constrained portfolios. Then, 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.

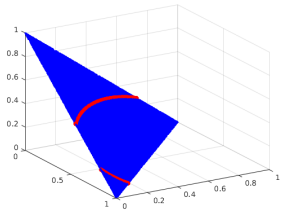
From a geometric point of view, the intersection between the canonical simplex and the boundary of an ellipsoid in $\mathbb{R}^d$ is a $(d-1)$-dimensional (geodesically) non-convex and non-connected body (Figure 3). Sampling from such a set is a very challenging problem. An additional challenge, in our 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 to sample from such a set relies on Markov Chain Monte Carlo (MCMC) sampling while uses efficient practical Multiphase Monte Carlo (MMC) volume approximation to sample from disconnected regions from the correct distribution.

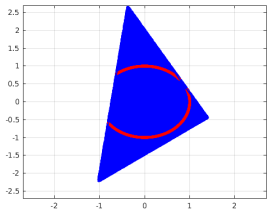
In previous approaches, quantile or MV portfolios corresponds to points belonging to the boundary of the set of constrained volatility portfolios. In particular, they lie on the intersection between several facets of the simplex and the boundary of the ellipsoid. Hence, the volume of the region around those portfolios is typically small, i.e., it is unlikely to be visited by the random walk. Thus, it is likely that the analysis of the volatility puzzle with those portfolios can done using outliers in our statistical framework.

High-dimensional sampling from multivariate distributions with MCMC algorithms is a fundamental problem with many applications in science and engineering 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 exactly in general dimension Dyer and Frieze (1988).

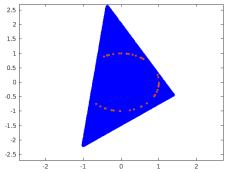
MCMC sampling algorithms have made remarkable progress in solving efficiently the problems of sampling and volume approximation of full-dimensional convex bodies in $\mathbb{R}^d$ while enjoying great theoretical guarantees Chen et al. (2018); Lee and Vempala (2018); Mangoubi and Vishnoi (2019). Sampling 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



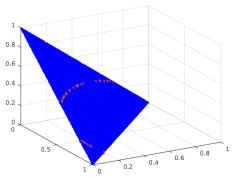
(a) Intersection of the ellipsoid with the unit simplex



(b) Transformed intersection



(c) Uniform points on the transformed intersection



(d) Uniformly distributed points on the ellipsoid

Figure 2: Illustration of the various phases of the algorithm: Initially, from the data we build the covariance matrix; this corresponds to an ellipsoid. Then we consider the intersection of the unit simplex with the ellipsoid; this intersection is the red curve in Fig. 2a. Then, we transform the ellipsoid to the unit sphere (one dimension lower); red curve in Fig. 2b. Next, we sample points from the intersection of the unit sphere with the (transformed) simplex (orange points in Fig. 2c). Finally, we apply the inverse transformation to map back the sampled points to the original ellipsoid (orange points in Fig. 2d). The latter correspond to (samples of) volatility constrained portfolios.

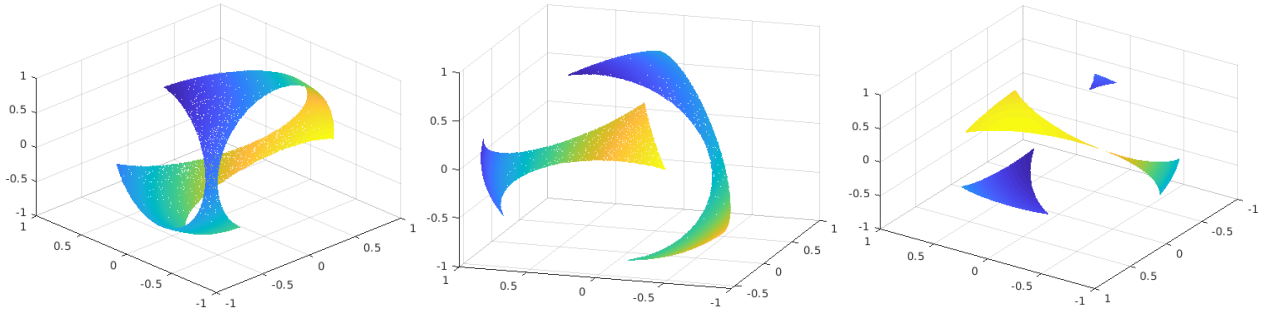


Figure 3: Three examples in $\mathbb{R}^3$ of the body $K = \mathcal{S}_2 \cap \Delta$ (the unit sphere intersected by the interior of a simplex) that we sample from. In general, K is a non-connected and geodesically non-convex body in $\mathbb{R}^d$.

