Untangling Universality and Dispelling Myths in Mean-Variance Optimization

Jerome Benveniste, Petter N. Kolm and Gordon Ritter

April 24, 2024

Jerome Benveniste is an adjunct professor at the Courant Institute of Mathematical Sciences, New York University, NY, USA.
ejb14@nyu.edu

Petter N. Kolm is a clinical professor and director of the Mathematics in Finance Master's Program at the Courant Institute of Mathematical Sciences, New York University, NY, USA.
petter.kolm@nyu.edu

Gordon Ritter is an adjunct professor at the Courant Institute of Mathematical Sciences, Baruch College, Columbia University and General Partner/CIO at Ritter Alpha LP, NY, USA.
ritter@post.harvard.edu

ABSTRACT

Following Markowitz's pioneering work on mean-variance optimization (MVO), such approaches have permeated nearly every facet of quantitative finance. In the first part of the article, the authors argue that their widespread adoption can be attributed to the universality of the mean-variance paradigm, wherein the maximum expected utility and mean-variance allocations coincide for a broad range of distributional assumptions of asset returns. Subsequently, they introduce a formal definition of mean-variance equivalence and present a novel and comprehensive characterization of distributions, termed mean-variance-equivalent (MVE) distributions, wherein expected utility maximization and the solution of an MVO problem are the same. In the second part of the article, the authors address common myths associated with MVO. These myths include the misconception that MVO necessitates normally-distributed asset returns, the belief that it is unsuitable for cases with asymmetric return distributions, the notion that it maximizes errors, and the perception that it underperforms a simple $1/n$ portfolio in out-of-sample tests. Furthermore, they address misunderstandings regarding MVO's ability to handle signals across different time horizons, its treatment of transaction costs, its applicability to intraday and high-frequency trading, and whether quadratic utility accurately represents investor preferences.

TOPICS

Investment management; Mean-variance optimization; Mean-variance equivalent distributions; Portfolio optimization; Portfolio theory; Robust portfolio management; Trading; Universality

KEY FINDINGS

- The authors argue that the widespread adoption of mean-variance approaches can be attributed to the universality of the mean-variance paradigm. This universality dictates that the maximum expected utility and mean-variance allocations coincide for a broad range of distributional assumptions of asset returns and utility functions.

- The authors present a novel and comprehensive characterization of distributions, termed mean-variance-equivalent distributions, where the maximization of expected utility coincides with the solution of an MVO problem for any standard utility function.

- A number of common myths associated MVO are addressed, including normally-distributed asset returns, asymmetric return distributions, error maximization, $1/n$ portfolios, signals across different time horizons, transaction costs, and intraday as well as high-frequency trading.

- A contemporary and practically implementable prescription is given for how to use value functions to solve multi-period optimization with a mean-quadratic-variation objective.

> *"It is now more than half a century since Markowitz [1959] first defended MV [mean-variance] analysis as a practical way to approximately maximize EU [expected utility]. In light of repeated confirmation since then of the efficacy of MV approximations to EU, the persistence of the Great Confusion – that MV analysis is applicable in practice only when return distributions are Gaussian or utility functions quadratic – is as if geography textbooks of 1550 still described the Earth as flat."* [Markowitz et al., 2013, Ch. 2]

With hindsight gained over almost 75 years, we are in the enviable position of being able to place Markowitz' pioneering work in historical context, and use subsequent developments to shed new light on his contributions. Rubinstein [2011] defines 1950 as the boundary between the "ancient" and "classical" periods in the history of financial economics. The critical juncture is the publication of Harry Markowitz's seminal article, "Portfolio Selection" in 1952 [Markowitz, 1952]. In this article, Markowitz addresses a fundamental question in investment management of how investors should allocate their capital among a diverse set of investment options. He formally quantifies the return and risk of assets through their expected returns and their pairwise return covariances. Then, Markowitz proposes that investors should consider the trade-off between portfolio risk and return when determining the allocation across investment alternatives.

His work is groundbreaking for several reasons. First, it represents a significant departure from traditional financial wisdom by challenging the notion that investors should solely prioritize assets based on which ones offer the highest future value relative to their current price, without taking risk into account. Second, it is arguably the first time a financial decision-making process is formalized as a mathematical optimization problem. Specifically, the *mean-variance optimization* (MVO) (or simply "mean-variance" for short) problem prescribes that among the infinite number of portfolios achieving a particular desired return, the investor should select the portfolio with the lowest risk, measured as portfolio variance. Portfolios with identical target returns but greater portfolio variances are deemed "inefficient," owing to their increased risk.

In some sense, Markowitz' work came far too early, thus contributing to *the Great Confusion* mentioned in the quote above. The optimization of the risk-reward trade-off in portfolio theory by means of MVO is not an accident; indeed, mean-variance is a necessary consequence of utility theory for the broad family of elliptical distributions, as demonstrated by Chamberlain [1983]. In the 1960s and 1970s, the foundation for multi-factor risk models are laid by Sharpe, Blume, King and Rosenberg [Sharpe, 1963; Blume, 1972; King, 1966; Rosenberg, 1974].[1] It is now known that mean-variance optimization remains numerically stable when we employ "reasonable" multi-factor risk models and

[1] For a recent discussion of multi-factor risk models see, for example, Connor et al. [2010], Fama et al. [2020], and Jacobs et al. [2021].

expected return forecasts. Therefore, Markowitz' pioneering work came about 25 years too early to stave off controversies around stability, and about thirty years too early to avoid questions about whether MVO is truly optimal for elliptical densities. That said, the evolution of the many theoretical and practical aspects of mean-variance optimization (MVO) is undeniably intriguing.[2]

The outline of the article is as follows. We start with a brief review of classical MVO, while also introducing some notation that will be used throughout. The rest of the article is split into two parts. Since Markowitz's pioneering work, mean-variance approaches have become ubiquitous across nearly every subfield of quantitative finance. In the first part, we argue that this widespread adoption can be attributed to the *universality* of the mean-variance paradigm, wherein the maximum expected utility and mean-variance allocations coincide. We refer to a utility function as *standard* if it is increasing, strictly concave, and continuously differentiable. A probability distribution on asset returns is said to be *mean-variance equivalent* (MVE) if the maximum expected utility and mean-variance allocations coincide for any standard utility function. We fully characterize MVE distributions, and find it to be a broad class that encompasses all elliptical distributions and a subset of asymmetric distributions. This is important because most people, including decision-making experts, cannot precisely articulate their utility function. Our results imply that for a broad class of distributions, the inability of the investor to specify their utility function *does not matter* – the optimal allocation is a mean-variance allocation for any standard utility function.

In the second part of the article, we shift our focus to common myths surrounding MVO. These include beliefs that MVO requires normally-distributed asset returns, cannot or should not be applied in cases where return distributions are asymmetric, is an error maximizer, and fails to outperform a $1/n$ portfolio out-of-sample. Additionally, there are misunderstandings regarding MVO's handling of signals at different horizons, transaction costs, applicability for intraday and high-frequency trading, and whether quadratic utility reflects investor preferences. While several of these myths have been addressed in prior literature (see, for example, [Kinlaw et al., 2017, Ch. 3–7]), we offer new insights and perspectives on them. We relegate mathematical proofs to the appendix, which readers can comfortably skip without losing sight of the article's main themes.

This work is not meant to be a survey, or a review of the pros and cons, of MVO and other approaches to portfolio construction. Nor do we advocate for MVO as the exclusive method for portfolio construction. Rather, while MVO may be considered the "grandparent" of other approaches, it is one of several techniques, each with its own advantages and disadvantages depending on the context and practical application at hand.

