# When Is One Asset a Substitute for Another?

A decision-theoretic criterion for replacement in optimized portfolios*

Agustín Shehadi Candela
QUAFI Research
research@quafi.io

August 2026

**Abstract**

Suppose an investor holds an optimized portfolio and wishes to replace one of its constituents — because of a mandate, an access restriction, or a preference — with an asset that plays the same role. Which asset should be offered? The intuitive answer is the one whose inclusion leaves the covariance matrix least disturbed, measured by a matrix norm. We show this answer is wrong in three specific and correctable ways, the sharpest being that a matrix norm is *sign-blind*: it penalises a candidate that strictly improves the portfolio exactly as much as one that strictly damages it. We derive the correct criterion from the envelope theorem — the first-order change in the *value* of the optimization problem, evaluated at unchanged optimal weights — and show it costs no more to compute than the naive one. We then address a limitation the covariance formulation cannot see at all: a covariance of one-period returns is a single-horizon object, and two assets can co-move at high frequency while their cumulative returns diverge permanently. We propose the $h$-period tracking variance and its associated spread variance ratio as the horizon-aware criterion, prove that its boundedness in $h$ is equivalent to cointegration of the log prices with a unit cointegrating vector, and argue this continuous statistic is preferable to hypothesis testing when thousands of candidates are screened. We formalise the selection bias induced by minimising a noisy criterion over a large universe, showing it grows like $\sqrt{2\log M}$ — so restricting the candidate set helps only logarithmically, a fact that has been used to justify more than it can support. A simulation study confirms the first-order criterion is correct to $O(\varepsilon^2)$, that it fails badly at discrete substitutions near a binding constraint, and that it nonetheless ranks well enough for a two-stage procedure to recover the exact optimum.

**Keywords:** asset substitution; portfolio optimization; envelope theorem; cointegration; variance ratio; selection bias; index tracking.

**JEL:** C13, C58, G11.

*QUAFI Working Paper Series, No. 2026-06. Preliminary; circulated to inform and to invite comment. *Nothing herein is investment advice, and nothing herein is a recommendation to buy or sell any security.* The paper is methodological: it states a criterion and proves its properties, and makes no claim about any particular instrument. Any errors are mine.

# Contents

# 1 Introduction

A portfolio optimizer returns a basket. The investor looks at it and objects to one name. Perhaps they already hold it through another vehicle, perhaps their mandate excludes it, perhaps their broker does not offer it, perhaps they simply do not want it. They ask a reasonable question: *what else would do the same job?*

The question is easy to state and surprisingly easy to answer badly. This paper is about what makes an asset a good substitute for another *inside a portfolio*, which is a different and more structured question than what makes two assets similar in the abstract.

## 1.1 The intuitive answer, and why it fails

The intuition most practitioners reach for first is that the portfolio's risk structure is summarised by its covariance matrix $\Sigma$, so a good substitute is one that leaves $\Sigma$ least disturbed. Replacing asset $i$ by candidate $y$ produces a new matrix $\tilde{\Sigma}$, and one selects

$$\arg\min_{y} \|\Sigma - \tilde{\Sigma}\|_F.$$

This is appealing, and it has an exact and efficient solution (Proposition 1) — the perturbation touches only one row and column, so the norm collapses to a closed form computable for thousands of candidates with a single matrix product.

The criterion is nevertheless wrong, in three ways of increasing severity.

1. It weights every asset in the portfolio equally, when the portfolio does not: a mismatch against a 1% position is not comparable to the same mismatch against a 30% one.

2. It contains no expected return, so a candidate with an identical risk profile and a materially worse mean scores perfectly.

3. **It is sign-blind.** A matrix norm is a function of $|\Delta|$. A candidate whose covariances with the rest of the portfolio are uniformly *lower* than the incumbent's — a strictly better diversifier, which improves the portfolio on every margin — receives exactly the same score as one whose covariances are uniformly higher by the same amount. The criterion cannot distinguish improvement from damage (Example 1).

The third point is decisive and, we think, not widely appreciated. It is not a matter of choosing a better norm; every norm has it, because every norm is symmetric about zero. The problem is that a distance between *inputs* is being used where a change in an *outcome* is meant.

## 1.2 What this paper does

**A criterion derived rather than posited.** Section 4 takes the object the investor actually cares about — the optimized value of their portfolio problem — and differentiates it. The envelope theorem does the work: because the first-order effect of re-optimising vanishes at an optimum, the derivative of the *value* equals the partial derivative of the objective at *unchanged* weights. The resulting criterion (Theorem 2) is signed, weighted by the portfolio's own weights, contains the mean, and — the practical point — costs exactly the same single matrix product as the naive criterion. Nothing is given up.

**A horizon the covariance cannot see.** A covariance of one-period returns describes co-movement at one frequency. Two assets can move together day by day while their cumulative returns separate permanently, and two others can look weakly related at daily frequency while tracking each other for years. For an investor holding one instead of the other over a horizon $h$,

the second pair is the better substitution and the covariance criterion cannot say so. Section 5 proposes the $h$-period tracking variance, proves (Proposition 3) that its boundedness in $h$ is *equivalent* to cointegration of the log prices with a unit cointegrating vector, and argues that measuring this continuous quantity is preferable to testing for cointegration when the candidate universe is large.

**An honest account of selection bias.** Choosing the minimum of a noisy criterion over $M$ candidates guarantees that the winner looks better than it is. Section 6 makes this precise: the bias grows like $\sqrt{2 \log M}$. The logarithm matters. Reducing a universe from 4,000 candidates to 150 — a factor of about 29 — removes only about 29% of the bias. Universe restriction is often defended as a cure for this problem; it is not one, and we quantify how far short it falls.

**Simulation evidence, including where the method fails.** Section 8 confirms the first-order criterion converges at the predicted $O(\varepsilon^2)$ rate, and also documents a regime where it is wrong by an order of magnitude: when the replaced asset sits at a binding constraint, substitution moves it to a bound and the true welfare change becomes insensitive to which candidate was chosen. We report this because it determines the architecture of the procedure in Section 7: the first-order criterion is used to *rank*, never to *decide*.

### 1.3 Related literature

The problem sits between several established areas without being any of them.

*Index tracking and replication* asks how to reproduce a benchmark with a subset of assets [5, 1, 9]. The objective is similar in spirit — minimise divergence from a target — but the target is a fixed portfolio rather than a single constituent inside an optimization that will be re-solved.

*Pairs trading and statistical arbitrage* is concerned with identifying assets that move together, by distance [16], cointegration [29], or factor decomposition [4]. The statistical question overlaps closely with ours, but the economic one does not: a pairs trader holds the spread and profits from convergence, whereas a substituting investor holds one leg and cares about divergence only as a risk. Section 5 shows this changes which notion of cointegration is relevant — specifically, that the cointegrating vector must be the unit vector.