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 hyperpshere 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). The most recent addition to this list is Chevallier et al. (2022) that is numerically robust and exploits a Piecewise Deterministic Markov Process.

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).

1.2 (brief) 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 Haugen and Heins (1975) were the first to recognize that risk does not generate a special reward following the early warning signs 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, e.g., Banz (1981); Rosenberg et al. (1985); Jegadeesh and Titman (1993); Asness et al. (2019). Even more, the outperformance of low-volatility stocks compared to high-volatility stocks has been shown to be robust among different markets, industries and sub-periods 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).

We refer the interested reader to the supplementary material where we present a more detailed overview of the financial background that help us to obtain accurate description of the various mathematical, algorithmic, and computational problems that we need to address.

2 GEOMETRIC MODELING AND ALGORITHMS

We introduce the geometric modeling of long-only volatility-constrained portfolios, we develop two sampling algorithms, and an MMC scheme for volume approximation. We denote a full dimensional (convex) body with a capital letter and if applicable with an index we denote its dimension; e.g., the unit ball in $\mathbb{R}^d$ is B^d. For lower dimensional, possibly (geodesically) non-convex, bodies we use a calligraphic letter; e.g., $\mathcal{S}_{d-1}$ is the $(d-1)$ dimensional sphere in $\mathbb{R}^d$. Usually, they subsets of the boundary of a full-dimensional convex body.

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. Thus, in our setting, the set of long-only portfolios, is the canonical d-dimensional simplex $\Delta_d := \{x \in \mathbb{R}^{d+1} \mid x_i \geq 0,\ \sum_{i=1}^{d+1} x_i = 1\} \subset \mathbb{R}^{d+1}$, 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 $x \in \Delta_d$, the return of x is $f_{ret}(x, R) = R^T x$. Similarly, for a positive definite covariance matrix $\Sigma \in \mathbb{R}^{(d+1)\times(d+1)}$ of the distribution of the assets' returns, let the portfolio volatility be $f_{vol}(x, \Sigma) = x^T \Sigma\, x$. Thus, to model portfolios' volatility we employ ellipsoids intersecting Δ_d, i.e.,

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

corresponds to the portfolios with volatility c. Notice that portfolios, that is the points, that belong to the (centered at the origin) ellipsoid, $\mathcal{E}_d^c := \{x \in \mathbb{R}^{d+1} \mid x^T \Sigma x = c,\ c \in \mathbb{R}_+\} \subseteq \mathbb{R}^{d+1}$, 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 $\mathcal{K} := \mathcal{S}_{d-1} \cap \Delta$. We can use the inverse transformations to obtain uniformly distributed portfolios with volatility c as the transformation is isometric, e.g., Chalkis et al. (2021).

In general, $\mathcal{K}$ is a set of geodesically non-convex[1] spherical surface patches (see Figure 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 vol-

[1] A subset C of a surface is 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 represents the shortest path or arc between two points on a surface.

ume. 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.

2.2 Sampling from a connected component of $\mathcal{K}$

We introduce two geometric random walks, namely the Great Cycle Walk and the Reflective Great Cycle Walk, 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 the $\mathcal{S}_{d-1}$ are the intersection of the $\mathcal{S}_{d-1}$ with 2-dimensional hyperplanes that pass through the origin in the $\mathbb{R}^d$. These great circles are the geodesics of the $\mathcal{S}_{d-1}$ Byrne and Girolami (2013).

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 6) that the (unique) stationary distribution of this algorithm is π and moreover, that GCW converges to π from any starting point in $\mathcal{K}$.