[2]We direct interested readers to Markowitz [1990], Markowitz [1991], Bernstein [1993], Fabozzi et al. [2007], Markowitz [2010], Rubinstein [2011], Markowitz et al. [2013], Kolm et al. [2014], Markowitz [2016], and Markowitz [2020].

MEAN-VARIANCE OPTIMIZATION

We consider a market with n investable assets and let $\mathbf{r}$ denote the n-dimensional random vector representing the one period asset returns over the investment horizon $[0, T]$. We *do not* assume that $\mathbf{r}$ follows a multivariate normal distribution, although in some later sections, we will impose restrictions on its distributional form. For notational convenience, we reserve special notation for the expected returns and covariance matrix of returns of $\mathbf{r}$

$$\boldsymbol{\mu} := \mathbb{E}[\mathbf{r}] \in \mathbb{R}^n, \quad \text{and} \quad \boldsymbol{\Sigma} := \mathbb{V}[\mathbf{r}] \in \mathbb{R}^{n \times n}, \tag{1}$$

which we assume exist.

In classical portfolio theory [Markowitz, 1952], a mean-variance investor determines their portfolio holdings $\mathbf{h} = (h_1, \ldots, h_n)'$ in units of dollars, or whatever currency the investor is using, by solving the MVO problem

$$\max_{\mathbf{h} \in \mathcal{C}} \mu_p - \frac{\lambda}{2}\sigma_p^2 \,, \tag{2}$$

where

$$\mu_p := \mathbb{E}[\mathbf{h}'\mathbf{r}] = \mathbf{h}'\boldsymbol{\mu} \quad \text{and} \tag{3}$$

$$\sigma_p^2 := \mathbb{V}[\mathbf{h}'\mathbf{r}] = \mathbf{h}'\boldsymbol{\Sigma}\mathbf{h} \,, \tag{4}$$

denote the *ex ante* portfolio return and variance respectively, $\mathcal{C} \subseteq \mathbb{R}^n$ is the set of constraints, $\lambda > 0$ is the investor's risk aversion, and prime denotes the transpose. Constraints may include long-only, trading, exposure and risk management constraints (see, for example, Fabozzi et al. [2007] and Kolm et al. [2014] for a discussion of common constraints used by portfolio managers).

Throughout the article, we assume that there are no perfectly redundant assets in the market.

Assumption 1. *Our marketplace cannot contain any linear combination of assets always returning zero in every state of the world.*

This assumption can be made without loss of generality, since the problem of how to allocate between redundant assets is trivial.

Bad forecasts, even when input into an excellent optimizer, should not be expected to lead to favorable results. In that spirit, we want to focus on the subset of the portfolios where the model of variances and covariances can be trusted.

Assumption 2. *We require that the covariance matrix* $\boldsymbol{\Sigma}$ *is such that* $\mathbf{h}'\boldsymbol{\Sigma}\mathbf{h}$ *is an admissible forecast of ex ante portfolio variance for all* $\mathbf{h} \in \mathcal{C}$.

If there is a non-trivial vector $\mathbf{v} \neq \mathbf{0}$ such that $\mathbf{\Sigma v} = \mathbf{0}$, then by extension there is a long-short portfolio whose forecasted variance is zero. Existence of such putative null vector would contradict our previous assumption of no redundant assets, and in the absence of redundant assets, fails Assumption 2 (zero is not an admissible forecast of portfolio variance) if $\mathbf{v}$ falls within the feasible set.

The unconstrained case is defined by the MVO problem (2) with $\mathcal{C} = \mathbb{R}^n$, and in this setting, Assumption 2 implies in particular that $\mathbf{\Sigma}^{-1}$ exists, and moreover, that the optimization problem is strictly convex with the unique solution given by the well-known formula

$$\mathbf{h}^* := (\lambda \mathbf{\Sigma})^{-1} \boldsymbol{\mu}\,. \tag{5}$$

Thus, in a world satisfying Assumption 2, unconstrained MVO is both as simple, and as complicated, as understanding the inverse of a *symmetric positive-definite* (SPD) matrix.

Suppose an investor has a fixed capital to allocate to a universe comprising risky assets, with a covariance matrix $\mathbf{\Sigma}$, and a riskless asset. Then the risk aversion parameter λ controls the investor's leverage. This leverage can be thought of as being subject to the preference of the investor (or that of their creditor), but in practice, certain levels of leverage may be deemed irrational. If leverage is excessively high, there is a high probability that an investor repeatedly following MVO will experience ruin (i.e., their portfolio value becoming negative) at some point. Perhaps for this reason, Markowitz himself suggested that the maximum point of leverage is the one that maximizes the expectation of the log of wealth, consistent with the Kelly betting criterion.[3] Jacobs et al. [2013] and Jacobs et al. [2014] propose an extension to MVO that allows investors to determine the right amount of portfolio leverage, given their trade-off between portfolio risk, return, and leverage risk.[4]

UNIVERSALITY

In the 75 years since Markowitz' pioneering work, mean-variance approaches have appeared in almost every subfield of quantitative finance. Almgren et al. [1999] introduce the optimal order execution problem, treating the investor's positions at different intraday time-slices as independent sources of risk and applying a mean-variance framework to determine the trade trajectory. The dynamic (multi-period) portfolio choice model of Gârleanu et al. [2013] and Gârleanu et al. [2016] and its extensions [Kolm et al., 2015; Moallemi et al., 2017; Ma et al., 2019; Collin-Dufresne et al., 2020; Baldacci et al., 2021] employ mean-variance and mean-quadratic variation formulation, resembling optimal order liquidation but incorporating additional complexities such as time-varying return

[3] John Mulvey, Princeton University, personal communication, March 2024.

[4] Edirisinghe et al. [2023] expand this approach to include price impact from trade execution. In the related work by Fan et al. [2012], the authors demonstrate that, under gross-exposure constraints, the sensitivity of the mean-variance form of the utility function to estimation errors can easily be bounded.

predictions ("alpha") and nonlinear price impact. Multi-period portfolio optimization models were quite familiar to Markowitz, who published several articles on the topic [Markowitz et al., 2011; Blay et al., 2020]. Methods of dynamic option hedging in the presence of transaction costs can be thought of as minimizing the mean loss due to transaction costs, while simultaneously minimizing variance introduced by imperfect replication [Almgren et al., 2016; Bank et al., 2017; Aliaga-Diaz et al., 2020; Kolm et al., 2019; Du et al., 2020; Fu et al., 2024]. Other examples include high-frequency trading with limit orders [Cartea et al., 2014], mean-reversion strategies subject to nonlinear price impact [Ritter, 2017], and market making [El Aoud et al., 2015; Bergault et al., 2021], which once again bear similarity to Almgren et al. [1999] insofar as the objective is mean-variance across intraday time-slices.

Therefore, it is of great interest to understand why the mean-variance paradigm is so universal. The best such explanation available at the time of writing seems to be that, under certain assumptions on the distribution of asset returns that shall be detailed below, the optimal allocation is the mean-variance allocation for any "reasonable" utility function. The framework of modeling investor preferences via utility functions is very well established, with seminal contributions due to Arrow [1971] and Pratt [1964]. Indeed the existence of a utility function can be derived from much more basic assumptions concerning the investor's preference relation, as we discuss next.