*Sensitivity analysis of optimized portfolios* established that mean–variance solutions react sharply to their inputs [7, 10, 24], motivating shrinkage [20, 21], constraints as implicit regularisation [18], and scepticism about optimization itself [12]. Our Theorem 2 is an application of the same envelope logic, used constructively rather than as a warning.

*Risk models* express similarity through factor exposures [27, 15, 28], which is how practitioners usually operationalise "comparable asset". We adopt the factor representation in Section 6, for statistical and computational reasons both.

What appears to be missing from this literature is the specific question of *single-constituent replacement inside a portfolio that will be re-optimized*, treated as a decision problem rather than a matching problem. That is the gap this paper addresses.

## 2 Setup and assumptions

We fix notation and state every assumption explicitly, so that later results can be read against the conditions they require.

### 2.1 Returns and portfolios

Let there be $N$ assets held over $T$ periods. Let $r_t \in \mathbb{R}^N$ denote the vector of period-$t$ returns and let

$$\mu = \mathbb{E}[r_t] \in \mathbb{R}^N, \qquad \Sigma = \mathrm{Var}(r_t) \in \mathbb{R}^{N \times N},$$

assumed to exist. A portfolio is a vector of weights $w \in \mathbb{R}^N$ lying in a feasible set $W$.

**Assumption 1** (Feasible set)**.** *$W \subseteq \{w \in \mathbb{R}^N : w^\top \mathbf{1} = 1\}$ is non-empty, compact and convex.*

Assumption 1 covers the budget constraint together with the restrictions used in practice — long-only ($w \geq 0$), position caps ($w_k \leq \bar{w}$), floors on a group of assets ($\sum_{k \in G} w_k \geq \underline{g}$), and any finite intersection of half-spaces. It excludes cardinality constraints, which are not convex; we return to this in Section 9.

**Assumption 2** (Objective)**.** *The investor's objective $U : \mathbb{R}^N \times \Theta \to \mathbb{R}$, written $U(w;\theta)$ for a parameter vector $\theta$, is concave and upper semicontinuous in $w$ for each $\theta$, and continuously differentiable in $\theta$ with $\nabla_\theta U$ continuous in $(w,\theta)$.*

**Assumption 3** (Unique optimum)**.** *For the $\theta$ of interest, $w^*(\theta) = \arg\max_{w \in W} U(w;\theta)$ is a singleton.*

Assumption 3 is used for differentiability of the value function; without it the results below hold with derivatives replaced by directional derivatives (Remark 3).

The leading example, used throughout for concreteness, is mean–variance utility with risk aversion $\gamma > 0$:

$$U(w;\mu,\Sigma) = \mu^\top w - \frac{\gamma}{2} w^\top \Sigma w, \qquad \theta = (\mu,\Sigma). \tag{1}$$

This satisfies Assumption 2 whenever $\Sigma$ is positive semidefinite. Nothing in Section 4 depends on the particular form (1); it is used because it makes the algebra transparent.

### 2.2 Substitution

**Definition 1** (Substitution)**.** Fix an index $i \in \{1,\dots,N\}$ and a candidate asset $y$ with mean $\mu_y$, variance $\sigma_{yy}$ and covariances $\sigma_{yk} = \mathrm{Cov}(r_y, r_k)$ for $k \neq i$. The *substitution of $i$ by $y$* replaces asset $i$ by $y$, producing parameters $\tilde{\theta} = (\tilde{\mu}, \tilde{\Sigma})$ with

$$\tilde{\mu}_i = \mu_y, \quad \tilde{\mu}_k = \mu_k \ (k \neq i), \qquad \tilde{\Sigma}_{ii} = \sigma_{yy}, \quad \tilde{\Sigma}_{ik} = \tilde{\Sigma}_{ki} = \sigma_{yk} \ (k \neq i), \quad \tilde{\Sigma}_{ab} = \Sigma_{ab} \ (a,b \neq i).$$

The last equality is the structural fact the whole paper rests on: replacing one asset leaves the entire $(N-1) \times (N-1)$ block of covariances among the *other* assets untouched. The perturbation $\Delta = \tilde{\Sigma} - \Sigma$ is supported on a single row and column.

**Definition 2** (Value function)**.** $V(\theta) = \max_{w \in W} U(w;\theta)$ is the optimized welfare attainable under parameters $\theta$.

A substitution is desirable to the extent that $V(\tilde{\theta})$ is close to $V(\theta)$, and undesirable to the extent that it is below it. This is the quantity the paper takes as primitive. It is worth being explicit that this is a *choice*, and an arguable one: it says the investor cares about the portfolio they end up holding and not about the identity of its constituents. An investor with a genuine preference over identity — an exclusion, a mandate, a taste — is expressing that through the choice of $i$, not through the criterion.

### 2.3 What is assumed about the return process

Sections 3–4 require only that $\mu$ and $\Sigma$ exist; they are statements about a single-period objective and are agnostic about dynamics. Section 5 is where dynamics enter, and we state there the weaker of the two standard assumptions:

**Assumption 4** (Covariance stationarity of returns)**.** *The bivariate return process $(r_{x,t}, r_{y,t})$ is covariance stationary: means, variances and autocovariances exist and depend on lag only.*

We do *not* assume returns are serially independent. That assumption — equivalent to log prices following a random walk — is precisely what Section 5 is about relaxing, and it is the implicit assumption under which a one-period covariance would be a sufficient description of co-movement at all horizons.

# 3 The matrix-distance criterion and its failure

We first give the naive criterion its due. It is exact, it is efficient, and its derivation is instructive — the same structural fact that makes it cheap also makes the correct criterion cheap.

## 3.1 An exact closed form

**Proposition 1** (Frobenius perturbation under substitution)**.** *Under Definition 1, the perturbation* $\Delta = \tilde{\Sigma} - \Sigma$ *satisfies*

$$\|\Delta\|_F^2 = \underbrace{(\sigma_{yy} - \sigma_{ii})^2}_{\text{variance term}} + 2 \underbrace{\sum_{k \neq i} (\sigma_{yk} - \sigma_{ik})^2}_{\text{cross-covariance term}} .$$

*Proof.* By Definition 1, $\Delta_{ab} = 0$ whenever $a \neq i$ and $b \neq i$. The non-zero entries are therefore $\Delta_{ii}$ and, for each $k \neq i$, the pair $\Delta_{ik} = \Delta_{ki}$. Since $\|\Delta\|_F^2 = \sum_{a,b} \Delta_{ab}^2$, and each off-diagonal entry is counted twice by symmetry, $\|\Delta\|_F^2 = \Delta_{ii}^2 + 2\sum_{k \neq i} \Delta_{ik}^2$. Substituting $\Delta_{ii} = \sigma_{yy} - \sigma_{ii}$ and $\Delta_{ik} = \sigma_{yk} - \sigma_{ik}$ gives the result. □

*Remark* 1 (Why this is computationally attractive)*.* Proposition 1 requires no $N \times N$ matrix to be formed per candidate. Writing $X \in \mathbb{R}^{T \times N}$ for the centred returns of the portfolio and $Y \in \mathbb{R}^{T \times M}$ for those of $M$ candidates, all the inner products $\sigma_{yk}$ for every candidate and every holding are obtained from the single product $X^\top Y \in \mathbb{R}^{N \times M}$. Screening $M$ candidates costs one matrix multiplication rather than $M$ matrix constructions — a reduction from $O(MN^2T)$ to $O(NMT)$. This observation is sound and we retain it; Section 4 applies it to a different integrand.