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

$$\ell(\theta) := \{p\cos\theta + v\sin\theta,\ \theta \in [-\pi, \pi]\}. \tag{2}$$

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

$$v = \frac{(I_d - pp^T)u}{\|(I_d - pp^T)u\|_2}, \tag{3}$$

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,

$$a_j^T \ell(\theta) = b_j \iff a_j^T x_i \cos\theta + a_j^T v_i \sin\theta = b_i, \tag{4}$$

with $\theta \in [-\pi, \pi]$, $j \in [d+1]$, 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

$$\theta_j^-,\ \theta_j^+ = \pm\cos^{-1}(z_j) + \tan^{-1}\left(\frac{a_j^T v_i}{a_j^T p}\right). \tag{5}$$

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

ALGORITHM 1: GCW(Δ, p, π)

Input : Simplex Δ; point p; PDF 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^+) the values s.t.
$\ell_p(\theta) := \{p\cos\theta + v\sin\theta,\ \theta \in [\theta^-, \theta^+]\}$ is the part of $\ell(\theta) \cap \Delta$ that contains p;
Pick $\tilde{\theta}$ from $\ell_p(\theta)$ according to π_ℓ;
return $p\cos\tilde{\theta} + v\sin\tilde{\theta}$;

Theorem 1. *The stationary distribution of Alg. 2, CGW(Δ, p, π), where the starting point p belongs to K, is π for any starting point p in 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. Thus, the cost per step is $O(d^2)$ operations.

GCW can also be used to sample uniformly from a component of $\mathcal{K}$. However, we also propose Reflective Great Cycle Walk that exploits boundary reflections and has superior practical performance compared to GCW for uniform sampling. We provide details in the supplementary material, while a pictorial overview of the algorithm is in Fig. 7.

Theorem 2. *The Reflective Great Cycle Walk has a unique stationary distribution, which is the uniform distribution; it converges from any starting point.*

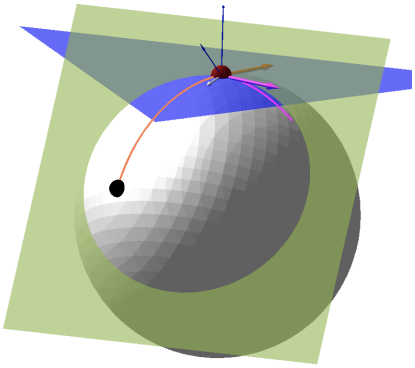


Figure 4: An illustration of the reflection of the ReGCW. The orange trajectory, starting from the black point, hits the boundary of the component defined by the intersection of the sphere with the green hyperplane (red point). The blue hyperplane is the tangent space at the intersection point. The resulted trajectory is in magenta.

2.3 Practical volume approximation

At a high level, our method is based on that of Lovász and Vempala (2006); Cousins and Vempala (2015) and on the practical variant 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 appeared in previous work Lovász and Vempala (2006); Cousins and Vempala (2015). In particular, given a connected component $\mathcal{K}_i,\ i \in [M]$, for any sequence of k functions $f_j : \mathcal{S}_{d-1} \to \mathbb{R}_+,\ j \in [k]$ we consider the following representation of $\text{vol}(\mathcal{K}_i)$, for $i \in [M]$,

$$\text{vol}(\mathcal{K}_i) = \int_{\mathcal{K}_i} f_k dx \frac{\int_{\mathcal{K}_i} f_{k-1} dx}{\int_{\mathcal{K}_i} f_k dx} \cdots \frac{\int_{\mathcal{K}_i} dx}{\int_{\mathcal{K}_i} f_1 dx} \tag{6}$$
$$= \Big(\frac{1}{\int_{\mathcal{K}_i} f_k dx} \frac{\int_{\mathcal{K}_i} f_k dx}{\int_{\mathcal{K}_i} f_{k-1} dx} \cdots \frac{\int_{\mathcal{K}_i} f_1 dx}{\int_{\mathcal{K}_i} dx}\Big)^{-1},$$

while we prefer the right hand of (6) for reasons we explain in the sequel. We set each f_j to be proportional to the von-Mises Fischer (vMF) distribution, i.e., $f_j(x) = e^{a_j(\mu^T x)},\ x, \mu \in \mathcal{S}_{d-1},\ a_j > 0$, 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).