From Preferences to Utility Functions

The theory of portfolio selection can be viewed within the broader framework of decision-making under uncertainty. Selecting a portfolio is a special case of choosing from among a class of "uncertain prospects." One can think of probabilities as objective or subjective. In the approach of von Neumann et al. [1947], probabilities are objective and there is given some set Z whose elements are are referred to as rewards or prizes. The choice set is effectively $X = \mathcal{P}$, a suitable space of probability distributions over Z, so when the agent chooses an uncertain prospect $x \in X$ they effectively lock themselves into a certain probability distribution p over rewards. In the application to portfolio choice, elements of Z correspond to possible values for the investor's final wealth at some future horizon, and the probability distribution of asset returns is assumed to be objective and known to the agent.

The key result we need from the theory of rational choice is: for $Z \subset \mathbb{R}$ a bounded interval, the three axioms of Preference, Substitution, and Continuity (as formulated in the appendix) are equivalent to the existence of a continuous function $u : Z \to \mathbb{R}$ such that the investor always ranks lotteries according to the value of $\mathbb{E}[u(z)]$, where the expectation is computed for each lottery with respect to the lottery's given probabilities. This holds both for continuous and discrete distributions, under appropriate mathematical conditions on the space of distributions as detailed in the appendix. To aid the reader in navigating the extensive literature, the appendix provides the precise set of axioms that give the desired expected-utility representation for preference relations over probability measures on a bounded interval. Such measures represent the probability distributions over the investor's final wealth relevant for portfolio optimization problems.

An investor whose choices satisfy the axioms of Preference, Substitution, and Continuity is frequently called a "rational investor." Informally, every rational investor has a utility function and behaves as if maximizing expected utility. This does not require that the investor does so *knowingly*, nor that they are able to articulate said utility function.

The domain of the utility function represents the investor's wealth level. It could of course be applied to the wealth in a single account, while the investor has other accounts. We denote the initial wealth level in the account by w_0, and the wealth one period ahead, in the absence of transaction costs, by

$$w_T = w_0 + \mathbf{h}'\mathbf{r}\,, \tag{6}$$

where $\mathbf{h} \in \mathbb{R}^n$ are the holdings. We assume the investor has access to effectively unlimited financing, allowing for negative wealth levels. Thus, the domain of the utility function is $\mathbb{R}$, encompassing both positive and negative values.

Assumption 3. *We refer to a utility function $u : \mathbb{R} \to \mathbb{R}$ as* standard *if it is increasing, strictly concave, and continuously differentiable. Throughout this article, we assume all utility functions are standard, unless explicitly noted otherwise.*

The properties required by Assumption 3 make economic sense. A utility function must be increasing. Even the most philanthropic charity would never optimize in such a way as to intentionally cause losses in the investment portfolio of its endowment. Concavity is equivalent to risk-aversion. If u were allowed to be discontinuous, any point of discontinuity would represent a particular level of wealth which is perceived by the investor as very different from nearby wealth levels, even if the difference were by one penny. This is not the case for most rational investors, and is nonsensical for investments in financial markets as portfolio values constantly fluctuate.

Mean-Variance Equivalent Distributions

A quadratic function can never satisfy Assumption 3 and so for all economically reasonable purposes, "quadratic utility" is a misnomer. Nevertheless, the function we are truly interested in, namely the *expected utility of* $\mathbf{h}$

$$\hat{u}(\mathbf{h}) := \mathbb{E}[u(w_0 + \mathbf{h}'\mathbf{r})] = \int u\,(w_0 + \mathbf{h}'\mathbf{r})\,p(\mathbf{r})d\mathbf{r}\,, \tag{7}$$

may represent an integral that is not available in closed form due to the high dimension n, the nonlinearity of u, and the complicated structure of the probability density function $p(\mathbf{r})$. Therefore, it would be convenient if we could somehow construct a quadratic function of $\mathbf{h}$ whose maximum is known to be located at the precise vector $\mathbf{h}^*$ which maximizes the expected utility (7). The associated quadratic function, if it exists, is not a utility function, but is instead referred to as a *mean-variance form of the utility* (or simply "mean-variance form" for short). The existence of a mean-variance form having the same solution as the original, often intractable, problem sounds almost too good to be true, but we must remember that due to strong concavity of u, maximizing the expected utility (7) is almost a convex optimization problem, and we can exploit a certain symmetry in $p(\mathbf{r})$ along with convexity.

Levy et al. [1979] show the effectiveness of what might be called large-E, low-V solutions, where E is the expected wealth and V is the variance of wealth. Later work of Kroll et al. [1984] focus on empirical studies in which the goal is to identify some function f, if such function exists, such that maximizing $f(E, V)$ yields almost as favorable results as maximizing expected utility of wealth. Markowitz [1991] asserts

> *"Of the various approximations tried in Levy-Markowitz the one which did best, almost without exception, was essentially that suggested in Markowitz [1959], namely*
>
> $$f(E, V) = u(E) + 0.5\, u''(E) V \,. \qquad \text{"}$$

This choice of $f(E, V)$ is of course the Taylor series of the expected utility (7) to second order [Markowitz, 2014].

In empirical work such as Kroll et al. [1984], we must concede some possibility that these results could be an artifact of the particular dataset that was chosen for the study. Nevertheless, their work inspires the question: when is expected utility a function of mean and variance, and what can we say concerning the shape of such function?

Definition 1. *For two scalars $E, V \in \mathbb{R}$, where $V > 0$, let $L(E, V)$ denote the space of allocations under which*

$$\mathbb{E}[w_T] = E \text{ and } \mathbb{V}[w_T] = V \,. \tag{8}$$

We say expected utility is a function of mean and variance *if there exists some function $f(E, V)$ such that*

$$\mathbb{E}[u(w_T)] = f(E, V) \tag{9}$$

for all allocations in $L(E, V)$.

Definition 2. *For n-dimensional random vectors possessing continuous probability density functions (pdfs), an* elliptical distribution *is one whose pdf on $\mathbb{R}^n$ takes the form*

$$f_n(\mathbf{x}) = |\mathbf{\Omega}|^{-1/2} g_n\left((\mathbf{x} - \boldsymbol{\mu})' \mathbf{\Omega}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right) \,, \tag{10}$$

where $\boldsymbol{\mu}$ is the median vector, $\boldsymbol{\Omega}$ is a positive definite matrix, $|\boldsymbol{\Omega}|$ denotes the determinant of $\boldsymbol{\Omega}$, and $g_n : \mathbb{R}^+ \to \mathbb{R}^+$ is a generator function (that does not depend on $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$).

We remark that $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Omega}$ is proportional to the covariance matrix, if they exist. For the multivariate normal distribution, the most well-known distribution within the elliptical family, the generator function is

$$g_n(s) = (2\pi)^{-n/2} \exp(-s/2)\,. \tag{11}$$

The elliptical class also encompasses several non-normal distributions, such as the multivariate Student-t, symmetric multivariate stable distribution, symmetric multivariate Laplace, and symmetric general hyperbolic distributions. These distributions can exhibit heavy tails. For example, the multivariate Student-t distribution with ν degrees of freedom has density of the form (10) with

$$g_n(s) \propto (\nu + s)^{-(n+\nu)/2}\,. \tag{12}$$

When $\nu = 1$, we obtain the multivariate Cauchy distribution, although not all of the remarks below apply to the Cauchy distribution due to the non-existence of moments of order one and greater.

Proposition 1. *If the distribution of $\mathbf{r}$ is elliptical, and if u is a standard utility function, then expected utility is a function of mean and variance, and*

$$\frac{\partial}{\partial E} f(E, V) \geq 0 \quad \textit{and} \quad \frac{\partial}{\partial V} f(E, V) \leq 0. \tag{13}$$

Proposition 1 is due to Chamberlain [1983], whose interest in this problem, and finance more broadly, was sparked by Michael Rothschild [Graham et al., 2023]. The form of this proposition is proven in Ingersoll [1987, App. B of Ch. 4]. Under the conditions of Proposition 1, let E be the level of the expected return at the optimal expected-utility solution. Because $\partial f(E, V)/\partial V \leq 0$, we know that V must be the minimum variance for the given expected return. In words, if expected utility is a function of mean and variance (9), and if furthermore, the function f is increasing in E for all V and decreasing in V for any E, then we can find the optimal solution by MVO.