## 3.2 Three failures

**(i) It ignores the portfolio's weights.** What a perturbation does to the portfolio is mediated by the weights. For fixed $w$, the induced change in portfolio variance is

$$w^\top \Delta w = w_i^2 \, \Delta_{ii} + 2 w_i \sum_{k \neq i} w_k \Delta_{ik}. \tag{2}$$

Every term in (2) is scaled by the weights of the assets involved, whereas Proposition 1 weights each $k$ equally. When weights are equal the two orderings roughly agree; when they are not — and position caps, minimum weights and group floors guarantee they are not — they can disagree arbitrarily.

**(ii) It ignores expected return.** The criterion is a function of second moments alone. A candidate with an identical covariance row and a materially lower mean is scored as a perfect substitute. Since [10] it has been understood that, for plausible risk aversions, errors in means dominate errors in second moments in certainty-equivalent terms; a criterion optimising only the latter is addressing the less important quantity.

**(iii) It cannot distinguish improvement from damage.** This is the failure that no choice of matrix norm repairs.

*Example* 1 (Sign-blindness)*.* Let $\varepsilon > 0$ and consider two candidates $y^+$ and $y^-$ with identical variance $\sigma_{yy} = \sigma_{ii}$, and

$$\sigma_{y^+k} = \sigma_{ik} + \varepsilon, \qquad \sigma_{y^-k} = \sigma_{ik} - \varepsilon, \qquad \text{for every } k \neq i.$$

By Proposition 1 both have $\|\Delta\|_F^2 = 2(N-1)\varepsilon^2$: they are equidistant, and any norm-based criterion is indifferent between them. Yet by (2), for any $w$ with $w_i > 0$ and $w_k > 0$,

$$w^\top \Delta^+ w = 2w_i\varepsilon \sum_{k\neq i} w_k > 0, \qquad w^\top \Delta^- w = -2w_i\varepsilon \sum_{k\neq i} w_k < 0.$$

Substituting $y^-$ *strictly reduces* portfolio variance at unchanged mean and unchanged weights — it is a strictly better diversifier and improves the objective (1). Substituting $y^+$ strictly worsens it. The matrix-distance criterion ranks them identically.

Example 1 is not pathological. Candidates that diversify better than the incumbent are common, and they are exactly the candidates an investor would most want to be shown. A criterion that is structurally incapable of preferring them is not a conservative choice; it is an incorrect one.

*Remark* 2 (Not a problem of the wrong norm). One might hope a better geometry helps. Covariance matrices form a Riemannian manifold rather than a vector space, and there is a mature literature on intrinsic distances between them: affine-invariant [25], log-Euclidean [3], and Bures–Wasserstein [8]. These are the right tools for the question "how far apart are two covariance matrices". They do not address failures (i)–(iii), because each remains a symmetric function of the perturbation and none involves $\mu$ or $w$. The difficulty is not the geometry of the input space; it is that the input space is the wrong place to measure.

# 4 The decision-theoretic criterion

The failures of Section 3 share a cause: a distance between *inputs* was used where a change in an *outcome* was meant. We therefore measure the outcome.

## 4.1 The envelope argument

The obstacle is apparently circularity: evaluating $V(\tilde{\theta})$ for each candidate seems to require re-solving the optimization $M$ times. The envelope theorem removes it.

**Theorem 1** (Marginal welfare of a parameter change)**.** *Under Assumptions 1–3, the value function $V$ of Definition 2 is differentiable at $\theta$ and*

$$\nabla_\theta V(\theta) \;=\; \nabla_\theta U(w;\theta)\Big|_{w=w^*(\theta)}.$$

*Proof.* Write $w^* = w^*(\theta)$ and let $\theta'$ be near $\theta$. Since $w^*$ is feasible for $\theta'$,

$$V(\theta') \;\geq\; U(w^*;\theta') \quad\Longrightarrow\quad V(\theta') - V(\theta) \;\geq\; U(w^*;\theta') - U(w^*;\theta).$$

Similarly, since $w^*(\theta')$ is feasible for $\theta$,

$$V(\theta') - V(\theta) \;\leq\; U\big(w^*(\theta');\theta'\big) - U\big(w^*(\theta');\theta\big).$$

Put $\theta' = \theta + s\,d$ for a direction $d$ and $s > 0$, divide by $s$, and let $s \downarrow 0$. The lower bound tends to $\nabla_\theta U(w^*;\theta)^\top d$. For the upper bound, Berge's maximum theorem together with Assumption 3 gives $w^*(\theta') \to w^*$ as $\theta' \to \theta$, and the continuity of $\nabla_\theta U$ in $(w,\theta)$ makes the upper bound tend to the same limit. The two bounds coincide, so the directional derivative exists and equals $\nabla_\theta U(w^*;\theta)^\top d$ for every $d$; as this is linear in $d$, $V$ is differentiable with the stated gradient. □

The economic content is worth stating plainly, because it is what makes the whole procedure practical: *to first order, the effect of a parameter change on an optimized outcome can be computed while holding the decision fixed.* Re-optimisation is a second-order refinement. The optimizer was already at a point where no small reallocation helps, so the first-order gain from adjusting is zero.

*Remark* 3 (Without uniqueness). If Assumption 3 fails, Danskin's theorem still gives the directional derivative $V'(\theta; d) = \max_{w \in \arg\max} \nabla_\theta U(w;\theta)^\top d$, and everything below holds with "derivative" replaced by "directional derivative". Non-uniqueness is non-generic here and we do not pursue it.