In our case, we cannot guarantee that the mass of the distribution corresponding to α_k is almost inside $\mathcal{K}_i$; thus, we start from the uniform distribution and then decrease the variance until we reach a distribution with sufficiently small variance. To terminate, at 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 (6) 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,let $Y_j = \frac{\int_{\mathcal{K}_i} f_j dx}{\int_{\mathcal{K}_i} f_{j-1} dx} = \int_{\mathcal{K}_i} \frac{f_j}{f_{j-1}} \frac{f_{j-1}}{\int_{\mathcal{K}_i} f_{j-1} dx}$. 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 $R_j = \frac{1}{N}\sum_{l=1}^{N} \frac{f_j(x_l)}{f_{j-1}(x_l)}$, while $E[Y_j] = \lim_{N\to\infty} R_j$. Using Chebyshev's inequality, when $Var[Y_j]/E[Y_j]^2 \leq 1$ we guarantee that $N = \widetilde{O}(1)$ points suffice to approximate Y_j 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 $\alpha_j = \alpha_{j-1}(1 + \frac{1}{d})$, then $Var[Y_j]/E[Y_j]^2 \leq 1$. To the best of our knowledge it is unclear if the 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 $\alpha_j = \alpha_{j-1}(1 + \frac{1}{d})^r,\ r \in \mathbb{R}_+$. 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 $\alpha_j = \alpha_{j-1}(1 + \frac{1}{d})^{r_{\max}}$ is larger than 1. Following Cousins and Vempala (2016) we set N' for 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, $\int_{\mathcal{K}_i} f_k dx = (1-\epsilon_0)\int_{\mathcal{S}_{d-1}} f_k dx$ 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 $\nu = O(\log(\frac{1}{1-\zeta}) \big/ \epsilon_0^2)$ 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 $1-\zeta$. In practice, $\epsilon_0 = 0.05$.

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, $(\max(W) - \min(W))/\min(W) \leq \epsilon_k/2$, 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, as in Cousins and Vempala (2016), we set it $4d^2 + 500$.

Sampling from a segment. To sample from the latter univariate distribution we use Metropolis-Hastings algorithm Chib and Greenberg (1995) (see section 3).

3 IMPLEMENTATION, EXPERIMENTS, RESULTS

We present the implementation of our algorithms and the tuning of various parameters. We provide a complete open-source software framework [2] to address low-volatility detection in stock markets with hundreds of assets. We will make our code publicly available. 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 below in this section). Last but not least, to obtain an efficient implementation for our methods we introduce an efficient parameterization. For reasons of space, we present part of our results, experiments, and analysis in this section. We refer the supplementary material for further details.

Construction of volatility-constrained random portfolios and backtesting framework. To perform experiments we construct volatility-constrained random portfolios. In particular, we sample 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 estimator is the non-linear shrinkage estimator 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, 1 000 portfolios are sampled[3]. The investments are held over the following three months until the process is repeated. Ultimately, we arrive at a total of 5 000 backtested portfolio price paths capturing the profits and losses endured over a period spanning from March 2002 to Dec 2021 by randomly concatenating time series within each volatility cluster at the 80 rebalancing dates. The entire procedure is repeated several times 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 MSCI[4] building on the Global Industry Classification Standards

To argue on our choice of non-linear shrinkage for the estimation of the covariance matrix was made particularly because of the robustness of the estimator (regulates the eigenvalues). It is a state-of-the-art robust estimator and particularly suitable to rather high-dimensional portfolio applications (see e.g., Ledoit and Wolf (2012, 2017)). In addition our simulations using the (i) sample covariance estimator and (ii) exponentially weighted covariance matrices (which overweighs the recent history) result the same qualitative conclusions. Also, our results are robust vis-à-vis the number of historical data points for the estimation of the covariances which we varied from one to five years for robustness checks.

Our conclusion that the sorting-based portfolios form bad proxies which could lead to misleading results stems from empirical observations (cross-checked across different sub-universes and sub-periods) that the risk-return statistics of

[2] `https://zenodo.org/record/7198256`

[3] Increasing the number of samples is not a bottleneck (linear complexity). However, we found no economic value by adding more samples as qualitatively, our results do not change.

[4] Defensive 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`

the sorting-based approach typically deviate substantially from the average sampling-based portfolio characteristics (this actually has a geometric explanation) and that the variation in the return among the five volatility clusters is much more erratic for the sorting-based method than for the mode of the sampling-based portfolio return distributions. Hence, depending on the (sub-) universe and time period one might over- or underestimate the low-vola anomaly whereas inference from the average sampled portfolio performance is more robust. The sorting-based approach (which ignores correlations) was used to conclude that stocks with higher volatility underperform stocks with low volatility and there is nothing wrong with that. However, on a portfolio level, the same conclusion (i.e., portfolios with low volatility outperform portfolios with high volatility) should not be drawn from the sorting-based portfolio only as it represents just one particular point out of the many possible portfolios with the same (ex-ante) volatility.