However, this result makes strong assumptions, and its conclusion is stronger than we need. Next, we define a weaker condition that we call *mean-variance equivalence*. This condition is merely a statement about the locations of the respective optima, rather than a broad statement about the expected utility function associated with any possible portfolio.

Definition 3. *The underlying asset return distribution, $p(\mathbf{r})$, is said to be* mean-variance equivalent *(MVE) if, for any standard utility function u at any initial wealth level w_0,* *there exists some constant $\kappa > 0$, such that*

$$\mathbf{h}^* := \operatorname{argmax}_{\mathbf{h}} \mathbb{E}[u(w_T)] = \operatorname{argmax}_{\mathbf{h}} \left(\mathbb{E}[w_T] - \frac{\kappa}{2} \mathbb{V}[w_T] \right) , \tag{14}$$

where $w_T = w_0 + \mathbf{h}'\mathbf{r}$.

We note that when expected utility is a function of the mean and variance, then the underlying asset return distribution is also MVE. We contend that a rational expected-utility-maximizer cares more about Definition 3 than that of Definition 1. Particularly, a utility-maximizer cares about their ability to find the exact solution to the expected utility problem by solving an equivalent MVO problem, rather than concerning themselves with the multitude of portfolios that represent sub-optimal expected-utility allocations. The risk-aversion parameter κ typically depends on w_0 and on the shape of the function u, but in the situation of Definition 3, other nuances of u are not needed to find the optimum. In order to apply the MVO formulation in (14), the investor does not need to know everything about their utility function. They only need to know enough to determine the appropriate κ given their current wealth. In practice, determining κ entails solving the MVO problem for all values of κ in some range, and selecting the optimal portfolio that is compatible with the investor's risk tolerance.

Then, which distributions are mean-variance equivalent? Tobin [1958] conjectures that for $n > 2$, *any* two-parameter distribution is mean-variance equivalent. Feldstein [1969] provides a counterexample to Tobin's idea. Many distributions, including heavy-tailed distributions, elliptical distributions and a subset of distributions exhibiting skew, are also mean-variance equivalent, but not all two-parameter families.

To motivate the following, let us consider a random vector $\mathbf{r}$ of one period asset returns from a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Of course, we know that for this distribution, expected utility is a function of mean and variance. We denote by Z the random variable that is the return of the Markowitz portfolio, i.e.

$$Z := \frac{\mathbf{r}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}} . \tag{15}$$

It is easy to see that the conditional expected return of $\mathbf{r}$ given Z is

$$\mathbb{E}[\mathbf{r}|Z] = Z\boldsymbol{\mu} . \tag{16}$$

Now, if we let $\boldsymbol{\varepsilon}$ be the residual random vector of returns

$$\boldsymbol{\varepsilon} := \mathbf{r} - Z\boldsymbol{\mu} , \tag{17}$$

Exhibit 1
A Mean-Variance Equivalent Distribution

Note: Level curves from the two-dimensional mean-variance equivalent distribution $p(x,y) \propto \left(x^2 \exp(-x^2/2) + \frac{(1-\exp\left(-x^2/2\right))^2}{x^3}\right) \cdot \exp(-y^2/2)$.

then the return of the corresponding Markowitz portfolio is zero,

$$\frac{\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}} = 0\,. \tag{18}$$

Moreover, we observe that all components of the residual return vector are independent of the return of the Markowitz portfolio. This has an interpretation in terms of portfolios. While it is always the case that a portfolio $\mathbf{h}$ whose returns are uncorrelated with those the Markowitz portfolio, has an expected return of zero (since $\mathbf{h}'\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \mathbf{h}'\boldsymbol{\mu} = 0$), we see that for the normal distribution such a portfolio must have zero expected return *conditional on the return of the Markowitz portfolio*.

Remarkably, it turns out that the existence of a stochastic representation of similar to the form above *fully* characterizes mean-variance equivalent distributions as per Definition 4. We formulate this new result next, and refer to the appendix for a formal proof.

We consider distributions of n-dimensional random vectors $\mathbf{r}$ with the stochastic representation

$$\mathbf{r} = \boldsymbol{\mu} Z + \boldsymbol{\varepsilon}\,, \tag{19}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$, Z is a scalar random variable, and $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is a random vector such that

$$\mathbb{E}[Z] = 1\,, \tag{20}$$

$$\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = 0\,, \tag{21}$$

$$\mathbb{E}[\boldsymbol{\varepsilon} \mid Z] = \mathbf{0}\,, \tag{22}$$

$$\mathbb{V}[\boldsymbol{\varepsilon}] = \boldsymbol{\Sigma} - \mathbb{V}[Z]\boldsymbol{\mu}\boldsymbol{\mu}'\,, \tag{23}$$

where $\boldsymbol{\Sigma}$ is an n-by-n SPD matrix.

We note that condition (23) is normalization to ensure that the covariance of $\mathbf{r}$ is $\boldsymbol{\Sigma}$.

Proposition 2. *A distribution is mean-variance equivalent if and only if it has the stochastic representation* (19) *that satisfies the conditions* (20)–(23)*.*

To the best of our knowledge, this proposition is the first to characterize mean-variance equivalence as per Definition 3. This new result motivates the following definition.

Definition 4. *Distributions with the stochastic representation* (19) *are said to be* mean-variance equivalent distributions*.*

Exhibit 1 presents an example of a MVE-distribution. In a forthcoming article [Benveniste et al., 2024], we examine a number of properties of MVE-distributions and present several implications for portfolio optimization. It is important to highlight that the main result in the recent work by Schuhmacher et al. [2021] and Auer et al. [2023] differs significantly from Proposition 2 above. In particular, they establish that the return distribution of portfolios that *include the risk-free asset* is determined by its mean and variance (as per Definition 1) *if and only if* asset returns follow a *skew-elliptical generalized location and scale* (SEGLS) distribution. Proposition 2 is more general and establishes mean-variance equivalence (as per Definition 3) for portfolios of any assets that are MVE-distributed.

Clearly, the above results disprove the prevailing belief that MVO cannot or should not be applied in asymmetric cases. In particular, these new findings shed fresh light on the universality of MVO, beyond that of Chamberlain's characterization (symmetric distributions) in Proposition 1. The reader may have noticed that the vector $\mathbf{r}$ could represent returns of different assets on the same time interval, or returns of the same asset over different time intervals, or a combination of the two. As long as the other assumptions of the proposition are not broken, it immediately extends to the discrete time versions of the many multi-period problems discussed earlier.

This completes a chain of ideas leading to Markowitz' famous result and its universality:

1. the axioms of Preference, Substitution, and Continuity lead to expected utility via rational choice theory by Proposition 3, and

2. expected utility leads to MVO for all MVE distributions by Proposition 2.

MYTHS AND FACTS

In order to be wrong, a statement must first be formulated precisely enough that it becomes falsifiable. Rigor and clarity are so important, that research neglecting them commands a lower status than ideas which were duly formulated, studied, and falsified by the scientific method or by mathematical proofs. Rejecting some hypotheses while finding evidence for others is part of the scientific method. Generating vague or ambiguous theories is not. Several of the memes that we shall unmask as myths in the present section, such as "mean-variance optimization is error maximizing" are, paraphrasing physicist Wolfgang Pauli, not *even* wrong [Peierls, 1960]. Our contribution will be to first formulate such myths rigorously enough that they can be falsified, and then, to falsify them or to ascertain some grain of truth.