### 4.2 The substitution criterion

**Theorem 2** (First-order welfare impact of a substitution)**.** *Consider the objective* (1) *and the substitution of Definition 1, and let $w^*$ be the optimal weights before substitution. Then the first-order change in optimized welfare is*

$$\boxed{\Delta V_1(y) \;=\; \underbrace{w_i^*\,(\mu_y - \mu_i)}_{\text{mean}} \;-\; \frac{\gamma}{2}\,\underbrace{(w_i^*)^2\,(\sigma_{yy} - \sigma_{ii})}_{\text{own variance}} \;-\; \gamma\, w_i^* \underbrace{\sum_{k \neq i} w_k^*\,(\sigma_{yk} - \sigma_{ik})}_{\text{cross-covariance}}} \tag{3}$$

*and $V(\tilde\theta) - V(\theta) = \Delta V_1(y) + o(\|\tilde\theta - \theta\|)$.*

*Proof.* Apply Theorem 1 with $\theta = (\mu, \Sigma)$. From (1), $\partial U/\partial \mu_i = w_i$ and, treating $\Sigma$ as symmetric, $\partial U/\partial \Sigma_{ii} = -\frac{\gamma}{2} w_i^2$ and $\partial U/\partial \Sigma_{ik} = -\gamma\, w_i w_k$ for $k \neq i$ (the factor 2 arising from the symmetric pair $(i,k),(k,i)$). By Definition 1 the only parameter changes are $\delta\mu_i = \mu_y - \mu_i$, $\delta\Sigma_{ii} = \sigma_{yy} - \sigma_{ii}$ and $\delta\Sigma_{ik} = \sigma_{yk} - \sigma_{ik}$. Contracting the gradient with these increments and evaluating at $w^*$ gives (3). □

Compare (3) with Proposition 1, term by term:

| | **Matrix distance** | **Criterion** (3) |
|---|---|---|
| Weighting over $k$ | uniform | by $w_k^*$ |
| Expected return | absent | present, at order $w_i^*$ |
| Sign | squared, so symmetric | signed |
| Units | squared covariance | welfare |
| Cost for $M$ candidates | one product $X^\top Y$ | one product $X^\top Y$ |

The last row is the practical claim. Writing $\tilde w = w^*$ with the $i$-th entry zeroed, the cross-covariance sum for *all* candidates simultaneously is the single matrix–vector product $\tilde w^\top (X^\top Y) \in \mathbb{R}^M$. The correct criterion is not more expensive than the incorrect one; it is the same computation applied to a different integrand.

**Corollary 1** (Order of the terms in a diversified portfolio)**.** *Suppose weights are of order $1/N$. Then in* (3) *the mean term is $O(1/N)$, the cross-covariance term is $O(1/N)$, and the own-variance term is $O(1/N^2)$.*

*Proof.* Immediate: $w_i^* = O(1/N)$; the cross term is $w_i^* \sum_{k \neq i} w_k^*(\cdot) = O(1/N) \cdot O(1)$ since the sum has $N-1$ terms each $O(1/N)$; the own-variance term carries $(w_i^*)^2$. □

Corollary 1 has a clean interpretation. In a diversified portfolio, *how volatile a candidate is on its own* matters an order $1/N$ less than *how it co-moves with everything else* and than its mean. The common practice of screening substitutes by matching volatility is therefore focused on the least consequential of the three quantities.

### 4.3 Where the linearisation fails

Theorem 2 is a first-order statement, and a substitution is a *discrete* change. Three effects are unmodelled.

1. **Re-optimisation.** Weights move after substitution. The omitted second-order term is non-negative, so (3) is conservative for small perturbations.

2. **Active-set changes.** Theorem 1 assumes the solution varies continuously. If substitution causes a constraint to bind or release — in particular, if the substituted asset is driven to a bound — the map $\theta \mapsto w^*(\theta)$ is non-differentiable there and the linearisation can fail badly. Section 8 exhibits an order-of-magnitude failure of exactly this kind.

3. **Parameter re-estimation.** If $\Sigma$ is estimated by a method whose output depends on the full asset set — shrinkage estimators do — then substituting one asset perturbs the estimated covariances of the others, violating the last clause of Definition 1 by a small amount.

The appropriate response is not a higher-order expansion but a division of labour, which Section 7 formalises: use (3) to *rank* candidates, which it does well and cheaply, and re-solve the optimization exactly for the few survivors, where the cost is affordable and the answer is exact.

# 5 Horizon: what a one-period covariance cannot see

Everything so far has treated the return distribution as a single-period object. This section argues that doing so omits something essential for substitution, identifies precisely what, and proposes a measure of it.

## 5.1 The gap

A covariance $\sigma_{xy}$ describes co-movement at the sampling frequency. It is silent about what happens to *cumulative* returns over a holding period. The two are related only under an assumption that is known to be false.

**Proposition 2** (When one period suffices)**.** *Let $z_t = r_{y,t} - r_{x,t}$ be the spread return and suppose Assumption 4 holds. Write $\gamma_\ell = \mathrm{Cov}(z_t, z_{t+\ell})$ and $\rho_\ell = \gamma_\ell/\gamma_0$. Then the variance of the $h$-period cumulative spread is*

$$D_h^2(x,y) \;\equiv\; \mathrm{Var}\Big(\textstyle\sum_{s=1}^{h} z_{t+s}\Big) \;=\; h\gamma_0 + 2\sum_{\ell=1}^{h-1}(h-\ell)\,\gamma_\ell. \tag{4}$$

*If $\gamma_\ell = 0$ for all $\ell \geq 1$, then $D_h^2 = h\,\gamma_0$.*

*Proof.* Expand the variance of the sum: $\mathrm{Var}(\sum_{s=1}^{h} z_{t+s}) = \sum_{s=1}^{h}\sum_{u=1}^{h}\mathrm{Cov}(z_{t+s}, z_{t+u})$. By stationarity the covariance depends only on $|s-u|$. There are $h$ pairs with $|s-u| = 0$ and $2(h-\ell)$ ordered pairs with $|s-u| = \ell$ for $1 \leq \ell \leq h-1$. Collecting terms gives (4). □

Proposition 2 isolates the assumption. *Only if the spread is serially uncorrelated* does the one-period quantity $\gamma_0$ determine divergence at every horizon, by the square-root-of-time rule. Serial correlation of the spread is exactly what a one-period covariance cannot observe, and it is not zero in practice: the stylised facts of asset returns include predictability structure at longer horizons [11, 22].

**Definition 3** ($h$-period tracking variance)**.** $D_h^2(x,y)$ in (4) is the *$h$-period tracking variance* of $y$ against $x$: the variance of the difference in their cumulative returns over $h$ periods. Its square root, annualised, is the expected tracking error at horizon $h$.

This is the decision-relevant quantity. An investor who substitutes $y$ for $x$ and holds for $h$ periods experiences exactly $\sum_{s \leq h} z_{t+s}$ as the difference in outcome.

### 5.2 Cointegration is the relevant structure — with a unit vector

**Definition 4** (Cointegration)**.** Two $I(1)$ series $p_{x,t}, p_{y,t}$ are *cointegrated* if there exists $\beta \neq 0$ such that $p_{y,t} - \beta p_{x,t}$ is stationary [14].