Experimental results and analysis. 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 Fig. 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 strategy using the procedure described in above. The points per volatility group are overlaid by bivariate (non-parametric 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 Fig. 5, as well as the figures in the supplementary material reveal the following results.

First, we can confirm the presence of the low-volatility effect in both universes and corresponding sub-markets. Second, 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 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. 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 of `LogConcDEAD` package Cule et al. (2009) that fits on the data a log-concave density function which logarithm is a tenant function and its support is the convex hull of the cluster. The fitted density gives us 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 vecotors of the sorting based quintile portfolios. In all 10 (sub-) universes, we take the volume of the small rectangle to be 1% of the volume of the support of the distribution. The plots in the supplementary material illustrate the PDF computed by `LogConcDEAD` package and the quintile portfolios for each volatility level that corresponds to the plots in Figure 5. We also report the measures of the rectangles in Table 1. 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 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 4 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 4 cases, the probability of the area around the quintile portfolio statistics is more than 10 times smaller than that of the average or the mode and in one case it is 4 times smaller. Consequently, according to the log-concave model in Cule et al. (2010), 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.

We also analyze 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, and 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

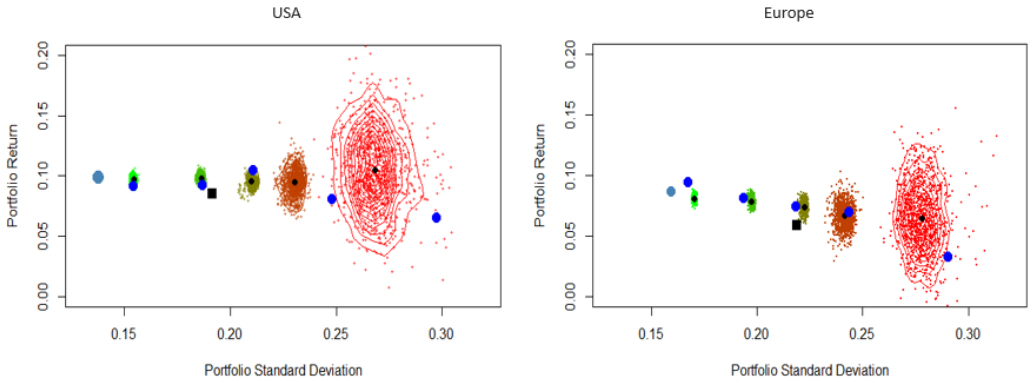


Figure 5: Risk-return profiles of the five clusters of backtested sampling-based portfolios for the U.S. (left) and European (right) market where clusters are color-coded by increasing variance from green to red. The black dots represent the cluster averages whereas blue dots depict the performance of 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.

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 1: For each volatility cluster in Figure 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 we 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.

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 and all significant test results come from the subset with positive Sharpe ratio differences.

We conclude that, while test statistics using 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.

Acknowledgements The authors would like to thank the anonymous reviewers for their constructive comments. ET is partially supported by ANR JCJC GALOP (ANR-17-CE40-0009).

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.

Althoff, M., Le Guernic, C., and Krogh, B. H. (2011). Reachable set computation for uncertain time-varying linear systems. In *Proc. 14th international conference on Hybrid systems: computation and control*, pages 93–102.

Asness, C. S., Frazzini, A., and Pedersen, L. H. (2019). Quality minus junk. *Review of Accounting Studies*, 24:34 – 112.

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.

Chevallier, A., Cazals, F., and Fearnhead, P. (2022). Efficient computation of the the volume of a polytope in high-dimensions using piecewise deterministic markov processes. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 10146–10160. PMLR.

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). Minimumvariance 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). Logconcdead: 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). Hyperspherical 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.

Fong, E. and Holmes, C. C. (2020). On the marginal likelihood and cross-validation. *Biometrika*, 107(2):489–496.

Friel, N. and Wyse, J. (2012). Estimating the evidence–a review. *Statistica Neerlandica*, 66(3):288–308.

Gelman, A. and Meng, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical science*, pages 163–185.

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. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, pages 1024–1060.

Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *The Review of Financial Studies*, 30(12):4349–4388.

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.

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.

A ADDITIONAL 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 Haugen and Heins (1975) were the first to recognize that risk does not generate a special reward following the early warning signs 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 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).