Myth 1. *MVO requires normally-distributed asset returns. MVO cannot or should not be applied in cases where return distributions are asymmetric.*

Proposition 1 debunks a prevalent misconception regarding MVO, which falsely links it to the assumption of a normal distribution for asset returns. At most, one could argue that an MVO user is implicitly assuming asset returns are MVE-distributed (Definition 4), a large class of asymmetric distributions which includes all elliptical families (which in turn are symmetric distributions) and some skew families.

Myth 2. *MVO is an error maximizer. MVO is not robust.*

Any optimization method can only be as good as its objective function. Assumption 2 demands that when we ask our optimizers to minimize predicted variance, the predictions we are using simply must be good predictions, or at least economically reasonable. If we consider variance forecasts of the form $\mathbf{h}'\mathbf{\Sigma}\mathbf{h}$ as in (4), then not all SPD matrices $\mathbf{\Sigma}$ are admissible. Specifically, not all SPD matrices can satisfy the more stringent requirement of always providing reasonable variance forecasts for all portfolios satisfying the constraints. For instance, when n is comparable to the size of the stock market, the widely recognized historical sample covariance estimator is rarely admissible for optimization, as we will illustrate below. This does not imply that MVO itself is not robust. It only implies the inadmissibility of the historical sample covariance as an estimator, for the purposes of computing the optimal portfolio.

Consider covariance matrix estimation as a statistical procedure which has $n(n+1)/2$ free parameters corresponding to the independent elements in a symmetric matrix. With k historical periods, we have nk data points, hence $2k/(n+1)$ data points per parameter. Thus $k = n + 1$ is the threshold to have two data points per parameter, implying an extreme loss of precision in the parameter estimation procedure. Practically, with $n = 3000$ we need about twelve years of daily returns to get two data points per parameter!

We can gain a clearer understanding of the inadmissibility of the sample covariance estimator by examining the rank of the matrix. Let $\mathbf{R}$ be a $k \times n$ matrix having the stock return time series as columns. De-mean each column of $\mathbf{R}$ and rescale by $1/(k-1)$ so that the sample covariance matrix is $\mathbf{R}'\mathbf{R}$. Due to its inadmissibility as an estimator for the portfolio optimization problem, we shall refer to $\mathbf{R}'\mathbf{R}$ as the *naive* estimator. We have

$$\operatorname{rank}(\mathbf{R}'\mathbf{R}) = \operatorname{rank}(\mathbf{R}) \leq \min(n, k)\,, \tag{24}$$

so if $k < n$ it is impossible that $\mathbf{R}'\mathbf{R}$ or $\mathbf{\Sigma}$ is invertible. Again with $n = 3000$, and assuming a market with 250 trading days per year, we need twelve years to get $k = 3000$ and hence to cross the threshold where $\mathbf{\Sigma}^{-1}$ begins to exist. However, sample covariance matrices based on k slightly larger than n continue to fail Assumption 2 due to their condition number, as is shown below.

By our above assumptions, $\mathbf{\Sigma}$ is an n-by-n SPD matrix, thus allowing us to express it as

$$\mathbf{\Sigma} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}' = \sum_{i=1}^{n} \lambda_i\, \mathbf{q}_i \mathbf{q}_i'\,, \tag{25}$$

where $\mathbf{\Lambda} := \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix of real non-negative eigenvalues, and the columns of $\mathbf{Q}$ form an orthonormal set of eigenvectors. The order is not important, so without loss of generality we may assume $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$.

If $\lambda_n = 0$, then eigenvector $\mathbf{q}_n$ is a long-short portfolio with predicted variance equal to zero. In the absence of redundant assets, such a prediction is not economically reasonable and fails Assumption 2. Henceforth, we consider the case where $\lambda_n > 0$, ensuring the existence of $\mathbf{\Sigma}^{-1} = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}'$ and, therefore, the existence of the unconstrained Markowitz portfolio. The eigenvectors of $\mathbf{\Sigma}^{-1}$ are the same as the eigenvectors of $\mathbf{\Sigma}$, while the eigenvalues of $\mathbf{\Sigma}^{-1}$ are $1/\lambda_i$ for all i. The largest eigenvalue of $\mathbf{\Sigma}^{-1}$ is $1/\lambda_n$, the reciprocal of the smallest eigenvalue of $\mathbf{\Sigma}$. The condition number of $\mathbf{\Sigma}^{-1}$ is

$$\operatorname{cond}(\mathbf{\Sigma}^{-1}) = \lambda_1/\lambda_n\,. \tag{26}$$

The condition number of a matrix is always at least one, but a matrix is said to be *ill-conditioned* if the condition number is much larger than one. Informally, as a common rule of thumb used by statisticians in regression analysis, if the condition number is greater than thirty, then multicollinearity and identifiability become important concerns, and parameter inference is called into question. As discussed above, if $k < n$ then the naive estimator $\mathbf{R}'\mathbf{R}$ is not invertible, but once we cross the threshold of $k > n$, it tends to happen that the smallest eigenvalue λ_n crosses from 0 to a tiny number, while λ_1 remains bounded away from zero. Hence the condition number λ_1/λ_n crosses over from ∞ to an astronomically large number as k crosses from just below n to just over n (cf. Laloux et al.

[1999] and Fan et al. [2008]).

The $\mathbf{q}_i$'s are the normalized principal component portfolios, so that $\|\mathbf{q}_i\| = 1$ for all i. The condition number (26) is the ratio of the largest and smallest principal component portfolio variances. What does Assumption 2 tell us about this ratio? Let us frame the question in trading language. Let us consider two portfolios, A and B, that are uncorrelated with each other. Both portfolios have the sum of their squared holdings equal to one. The trader asks the risk department to forecast the variance of each, and the risk department forecasts variance λ_A, λ_B respectively, with $\lambda_B < \lambda_A$. How high can the ratio λ_A/λ_B become, before the trader responds incredulously to the risk department that their forecasts cannot possibly make economic sense, regardless of the actual portfolios' contents? This would represent the maximum economically reasonable condition number, according to this trader. More generally, any trader should be suspicious that a so-called "perfect hedge" will always continue to work out as such in the future, and hence should demand of their risk departments to use forecasting methods that do not predict perfect hedges among companies that are truly distinct businesses. Therefore, the practical implication of Assumption 2 is that the condition number should, at very least, be bounded by the maximum λ_A/λ_B ratio deemed reasonable by any conservative, risk-averse trader.

In summary, matrices with high condition number fail Assumption 2 and for n the size of the equity market, the naive historical estimator will almost always have high condition number. The procedure of MVO is a robust procedure, but the naive historical covariance estimator is inadmissible as an estimator for the purposes of optimization, and should not be used. Error maximization could occur, for example, if λ_1/λ_n were huge, and if the expected return vector $\boldsymbol{\mu}$ has a nonzero component in the direction $\mathbf{q}_n$, then that component would get multiplied by the huge factor $1/\lambda_n$ when computing $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ (see also the discussions in Best et al. [1991] and Palczewski et al. [2014]). However, this form of error maximization would only arise if we attempt to apply MVO to a covariance matrix estimator that Assumption 2 would rule out.