**Proposition 3** (Bounded tracking $\iff$ unit-vector cointegration)**.** *Let $p_{x,t}, p_{y,t}$ be log prices with $r_{\cdot,t} = \Delta p_{\cdot,t}$, and let $s_t = p_{y,t} - p_{x,t}$ be the log spread. Then*

$$\sup_{h \geq 1} D_h^2(x,y) < \infty \qquad \iff \qquad s_t \text{ is covariance stationary,}$$

*that is, $p_x$ and $p_y$ are cointegrated with cointegrating vector $(1,-1)$. Moreover if $s_t$ is stationary with variance $\gamma_0^s$ and autocorrelation $\rho_h^s$, then*

$$D_h^2(x,y) = 2\gamma_0^s(1-\rho_h^s) \;\leq\; 4\gamma_0^s.$$

*Proof.* The cumulative spread return telescopes: $\sum_{s=1}^{h} z_{t+s} = \sum_{s=1}^{h}(\Delta p_{y,t+s} - \Delta p_{x,t+s}) = s_{t+h} - s_t$. Hence $D_h^2 = \mathrm{Var}(s_{t+h} - s_t)$.

($\Leftarrow$) If $s_t$ is covariance stationary, $\mathrm{Var}(s_{t+h} - s_t) = 2\gamma_0^s - 2\,\mathrm{Cov}(s_{t+h}, s_t) = 2\gamma_0^s(1-\rho_h^s)$, which lies in $[0, 4\gamma_0^s]$ for every $h$, so the supremum is finite.

($\Rightarrow$) Suppose $\sup_h \mathrm{Var}(s_{t+h} - s_t) = B < \infty$. Since $z_t$ is stationary by Assumption 4, $s_t$ is an integrated process whose increments are stationary; if $s_t$ were $I(1)$ its Beveridge–Nelson decomposition would contain a random walk component with strictly positive innovation variance $\sigma_\eta^2$, giving $\mathrm{Var}(s_{t+h} - s_t) \geq \sigma_\eta^2 h \to \infty$ and contradicting boundedness. Hence the random walk component is degenerate and $s_t$ is $I(0)$. □

Proposition 3 makes two points that we have not seen stated together.

First, it identifies *cointegration* as the exact structural condition under which a substitution is safe at long horizons, and shows the condition is equivalent to a statement about an observable second moment — no unobserved common trend needs to be estimated.

Second, and less obviously: **the cointegrating vector must be** $(1,-1)$. General cointegration with $\beta \neq 1$ does not bound $D_h$. This distinguishes substitution from pairs trading, where any $\beta$ is usable because the trader sizes the two legs to $\beta$ and holds the stationary combination. A substituting investor holds *one unit of one asset instead of one unit of another* and cannot rescale; only the unit combination is relevant to them. The distinction is easy to lose because both problems are described as "finding cointegrated pairs".

### 5.3 Measuring, not testing

**Definition 5** (Spread variance ratio)**.** $\mathrm{VR}_h(x,y) = \dfrac{D_h^2(x,y)}{h\,\gamma_0} = 1 + 2\displaystyle\sum_{\ell=1}^{h-1}\Big(1 - \frac{\ell}{h}\Big)\rho_\ell$, the ratio of realised $h$-period tracking variance to what it would be under serially uncorrelated spread returns.

The second equality follows by dividing (4) by $h\gamma_0$. This is the Lo–MacKinlay variance ratio [22] applied to the spread rather than to a single series, and it reads directly:

| Regime | Spread autocorrelation | $\mathrm{VR}_h$ | $D_h^2$ as $h \to \infty$ |
|---|---|---|---|
| Uncorrelated increments | $\rho_\ell \approx 0$ | $\approx 1$ | grows $\propto h$ |
| **Cointegrated** $(1,-1)$ | net negative | $< 1$ | **bounded** |
| Divergent | net positive | $> 1$ | grows faster than $h$ |

We argue this continuous statistic should be preferred to a cointegration *test* whenever many candidates are screened, for four reasons.

1. **Multiplicity.** Screening one incumbent against $M$ candidates is $M$ tests. At $M = 4{,}000$ and a 5% level, roughly 200 spurious rejections occur by construction. False-discovery control [6] restores validity but costs power, which is already the binding problem:

2. **Power.** Unit-root and cointegration tests are known to have low power against persistent alternatives at sample sizes typical of financial panels. Failure to reject is close to uninformative.

3. **Discretisation.** A test returns a binary verdict. It does not distinguish a spread that reverts within weeks from one that reverts within a decade — yet that difference is what determines whether the substitution is safe over the investor's actual horizon. $D_h$ answers the question at the horizon asked.

4. **Decision relevance.** Cointegration is a modelling assumption that *implies* bounded divergence. By Proposition 3 bounded divergence is what we care about, and it is directly measurable. Estimating the quantity of interest is preferable to testing for a model that implies a property of it.

*Remark* 4 (Estimation warnings). Two practical cautions. (i) With overlapping $h$-period windows the naive estimator of (4) is biased downward in finite samples; the [22] bias correction should be applied, and confidence intervals obtained by a block or stationary bootstrap [26] rather than under an i.i.d. null. (ii) The normalisation in Definition 5 is a common source of error: dividing by $h$ twice — once in forming the $h$-period variance and again in the ratio — produces the artefact $\mathrm{VR}_h \approx 1/h$ for every pair, which is superficially plausible and uniformly wrong. Any implementation should be validated against a simulated process with a known theoretical ratio before being applied to data.

### 5.4 Choosing $h$

$D_h$ requires a horizon, and the horizon should be the investor's. As $h \to 1$, $D_h^2 \to \gamma_0 = \mathrm{Var}(r_y - r_x)$, which is the one-period criterion: the covariance-based notion of similarity is recovered as the short-horizon special case rather than being discarded. As $h$ grows, the autocovariance terms in (4) accumulate and the criterion increasingly rewards pairs whose spreads revert — that is, cointegrated pairs. A single formula therefore spans both regimes, with the investor's horizon selecting between them.

## 6 Estimation: noise, factors, and the winner's curse

The criteria of Sections 4–5 are stated in population moments. In practice they are estimated, and the estimation problem has a feature that is easy to overlook: we are not estimating a quantity, we are *selecting the minimum* of many estimated quantities. That changes what can be claimed about the winner.

### 6.1 Selection bias from screening

**Proposition 4** (Order of the selection bias)**.** *Let $d_1, \dots, d_M$ be true dissimilarities and $\hat{d}_j = d_j + \varepsilon_j$ their estimates, with $\varepsilon_j$ i.i.d. $N(0, s^2)$ independent of $d$. Let $\hat{\jmath} = \arg\min_j \hat{d}_j$. If the $d_j$ are equal, then*

$$\mathbb{E}[d_{\hat{\jmath}} - \hat{d}_{\hat{\jmath}}] \;=\; s\,\mathbb{E}[\max_j Z_j] \;=\; s\Big(\sqrt{2\log M} - \frac{\log\log M + \log 4\pi}{2\sqrt{2\log M}} + o(1)\Big),$$

*where $Z_j$ are i.i.d. standard normal. That is, the selected candidate's estimated dissimilarity understates its true dissimilarity by an amount of order $s\sqrt{2\log M}$.*

*Proof.* With all $d_j$ equal to $d$, $\arg\min_j \hat{d}_j = \arg\min_j \varepsilon_j$ and $\hat{d}_{\hat{\jmath}} = d + \min_j \varepsilon_j = d - s \max_j(-\varepsilon_j/s)$. Since $-\varepsilon_j/s$ are i.i.d. standard normal, taking expectations gives $\mathbb{E}[d - \hat{d}_{\hat{\jmath}}] = s\,\mathbb{E}[\max_j Z_j]$. The asymptotic expansion of the expected maximum of $M$ i.i.d. standard normals is classical [13]. □

The practical content is in the logarithm, and it is worth being blunt about it because the implication is frequently overstated in the opposite direction.