The typical approach pursued by the studies on the low-volatility anomaly is to sort stocks according to their historical volatility and to form portfolios, either weighted equally or proportional to market capitalization, within quantiles 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.

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.

However, the ignorance of correlations between assets' returns in the formation of volatility-ranked portfolios poses a drawback in the existing literature analyzing the volatility puzzle Haugen and Baker (1991); Clarke et al. (2006, 2011). Instead of clustering stocks according to volatility and representing the subgroups by a single portfolio, we propose an alternative. That is, pre-define certain volatility targets and to sample portfolios with exactly those ex-ante volatilities from the entire collection of investable firms. Using sampling, we consider correlation but we also overcome 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).

B MEMBERSHIP ORACLE FOR A CONNECTED COMPONENT

Lemma 3. *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. The Figure 6 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. □

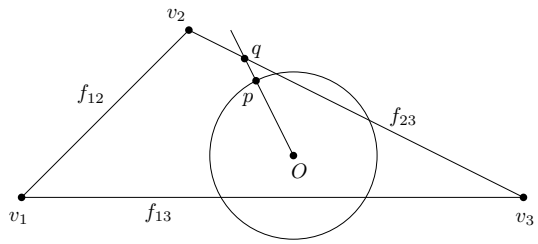


Figure 6: Illustration of how the membership oracle works. In this example, there are two connected subsets of $\mathcal{S}_1$. We would like to answer in which component the point p belongs. The half-line l_{op} intersects ∂P on q which lies in the facet f_{23}. The point q has visible contact with 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.

C SAMPLE 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

$$\phi(\theta) \propto \|\ell'(\theta)\|_2 = \|-p\sin\theta + v\cos\theta\,\|_2 = 1,\ \theta \in [\theta^-, \theta^+], \tag{7}$$

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 $[\theta^-, \theta^+]$.

D 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 it 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 is also a (part of a) great cycle of $\mathcal{S}_{d-1}$. ReGCW returns the next Markov point when it travels 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 detail our choices for τ and ρ.

ALGORITHM 2: GCW(Δ, p, π)

Input : Simplex Δ; point p; PDF 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^+) the values s.t.
$\ell_p(\theta) := \{p\cos\theta + v\sin\theta,\ \theta \in [\theta^-, \theta^+]\}$ is the part of $\ell(\theta) \cap \Delta$ that contains p;
Pick $\tilde{\theta}$ from $\ell_p(\theta)$ according to π_ℓ;
return $p\cos\tilde{\theta} + v\sin\tilde{\theta}$;

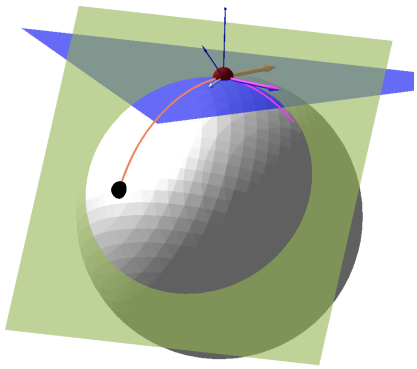


Figure 7: An illustration of the reflection of the ReGCW. The orange trajectory, starting from the black point, hits the boundary of the component defined by the intersection of the sphere with the green hyperplane (red point). The blue hyperplane is the tangent space at the intersection point. The resulted trajectory is in magenta.

Some details of the computations are in order. At each step of ReGCW, 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 has 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

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

ALGORITHM 3: 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 ;
return p

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 $a \in \mathbb{R}^d$ 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 the previous equation. Then, the reflection of $\ell(\theta)$ at q is,

$$\ell_r(\theta) = \{q \cos\theta + v_r \sin\theta,\ \theta \in [0, L - \tilde{\theta}]\}, \qquad (8)$$

where $v_r = v - 2(v^T \alpha')\alpha'$. We also update the length travelled so far, i.e., $L \leftarrow L - \tilde{\theta}$. When $L < \tilde{\theta}$ 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 7.