If the historical covariance $\mathbf{R}'\mathbf{R}$ is a naive estimator for optimization, what then is a proper, admissible estimator? Fortunately, there are readily available covariance matrix estimators which are admissible for MVO in the form of factor models. Let r_i denote the ex ante excess return of the i-th asset in the market, as before. Instead of allowing very complex forms for the multivariate distribution $p(\mathbf{r})$, representations based on arbitrage pricing theory (APT) [Ross, 1976; Roll et al., 1980] constrain the distribution's complexity by assuming a linear structure such that the return in excess of the risk-free rate of the i-th asset is given by

$$r_i = x_{i,1}f_1 + x_{i,2}f_2 + \cdots + x_{i,p}f_p + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)\,, \tag{27}$$

where $f_j, j = 1, \ldots, p$, are *factors*, $x_{i,j}$ denote the j-th *factor loading* of the i-th asset, and

ε_i and σ_i are the *residual return* (or *idiosyncratic return*) and *idiosyncratic volatility* of the i-th asset, respectively. We assume (27) represents a *strict factor model*, such that the residual returns are mutually uncorrelated with each other and across time, as well as uncorrelated with all factors.

Jointly estimating all loadings $x_{i,j}$ *and* factors f_j is challenging, so typically, one assumes that one of them is known and calculates the other through regression. In what follows, we focus on the Rosenberg approach [Rosenberg, 1974] where the loadings $x_{i,j}$ are taken to be known (or exogenous), and the factors f_j are hidden (also known as latent) variables. If we denote by $\mathbf{f} := (f_1, \ldots, f_p)'$ the random factor process, then we can succinctly express model (27) as

$$\mathbf{r} = \mathbf{X}\mathbf{f} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{D})\,, \tag{28}$$

where $\mathbf{X}$ is the matrix with the loadings $x_{i,j}$ as elements, and

$$\mathbf{D} := \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)\,, \tag{29}$$

is the diagonal matrix of the variances of the residual returns.

Since one cannot directly observe the $\mathbf{f}$-process, we must obtain information about it through statistical inference. We assume that the $\mathbf{f}$-process has finite first and second moments given by

$$\boldsymbol{\mu}_{\mathbf{f}} := \mathbb{E}[\mathbf{f}] \in \mathbb{R}^p, \quad \text{and} \quad \mathbf{F} := \mathbb{V}[\mathbf{f}] \in \mathbb{R}^{p\times p}\,. \tag{30}$$

The model (28)–(30) entails associated reductions of the first and second moments of the asset returns

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\mu}_{\mathbf{f}}, \quad \text{and} \quad \boldsymbol{\Sigma} = \mathbf{D} + \mathbf{X}\mathbf{F}\mathbf{X}'\,. \tag{31}$$

In the usual case $p \ll n$, this entails a substantial dimension reduction over the naive model in which all of the $n(n+1)/2$ elements of $\boldsymbol{\Sigma}$ are treated as independent parameters. Indeed, the two moments in (30) together with the volatilities from (29) entail a total of $n + p + p(p+1)/2$ parameters. In an industrial example, the Barra USE4 model is of the above form, and uses $p = 72$ in the form of twelve style factors and sixty industry factors. Hence, with $n = 3,000$ the naive model entails

$$n(n+1)/2 = 4,501,500 \text{ parameters}\,, \tag{32}$$

while the $p = 72$ model entails

$$n + p + p(p+1)/2 = 5,700 \text{ parameters}\,. \tag{33}$$

Equation (31) implies that $\boldsymbol{\Sigma}$ is in a diagonal plus low-rank form, which is quite useful for portfolio construction and analyzing existing portfolios. For example, for a portfolio with dollar holdings $\mathbf{h} \in \mathbb{R}^n$, from (31) we have

$$\mathbf{h}'\mathbf{\Sigma}\mathbf{h} = \mathbf{h}'\mathbf{D}\mathbf{h} + \mathbf{h}'\mathbf{X}\mathbf{F}\mathbf{X}'\mathbf{h}\,, \tag{34}$$

which expresses the portfolio's variance using the *idiosyncratic variance* $\mathbf{h}'\mathbf{D}\mathbf{h} = \sum_{i=1}^{n} h_i^2 \sigma_i^2$, and a second term which depends on the portfolio through the portfolio exposure vector $\mathbf{X}'\mathbf{h}$.

We illustrate how the Woodbury matrix-inversion lemma enables us to solve the unconstrained MVO problem with computational complexity that scales linearly with the number of assets. First, let us recall the Woodbury matrix-inversion lemma

$$(\mathbf{A} + \mathbf{U}\mathbf{C}\mathbf{V})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{C}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U}\right)^{-1}\mathbf{V}\mathbf{A}^{-1}\,, \tag{35}$$

where $\mathbf{A}$, $\mathbf{U}$, $\mathbf{C}$ and $\mathbf{V}$ denote matrices of the correct (conformable) dimensions. The lemma holds when all of the matrix inverses in (35) exist. Now, by applying (35) with $\mathbf{A} := \mathbf{D}$, $\mathbf{U} := \mathbf{X}$, $\mathbf{V} := \mathbf{X}'$ and $\mathbf{C} := \mathbf{F}$, we obtain

$$\mathbf{\Sigma}^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{X}\left(\mathbf{F}^{-1} + \mathbf{X}'\mathbf{D}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{D}^{-1}\,. \tag{36}$$

The right hand side of (36) is computationally efficient. Note that $\mathbf{D}^{-1}$ merely involves n scalar reciprocals of the idiosyncratic variances σ_i^2, and hence calculating $\mathbf{D}^{-1}$ is $O(n)$ and not $O(n^3)$ like standard matrix inversion algorithms. The remaining inverses in (36) all involve $p \times p$ matrices where $p \ll n$. Notably, Jacobs et al. [2005] examine the *critical line algorithm* (CLA) [Markowitz, 1987] for long-short portfolios under linear constraints. They demonstrate that the CLA is a fast algorithm when the covariance matrix is of the form (31).

We can use (36) to obtain a strong lower bound on eigenvalues of $\mathbf{\Sigma}^{-1}$, and hence an upper bound on the condition number. Hence, one of the main issues with covariance matrices and optimization – namely, the existence of portfolios with very low forecasted volatility – does not occur when the covariance matrix comes from a reasonably-constructed APT model. In some sense, factor models are the perfect complement to Markowitz optimization.

One of the key assumptions of the Rosenberg approach is that the factor loadings $x_{i,j}$ represent exposures to *common sources of risk*. For example, many stocks have exposure (either positive or negative) to the prices of energy commodities because they are net producers or consumers of energy, so one could argue that the price volatility of a bundle of energy represents a common source of risk. In contrast to this, top stock analysts conduct in-depth analyses of the companies they cover, often making predictions that cannot be attributed to any common factor. In so doing, they are effectively predicting the next realization of the residual ε_i from (27). More generally, one could take any expected return prediction vector $\boldsymbol{\mu}$ and remove the component spanned by $\boldsymbol{X}$, thus forming

$$\boldsymbol{\alpha} := \mathbf{P}_{\mathbf{X}^\perp}\boldsymbol{\mu} := (\mathbf{I} - \mathbf{X}\mathbf{X}^+)\boldsymbol{\mu}. \tag{37}$$

where $\mathbf{X}^+$ is the pseudo-inverse and $\mathbf{P}_{\mathbf{X}^\perp}$ denotes the orthogonal projection onto the kernel of $\mathbf{X}'$. Form an augmented matrix $\mathbf{X}_\alpha$, which has $\boldsymbol{\alpha}$ as the first column, and $\mathbf{X}$ in the remaining columns. If we regress a cross section $\mathbf{r}$ of realized asset returns on $\mathbf{X}_\alpha$, the coefficient of the $\boldsymbol{\alpha}$ column can be interpreted as the returns on a specific MVO portfolio, so MVO portfolio returns are equivalent to alpha factor returns in cross-sectional asset return models [Ritter, 2016; Raponi et al., 2023]. This shows that in a special case, MVO is simply computing a multivariate regression. Consequently, any instabilities are the same and of course, statisticians are comfortable with the instabilities in regression, and know how to control them.