*Example* 2 (Restricting the universe helps only logarithmically). Take $M = 4{,}334$ candidates. Proposition 4 gives an expected optimism of about 3.5 standard errors. Restricting the search to, say, the 150 candidates sharing the incumbent's industry — a *29-fold* reduction — lowers this only to about 2.5 standard errors, a reduction of roughly 29%. Simulation agrees with the asymptotic formula to about 0.1 standard errors at both sizes.

Universe restriction is often defended on the grounds that it controls data-mining bias. Example 2 shows the effect is real but modest: because the bias grows like $\sqrt{\log M}$, order-of-magnitude reductions in $M$ buy only tens of percent in bias. Restriction should be justified on other grounds — relevance to the investor's constraints, computational cost — and the selection bias addressed directly:

1. **Sample splitting.** Rank on one sub-sample, verify the ranking on a disjoint one. A candidate whose ranking does not survive is rejected. This targets the bias at its source.
2. **Resampling stability.** Require the candidate to remain highly ranked across bootstrap resamples [26], which for dependent data should be a block or stationary bootstrap.
3. **Shrinkage of the score.** Rank on an estimate shrunk toward a pooled value, so a candidate must beat its peers by a margin exceeding the noise.

### 6.2 Factor representation

Estimating $\sigma_{yk}$ for every candidate against every holding by sample covariance is both noisy and, when the candidate universe is large, awkward: it requires the full candidate return panel. A factor model resolves both.

Assume the linear factor structure

$$r_{j,t} = \beta_j^\top f_t + \epsilon_{j,t}, \qquad \mathbb{E}[\epsilon_{j,t} \mid f_t] = 0, \qquad \mathrm{Cov}(\epsilon_j, \epsilon_k) = 0 \ (j \neq k), \tag{5}$$

with $f_t \in \mathbb{R}^K$, $K \ll N + M$. Then

$$\sigma_{jk} = \beta_j^\top \Sigma_f \, \beta_k \ (j \neq k), \qquad \sigma_{jj} = \beta_j^\top \Sigma_f \, \beta_j + s_j^2. \tag{6}$$

Three consequences.

1. *Conditioning.* (6) replaces $O((N+M)^2)$ free parameters with $O((N+M)K)$, which is the standard remedy for the instability of sample covariance matrices in high dimension [20, 19].
2. *Sufficiency.* Every quantity in criterion (3) depends on the candidate only through $(\beta_y, s_y, \mu_y)$. A compact per-candidate summary is therefore *sufficient* for screening: the candidate return panel need not be retained. For a universe of a few thousand instruments and $K \approx 15$, the summary is under a megabyte.
3. *Interpretability.* Under (5), similarity becomes a statement about factor exposures,

   $$d_{\text{factor}}(x, y)^2 = (\beta_x - \beta_y)^\top \Sigma_f (\beta_x - \beta_y) + (s_x - s_y)^2,$$

   which is how similarity is usually operationalised in practice [27, 28] and which supports an explanation to the investor of *why* two assets were judged comparable.

*Remark* 5 (Second moments are not the whole story). Criterion (3) and the factor representation are second-moment constructions. There is substantial evidence that dependence between assets is asymmetric and stronger in the lower tail than a covariance implies [23, 2], and that return distributions are heavy-tailed enough for the fourth moment to be questionable [11]. Two candidates matched on (3) may therefore differ materially in downside co-movement. A practical procedure should carry a tail-risk screen — a comparison of tail indices [17] or downside semideviations — as an admissibility condition rather than as another term in the objective, since the sample sizes that support tail estimation rarely support optimising over it.

# 7 A two-stage procedure

We now assemble the pieces. The design responds to a specific asymmetry: criterion (3) is cheap and approximately right, while re-solving the optimization is expensive and exactly right. Section 4.3 showed the approximation can fail badly at discrete substitutions. The resolution is to let each do what it is good at.

## 7.1 Statement

**Stage 0 — admissibility.** Restrict to candidates satisfying the investor's constraints, excluding: assets already held; assets representing the *same underlying* through a different listing; assets failing data-quality requirements; and assets failing the tail screen.

**Stage 1 — rank (all $M$ candidates).** Evaluate (3) for every admissible candidate via a single matrix product, together with the horizon-matched tracking error $D_h$ of Definition 3. Retain the $K$ best, $K$ small.

**Stage 2 — decide ($K$ candidates).** Re-solve the optimization exactly for each retained candidate under the *same* constraint set, obtaining the true $V(\tilde{\theta}) - V(\theta)$. Rank by structural similarity subject to a welfare floor:

$$\min_{y} \; D_h(x_i, y) \quad \text{s.t.} \quad V(\tilde{\theta}_y) - V(\theta) \;\geq\; -\tau.$$

Report the best few, with the realised welfare change disclosed.

## 7.2 Why the objective and the constraint are ordered this way

It would be simpler to rank by $\Delta V$ alone. It would also answer the wrong question. The candidate maximising $\Delta V$ over the whole universe is, approximately, the best available asset — which the optimizer would already have selected had it been admissible. Ranking that way returns a *re-optimization*, not a *substitution*, and it will happily propose an asset with no resemblance to the one the investor asked to replace.

Conversely, ranking by similarity alone ignores cost. The constrained form separates the two roles cleanly: welfare is a floor the investor is entitled to, and similarity is what the word "alternative" means. The tolerance $\tau$ is then an interpretable quantity — "no worse than $\tau$ per year" — rather than an arbitrary exchange rate between incommensurable units, which is what a weighted sum of the two would require.

## 7.3 Cost

Let $T$ be the sample length, $N$ the portfolio size, $M$ the candidate count, $K$ the shortlist, $C_{\text{opt}}$ the cost of one optimization.

| Stage | Work | Cost | Typical |
|---|---|---|---|
| 1, sample covariance | one product $X^\top Y$ | $O(NMT)$ | $10^8$ flops |
| 1, factor form (6) | one product $B(\Sigma_f \beta)$ | $O(MK)$ | $10^4$ flops |
| 2 | $K$ exact re-solves | $K\,C_{\text{opt}}$ | **dominant** |

Stage 1 is negligible under the factor representation. *The entire cost of the procedure is $K$ optimizations*, which makes $K$ the only tuning parameter that matters and makes the ranking quality of Stage 1 — not its accuracy — the property to verify. Section 8 verifies it.