Theorem 4. *The Reflective Great Cycle Walk of Alg. (3) has a unique stationary distribution, which is the uniform distribution; it converges from any starting point.*

Proof. We build upon the results and the methodology in Gryazina and Polyak (2014). Let K_i a connected component of $K = \Delta \cap \mathcal{S}_{d-1}$. The 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,

$$\Pr[\delta q|p] \propto \Pr[\delta\theta] \Pr[\delta L], \qquad (9)$$

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. □

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

Parameters of ReGCW. To employ ReGCW (see section D), we have to efficiently select values for the parameter τ that controls the length of the trajectory in each step, for the maximum number of reflections per step ρ, and for the walk length of the random walk. We have experimentally found that setting the walk length equal to 1, is the

fastest choice so that the empirical distribution converges to the uniform distribution. 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 (5), in the main text, computing in those $20d$ steps of GCW. For the maximum number of reflections, we experimentally found that $\rho = 100d$ is violated in less than 0.1% of the total number of ReGCW steps in our experiments.

E PROOF OF THEOREM 1 IN THE MAIN TEXT

Theorem 6. *The stationary distribution of Alg. 2, CGW(Δ, p, π), where the starting point p belongs to K, is π for any starting point p in K.*

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

$$\int_A P(p, B)\pi(dp) = \int_B P(r, A)\pi(dr), \text{ for all } A, B \in \mathcal{D} \tag{10}$$

Then, the stationarity of π follows at once since we set $B = K$ in Equation (10) and we get,

$$\pi(A) = \int_K P(r, A)\pi(dr), \text{ for all } A \in \mathcal{D}, \tag{11}$$

which implies stationarity of π Vempala (2005).

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

$$\Phi(p, q) = \{\theta \in [-\pi, \pi] \mid p\cos\theta + q\sin\theta \in K\}. \tag{12}$$

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

$$f_{(p,q)}(\theta) = \begin{cases} \frac{f(p\cos\theta + q\sin\theta)}{\int_{\Phi(p,q)} f(p\cos t + q\sin t)\ dt} & \text{if } \theta \in \Phi(p, q), \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

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

$$\begin{aligned} P(p, A) &= \Pr[F^{-1}_{(p,Q)}(U) \in A] \\ &= \int_{\mathcal{S}_{d-1}\cap\mathcal{H}_p} \frac{1}{\text{vol}(\mathcal{S}_{d-1} \cap \mathcal{H}_p)} \Pr[F^{-1}_{(p,q)}(U) \in A] dq \\ &= \int_{\mathcal{S}_{d-1}\cap\mathcal{H}_p} \frac{1}{\text{vol}(\mathcal{S}_{d-1} \cap \mathcal{H}_p)} \int_{\Phi(p,q)} \mathbf{1}_A(p\cos\theta + q\sin\theta) f_{(p,q)}(\theta) d\theta\ dq, \end{aligned} \tag{14}$$

where $\mathbf{1}_A(\cdot)$ is the indicator function of the set A and $F^{-1}_{(p,q)}$ 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

$$g(p, r) = \frac{\text{vol}(W \cap \mathcal{S}_{d-1} \cap \mathcal{H}_p)}{\text{vol}(\mathcal{S}_{d-1} \cap \mathcal{H}_p) \int_{\Phi_{pr}} p\cos t + q\sin t\ dt}, \tag{15}$$

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 (10) 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. □

F EFFICIENT PARAMETERIZATION OF THE IMPLEMENTATION

To obtain 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,

$$\max\ r, \quad \text{subject to : } a_i^T x_c + r||a_i|| \le b_i,\ ||x_c|| \le 1 - r, \tag{16}$$

where Δ is defined as the intersection of the half-spaces $a_i^T x \le b_i,\ i = 1, \ldots, d + 1$. Then, any vertex v of Δ that also belongs to the connected component of graph G 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 Markov Chain Monte Carlo diagnostic, namely the potential scale reduction factor (PSRF) Gelman and Rubin (1992). In particular, we compute the PSRF for each univariate marginal of the sample that both GCW and ReGCW output. Following Gelman and Rubin (1992), and plenty of other works, a convergence is satisfying according to PSRF when all the marginals have PSRF smaller than 1.1.

mVF restricted a segment (GCW). To sample from mVF using GCW, in each step of the random walk, we have to

sample from the mVF constrained on a part of a great cycle. Thus, 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. We found that this value is an efficient choice; in our experiments the Markov point of GCW changes after the 95% of the total number of steps we performed in our experiments. Thus, the empirical probability of stacking at a certain point is quite small.

G 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 indices[5] 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 total[6] returns using Wednesday closing 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.

[5] see https://www.msci.com/our-solutions/indexes/developed-markets

[6] Returns, 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).