Myth 3. *MVO fails to outperform a $1/n$ portfolio out-of-sample.*

It is hardly surprising that this and the previous myth have sparked significant controversy and intrigue among practitioners and academics alike. The $1/n$ portfolio "debate" is extensively covered in a number of articles, so we will keep our discussion brief, focusing primarily on the main ideas.

In this battle of ideas, we playfully dub the opposing sides as "team-$1/n$" and "team-MVO," respectively. On the team-$1/n$ side of the debate, several studies suggest that MVO fails to outperform basic portfolio rules such as equal weights in out-of-sample tests (see, for example, Jobson et al. [1981], Jorion [1985], and DeMiguel et al. [2009]). In contrast, on team-MVO's side, research such as Kinlaw et al. [2017, Ch. 7], Kritzman et al. [2010] and Allen et al. [2019] disprove these findings.

So how can the findings of these two teams be so different? We argue that the majority of team-$1/n$ studies arrive at their conclusions primarily due to the reliance on using sample estimates of historical returns and covariances as inputs in MVO. As we saw earlier, the sample covariance estimator does not satisfy Assumption 2. Moreover, historical return-based estimates are notoriously inadequate predictors of expected returns [Merton, 1980; Campbell et al., 2008]. Rather, it is crucial to incorporate anomalies, style and risk factors, and other sources of information beyond historical returns when estimating expected return [Kelly et al., 2013; Pettenuzzo et al., 2014; Gu et al., 2020]. In contrast to those of team-$1/n$, studies by team-MVO demonstrate that investors endowed with *some forecasting ability* benefit from using MVO in constructing their portfolios, and outperform $1/n$ portfolios and other basic portfolio rules.

To implement MVO in practice, portfolio managers need an admissible covariance matrix and reasonable estimates of expected returns. Interestingly, Markowitz himself did not address this issue. In his seminal article, Markowitz [1952] begins with the declaration

> *"The process of selecting a portfolio may be divided into two stages. The first stage starts with observation and experience and ends with beliefs about the future performances of available securities. The second stage starts with the relevant beliefs about future performances and ends with the choice of portfolio. This paper is concerned with the second stage."*

However, on page 91 of the same article, Markowitz writes

> *"...we must have procedures for finding reasonable μ_i and σ_{ij}. These procedures, I believe, should combine statistical techniques and the judgment of practical men."*
>
> *"One suggestion as to tentative μ_i, σ_{ij} is to use the observed μ_i, σ_{ij} for some period of the past. I believe that better methods, which take into account more information, can be found."*

Markowitz advice is clear: construct forecasts that incorporate information beyond historical samples of returns. When queried about how to construct these, frequently his retort was: *"That's your job, not mine."* [Sexauer et al., 2024] However, in his foreword to Jacobs et al., 2000, he recounts how he and his colleagues developed the Daiwa Portfolio Optimization System in 1990, leveraging expected return estimation techniques based on Jacobs et al., 1988.[5]

Beyond their simplicity and explainability, $1/n$ portfolios present several practical issues. First, they employ a "one-size-fits-all" approach, ignoring investors' risk tolerances. Second, rebalancing such portfolios involves selling winners and buying losers, a strategy that may yield profits in mean-reverting markets but can lead to losses in trending environments. Third, a $1/n$ portfolio fails to account for liquidity and trading frictions, such as price impact and cost to borrow.

Under what circumstances would a $1/n$ portfolio actually be optimal? We mention two situations. From equation (5), it is evident that when the expected returns are given by $\boldsymbol{\mu}_{1/n} := \frac{\lambda}{n}\boldsymbol{\Sigma}\mathbf{e}$, where $\mathbf{e} := (1, \ldots, 1)' \in \mathbb{R}^n$, the $1/n$ portfolio is optimal. This is quite a special case that is unlikely in practice. Another case where the $1/n$ portfolio emerges as the optimal strategy is for an investor who (i) is faced with ambiguous asset return distributions and, (ii) adopts a worst-case approach considering all measures within an ambiguity set. That said, such an investor's preferences represent a most pessimistic view of the world, one which cannot be formulated by standard utility functions that satisfy Assumption 3.

Myth 4. *Quadratic utility does not reflect investor preferences.*

[5] The model's inputs are drawn from the anomalies literature of its time (see, Jacobs et al. [1988] for a comprehensive overview), using cross-sectional characteristics such as earnings-to-price and book-to-price [Fama et al., 1992]. We refer the reader to Guerard Jr. et al. [1991], Miller et al. [1992], and Bloch et al. [1993] for details.

As discussed above, the properties required by Assumption 3 make economic sense, at least for utility functions of wealth that describe choices among different portfolios of financial assets. We shall not consider utility theory as applied to problems with multiple conflicting objectives, life-and-death choices or other such extreme examples. A quadratic function does not satisfy Assumption 3 and so for all economically reasonable purposes, "quadratic utility" is a misnomer as already stated above. Nevertheless, the function we are truly interested in, namely the expected utility (7), often has its maximum located at the same portfolio as a certain mean-variance form. The mean-variance form is not itself a utility function in any economic sense.

Myth 5. *MVO does not handle signals at different horizons, or transaction costs.*

Baldacci et al. [2021] study the multi-horizon version of the mean-variance objective, which is known as *mean-quadratic-variation (MQV).* The authors show that for any initial portfolio $\mathbf{h}_0 \in \mathbb{R}^n$, there is a unique solution to the optimization problem

$$\max_{\mathbf{h}\in\mathcal{A};\ \mathbf{h}(0)=\mathbf{h}_0} \mathbb{E} \int_0^\infty L(\mathbf{h}_t, \dot{\mathbf{h}}_t)\, dt\,, \tag{38}$$

with

$$L(\mathbf{h}_t, \dot{\mathbf{h}}_t) := \mathbf{h}_t' \boldsymbol{\mu}_t - \frac{1}{2}\dot{\mathbf{h}}_t' \boldsymbol{\Lambda} \dot{\mathbf{h}}_t - \frac{\kappa}{2}\mathbf{h}_t' \boldsymbol{\Sigma} \mathbf{h}_t\,, \tag{39}$$

where $\mathcal{A}$ is a large space containing admissible stochastic processes representing a diverse range of trading strategies[6], and a dot above a variable denotes the time derivative $\dot{\mathbf{h}}_t := \frac{d}{dt}\mathbf{h}_t$. The first two terms in (39) represent the expected portfolio return net of cost, while the third term is the quadratic variation. The above formulation generalizes the work of Gârleanu et al. [2016], who also use a mean-quadratic-variation approach to generalize their earlier discrete time model [Gârleanu et al., 2013] to continuous time.

We emphasize that the maximization in (38) is not simply over static, predetermined trading trajectories, but over the substantially larger class of all admissible stochastic processes. A wide variety of optimal trading and optimal execution models with alphas and/or price impact are special cases of this model. For example, the continuous-time model of Almgren et al. [1999] is the special case with zero alpha. The model of Gârleanu et al. [2016] is a special case where the return-predicting factors follow a Markovian jump diffusion.