*Remark* 6 (Eager versus lazy evaluation). Because the cost is $KC_{\text{opt}}$ per incumbent, computing alternatives for *every* holding in advance costs $NKC_{\text{opt}}$, paid whether or not anyone asks. For a ten-asset portfolio with $K = 5$ this is fifty optimizations per portfolio construction — an order-of-magnitude increase in the cost of the base service, incurred on behalf of a request that may never arrive. Lazy evaluation, on request, is the appropriate default; results should be cached against the basket, since substituting one asset invalidates the weights $w^*$ at which (3) was evaluated for all the others.

# 8 Simulation evidence

We verify three claims: that criterion (3) is first order in the sense of Theorem 2; that it fails where Section 4.3 predicts; and that it nonetheless ranks well enough for the procedure of Section 7 to recover the exact optimum with a small shortlist.

**Design.** Returns follow the factor structure (5) with $K = 3$ factors, $T = 1{,}500$ periods, $N = 8$ portfolio assets and, for the ranking experiment, $M = 300$ candidates sharing the same factor structure. The objective is of the form (1) with a state-dependent effective risk coefficient, maximised over the simplex intersected with a position cap. Candidate universes and portfolios are drawn independently across replications.

## 8.1 Is the criterion first order?

We interpolate between the incumbent and a candidate, $y(\varepsilon) = (1 - \varepsilon)x_i + \varepsilon\, y$, and compare (3) against the exact re-optimized change. If the criterion is first order, the error is $O(\varepsilon^2)$ and must fall by a factor of about 4 each time $\varepsilon$ is halved; a criterion that were merely *correlated* with the truth would show a factor of about 2.

Table 1: Approximation error as the perturbation shrinks.

| $\varepsilon$ | predicted $\Delta V_1$ | exact $\Delta V$ | \|error\| |
|---|---|---|---|
| 1.000 | $-5.83 \times 10^{-4}$ | $-5.27 \times 10^{-5}$ | $5.31 \times 10^{-4}$ |
| 0.500 | $-2.78 \times 10^{-4}$ | $-5.27 \times 10^{-5}$ | $2.25 \times 10^{-4}$ |
| 0.250 | $-1.36 \times 10^{-4}$ | $-5.27 \times 10^{-5}$ | $8.30 \times 10^{-5}$ |
| 0.125 | $-6.70 \times 10^{-5}$ | $-4.49 \times 10^{-5}$ | $2.22 \times 10^{-5}$ |
| *error shrink factor per halving:* 2.4, 2.7, 3.7 | | | |

The shrink factors are super-linear and approach 4, consistent with $O(\varepsilon^2)$ convergence. Theorem 2 is confirmed.

## 8.2 Where it fails — and why we report it

Table 1 also shows the criterion badly wrong at $\varepsilon = 1$: it predicts $-5.83 \times 10^{-4}$ against a true $-5.27 \times 10^{-5}$, an error of an order of magnitude. The diagnosis is visible in the table: the exact $\Delta V$ is *identical* for $\varepsilon = 1$, 0.5 and 0.25. Three genuinely different candidates produced the same welfare outcome.

The cause is the active-set effect of Section 4.3, item 2. The replaced asset sat at its position cap; after substitution the optimizer moved it to a bound, so its weight — and hence the portfolio — became independent of which candidate was used. Where the solution map is non-differentiable, a first-order expansion carries no information.

This is not a defect to be engineered away, and we report it prominently because it determines the architecture. It is the reason Section 7 uses (3) to rank and never to decide. An implementation that reported $\Delta V_1$ to a user as the cost of a substitution would be wrong by an order of magnitude precisely in the case that matters most — a large position at a binding constraint.

### 8.3 Does it rank well enough?

Ranking is the only property Stage 1 requires. For each replication we compute (3) for all $M = 300$ candidates and, separately, the exact re-optimized $\Delta V$ for all 300 — the brute-force ground truth — then ask two questions: how well does the criterion rank, and does a shortlist of $K$ contain the true best three?

Table 2: Ranking quality of criterion (3) against brute force, $M = 300$ candidates.

| Replication | $w_i^*$ | Spearman $\rho$ | recall@12 | recall@25 |
|---|---|---|---|---|
| 1 | 0.100 | +0.705 | 100% | 100% |
| 2 | 0.345 | +0.982 | 100% | 100% |
| 3 | 0.100 | +0.919 | 100% | 100% |
| mean | | +0.869 | **100**% | **100**% |

"recall@$K$" is the fraction of the true best three that appear in the criterion's top $K$. Across replications it is 100% at $K = 12$, and rank correlation with the exact outcome averages +0.87.

The interpretation matters. The criterion is *unreliable as an estimate* and *reliable as an ordering* — these are different properties, and the two-stage design needs only the second. Since Stage 2 evaluates the shortlist exactly, an ordering good enough to place the true winners inside a shortlist of a dozen is sufficient to recover the exact optimum at a cost of a dozen optimizations rather than several thousand.

*Remark* 7 (Limits of this evidence). The design is synthetic: a correctly specified factor model, no estimation error in the moments, 300 candidates rather than thousands, and a feasible set consisting only of the simplex and a box — not the group floors and budget constraints of a production problem. Each of these makes the task easier than reality. What the experiment establishes is that the *design* is sound and that the failure mode is the predicted one; it does not establish a value of $K$ for any particular application, which should be determined empirically on the data at hand.

## 9 Discussion

### 9.1 Summary

The question "which asset is a good substitute for this one?" invites a matching answer and requires a decision-theoretic one. We have argued four things.

1. A matrix distance between covariance matrices is the wrong criterion, most fundamentally because it is sign-blind: it cannot prefer a candidate that strictly improves the portfolio to one that strictly damages it (Example 1). No choice of norm repairs this, because the defect is in measuring inputs rather than outcomes.

2. The envelope theorem yields the right criterion at the same computational cost (Theorem 2). It is signed, weighted by the portfolio's own weights, and contains the expected return. A corollary is that a candidate's own volatility is an order $1/N$ less consequential than its co-movement with the rest of the portfolio — which inverts the usual screening practice.

3. A one-period covariance is a single-horizon object. The $h$-period tracking variance is the horizon-aware replacement, and its boundedness is *equivalent* to cointegration of the log prices with a unit cointegrating vector (Proposition 3) — a stricter requirement than the cointegration used in pairs trading, because a substituting investor cannot rescale the legs. Measuring this continuous quantity dominates testing for cointegration when many candidates are screened.

4. Selecting the best of many noisy candidates biases the winner by $O(s\sqrt{2\log M})$. The logarithm is the point: restricting the candidate universe by an order of magnitude removes only tens of percent of the bias (Example 2), so restriction is not a substitute for sample splitting or resampling.

### 9.2 Limitations

**Convexity.** Assumption 1 excludes cardinality constraints and other combinatorial restrictions, under which the value function is not differentiable and Theorem 1 does not apply. Substitution is itself a discrete operation, so the theory here is best read as covering the *continuous relaxation* in which the substitution has been made and the remaining weights re-optimized.

**Stationarity.** Assumption 4 underlies Section 5. Structural breaks — corporate actions, regime changes, index reconstitution — break both the covariance and the cointegrating relationship, and a pair that tracked historically may not continue to. Nothing here is a forecast; every quantity is an unconditional in-sample moment.

**Second moments.** Criterion (3) follows from a mean–variance objective. For objectives sensitive to higher moments the envelope argument is unchanged — Theorem 1 requires no particular functional form — but the explicit expression (3) would acquire terms in the higher moments, and the evidence on asymmetric tail dependence [23, 2] suggests those terms would not be negligible.

**Estimation error is treated, not solved.** Proposition 4 quantifies selection bias under a convenient Gaussian model with equal true dissimilarities. Real candidate sets have heterogeneous $d_j$ and dependent errors, both of which change the constant though not the $\sqrt{\log M}$ rate.

### 9.3 Directions

Three extensions seem worthwhile. First, a *second-order* criterion using the sensitivity $\partial w^*/\partial\theta$ from the KKT system, which would capture re-optimisation without a full re-solve and might shrink the shortlist further. Second, a treatment of *simultaneous* substitution of several holdings, where the first-order terms interact and the sequential procedure used here is not obviously optimal. Third, replacing the tail *screen* with a tail-aware objective, which would require an estimator of lower-tail dependence stable enough to optimise over — currently the binding obstacle.

### Reproducibility

All results in Section 8 come from simulated data generated by the process described there; no proprietary data or procedure is used anywhere in this paper. The propositions are self-contained and their proofs are given in full.

## References


[1] Alexander, Carol; Dimitriu, Anca (2005). *Indexing and Statistical Arbitrage*. Journal of Portfolio Management 31(2), 50–63.

[2] Ang, Andrew; Chen, Joseph (2002). *Asymmetric Correlations of Equity Portfolios*. Journal of Financial Economics 63(3), 443–494.

[3] Arsigny, Vincent; Fillard, Pierre; Pennec, Xavier; Ayache, Nicholas (2006). *Log-Euclidean Metrics for Fast and Simple Calculus on Diffusion Tensors*. Magnetic Resonance in Medicine 56(2), 411–421.

[4] Avellaneda, Marco; Lee, Jeong-Hyun (2010). *Statistical Arbitrage in the US Equities Market*. Quantitative Finance 10(7), 761–782.

[5] Beasley, J. E.; Meade, N.; Chang, T.-J. (2003). *An Evolutionary Heuristic for the Index Tracking Problem*. European Journal of Operational Research 148(3), 621–643.

[6] Benjamini, Yoav; Hochberg, Yosef (1995). *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing*. Journal of the Royal Statistical Society, Series B 57(1), 289–300.

[7] Best, Michael J.; Grauer, Robert R. (1991). *On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means*. Review of Financial Studies 4(2), 315–342.

[8] Bhatia, Rajendra; Jain, Tanvi; Lim, Yongdo (2019). *On the Bures–Wasserstein Distance Between Positive Definite Matrices*. Expositiones Mathematicae 37(2), 165–191.

[9] Brodie, Joshua; Daubechies, Ingrid; De Mol, Christine; Giannone, Domenico; Loris, Ignace (2009). *Sparse and Stable Markowitz Portfolios*. Proceedings of the National Academy of Sciences 106(30), 12267–12272.

[10] Chopra, Vijay K.; Ziemba, William T. (1993). *The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice*. Journal of Portfolio Management 19(2), 6–11.

[11] Cont, Rama (2001). *Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues*. Quantitative Finance 1(2), 223–236.

[12] DeMiguel, Victor; Garlappi, Lorenzo; Uppal, Raman (2009). *Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?*. Review of Financial Studies 22(5), 1915–1953.

[13] Embrechts, Paul; Klüppelberg, Claudia; Mikosch, Thomas (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.

[14] Engle, Robert F.; Granger, C. W. J. (1987). *Co-Integration and Error Correction: Representation, Estimation, and Testing*. Econometrica 55(2), 251–276.

[15] Fama, Eugene F.; French, Kenneth R. (1993). *Common Risk Factors in the Returns on Stocks and Bonds*. Journal of Financial Economics 33(1), 3–56.

[16] Gatev, Evan; Goetzmann, William N.; Rouwenhorst, K. Geert (2006). *Pairs Trading: Performance of a Relative-Value Arbitrage Rule*. Review of Financial Studies 19(3), 797–827.

[17] Hill, Bruce M. (1975). *A Simple General Approach to Inference About the Tail of a Distribution*. The Annals of Statistics 3(5), 1163–1174.

[18] Jagannathan, Ravi; Ma, Tongshu (2003). *Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps*. Journal of Finance 58(4), 1651–1683.

[19] Laloux, Laurent; Cizeau, Pierre; Bouchaud, Jean-Philippe; Potters, Marc (1999). *Noise Dressing of Financial Correlation Matrices*. Physical Review Letters 83(7), 1467–1470.

[20] Ledoit, Olivier; Wolf, Michael (2004). *A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices*. Journal of Multivariate Analysis 88(2), 365–411.

[21] Ledoit, Olivier; Wolf, Michael (2017). *Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks*. Review of Financial Studies 30(12), 4349–4388.

[22] Lo, Andrew W.; MacKinlay, A. Craig (1988). *Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test*. Review of Financial Studies 1(1), 41–66.

[23] Longin, François; Solnik, Bruno (2001). *Extreme Correlation of International Equity Markets*. Journal of Finance 56(2), 649–676.

[24] Michaud, Richard O. (1989). *The Markowitz Optimization Enigma: Is Optimized Optimal?*. Financial Analysts Journal 45(1), 31–42.

[25] Pennec, Xavier; Fillard, Pierre; Ayache, Nicholas (2006). *A Riemannian Framework for Tensor Computing*. International Journal of Computer Vision 66(1), 41–66.

[26] Politis, Dimitris N.; Romano, Joseph P. (1994). *The Stationary Bootstrap*. Journal of the American Statistical Association 89(428), 1303–1313.

[27] Rosenberg, Barr (1974). *Extra-Market Components of Covariance in Security Returns*. Journal of Financial and Quantitative Analysis 9(2), 263–274.

[28] Sharpe, William F. (1992). *Asset Allocation: Management Style and Performance Measurement*. Journal of Portfolio Management 18(2), 7–19.

[29] Vidyamurthy, Ganapathy (2004). *Pairs Trading: Quantitative Methods and Analysis*. John Wiley & Sons.