Baldacci et al. [2021] show that the unique solution to (38) satisfies the stochastically-forced ODE system

$$\dot{\mathbf{h}}_t = -\boldsymbol{\Gamma}\mathbf{h}_t + \mathbf{b}_t\,, \tag{40}$$

[6] Mathematically, the stochastic processes they consider are square integrable and adapted to the filtration $\mathcal{F} = \{\mathcal{F}_t : t \geq 0\}$, representing the available information at time t. This space encompasses a wide range of processes, including non-smooth and non-Markovian ones.

where

$$\boldsymbol{\Gamma} := (\kappa \boldsymbol{\Lambda}^{-1} \boldsymbol{\Sigma})^{1/2}\,, \tag{41}$$

$$\mathbf{b}_t := \int_t^\infty \exp\big(\boldsymbol{\Gamma}(t-s)\big)\boldsymbol{\Lambda}^{-1}\mathbb{E}_t[\boldsymbol{\mu}_s]\,ds\,. \tag{42}$$

Given that the model (38)–(39) maximizes a mean-quadratic-variation objective, it should not be surprising that the process $\mathbf{b}_t$ from (42) is related in a straightforward way to an integral of expected future Markowitz portfolios

$$\boldsymbol{\Gamma}^{-1}\mathbf{b}_t = \boldsymbol{\Gamma}\int_t^\infty \exp\big(\boldsymbol{\Gamma}(t-s)\big)\,\mathbb{E}_t[\mathbf{M}_s]\,ds\,, \tag{43}$$

$$\text{where}\ \ \mathbf{M}_s := (\kappa\boldsymbol{\Sigma})^{-1}\boldsymbol{\mu}_s\,. \tag{44}$$

Following Gârleanu et al. [2016], we refer to (43) as the *aim portfolio*, because by (40), at optimality, the trading process $\dot{\mathbf{h}}_t$ is aiming towards it. Hence the optimal trade aims towards an integral of future expected Markowitz portfolios. With fond recollections of Markowitz's work on multi-period portfolio models [Markowitz et al., 2011; Blay et al., 2020], we hope that, if he were still here, he would look upon equations (43) and (44) with approval.

Myth 6. *MVO is not suitable for intraday and high-frequency trading.*

We distinguish between *ultra-high-frequency trading* (UHFT) and everything else. There is a class of trading which attempts to render markets more efficient by removing pure arbitrages created by simple physics such as the speed of light between cities. As such, UHFT strategies do not represent risk-taking in the usual sense as described by modern portfolio theory. Hardware or communication problem aside, such trading is effectively riskless. While important to ensure markets continue to function smoothly and efficiently, this form of trading is a niche.

Once we look beyond the niche of UHFT latency arbitrage, we enter the realm of strategies that are intraday, with holding periods of minutes to hours, or longer. For strategies at these slightly longer timescales, we can use the mean-quadratic-variation model (38) to determine a continuous-time trading path by solving (40). A given strategy can then use individual orders, which are necessarily discrete, to minimize deviations from the continuous-time optimal path.

The model of Baldacci et al. [2021] admits an an explicit form for the *value function*, namely the following quadratic value function

$$V_s(\mathbf{p}) := \sup_{\mathbf{h}\in\mathcal{A};\ \mathbf{h}(s)=\mathbf{p}} \mathbb{E}\int_s^\infty L(\mathbf{h}_t, \dot{\mathbf{h}}_t)\,dt = -\frac{1}{2}\mathbf{p}'\boldsymbol{\Lambda}\boldsymbol{\Gamma}\mathbf{p} + \mathbf{p}'\mathbf{b}_s + C_s\,, \tag{45}$$

where C_s is a constant. Economically, the number $V_s(\mathbf{p})$ represents the value to a rational

MQV-maximizer at time s of the arbitrary portfolio $\mathbf{p}$. This directly leads to a practically implementable algorithm for applying continuous-time MQV in a trading strategy. If trading were truly possible in continuous time, we would simply trade to $\mathbf{h}_t$ that solves (40). For simplicity, let us denote the current time by $t = 0$ in the following. Since real trading is discrete, we choose an interval of length τ minutes, over which the next wave of child orders will execute (say, 30 minutes). We observe that

$$\max \mathbb{E} \int_0^\infty L(\mathbf{h}_t, \dot{\mathbf{h}}_t)\, dt = \max \left(\mathbb{E} \int_0^\tau L(\mathbf{h}_t, \dot{\mathbf{h}}_t)\, dt + \mathbb{E} \int_\tau^\infty L(\mathbf{h}_t, \dot{\mathbf{h}}_t)\, dt \right). \tag{46}$$

Since any sub-part of an optimal path is optimal with respect to its endpoints, we can replace the second term in (46) with $V_\tau(\mathbf{h}_t)$ for optimal paths. Let $\boldsymbol{\delta}$ be the list of orders we intend to trade over the first τ minutes. The first term of (46) then contains our expected costs, short-term alpha, and risk for the trade list $\boldsymbol{\delta}$, while the second term is the value function $V_\tau(\mathbf{h}_0 + \boldsymbol{\delta})$. Therefore, we obtain the trade list by solving

$$\max_{\boldsymbol{\delta} \in \mathbb{R}^n} \left(f(\boldsymbol{\delta}) + \mathbb{E}_{t=0}[V_\tau(\mathbf{h}_0 + \boldsymbol{\delta})] \right), \tag{47}$$

where $f(\boldsymbol{\delta})$ is an appropriate representation of the expected cost, alpha, and risk over the short interval $[0, \tau]$. The number of decision variables in (47) is equal to the number of assets, which makes this formulation significantly more attractive than standard approaches that work by discretizing time into many sub-intervals and suffering from a corresponding "blowup" in the number of decision variables. Moreover, in simpler models such as (45), the second term in (47) is quadratic and easy to optimize. Colloquially, we refer to this formulation as the *"from here to eternity"* method since the second term in (47) extends from our current target to infinity.

CONCLUSION

Since Markowitz's pioneering work on mean-variance optimization (MVO), mean-variance approaches have permeated nearly every facet of quantitative finance. In this article, we argued that their widespread adoption can be attributed to the universality of the mean-variance paradigm, wherein the maximum expected utility and mean-variance allocations coincide for a broad range of distributional assumptions of asset returns.

We introduced a formal definition of mean-variance equivalence and provided a novel and comprehensive characterization of distributions, termed mean-variance-equivalent (MVE) distributions, wherein expected utility maximization and the solution of an MVO problem are the same.

In the second part of the article, we shifted focus toward common myths associated with MVO. The myths we debunked include the misconception that MVO necessitates normally-distributed asset returns, the belief that it is unsuitable for cases with asymmetric return distributions, the notion that it maximizes errors, and the perception that it underperforms a simple $1/n$ portfolio in out-of-sample tests. In addition, we addressed misunderstandings regarding MVO's ability to handle signals across different time horizons, its treatment of transaction costs, its applicability to intraday and high-frequency trading, and whether quadratic utility accurately represents investor preferences. We provided a contemporary and practically implementable prescription for how to use value functions to solve multi-period optimization with a mean-quadratic-variation objective.

While Harry Markowitz will forever be remembered for his contributions to financial economics and investment management, he was first and foremost a philosopher. Descartes and Hume were two of his favorite philosophers. In an interview, he shared

> *"(But) the greatest philosopher ever, in my mind, is Aristotle. He spoke of eudaimonia, which means living a good life, so that people think well of you after you are dead." [Powell, 2023]*

In conclusion, we continue to be deeply inspired by Harry Markowitz's contributions, which have left an indelible mark on us and countless others. We believe that by prioritizing the cultivation of virtues, the pursuit of meaningful goals, and the promotion of well-being, we can lead lives full of fulfillment and purpose, echoing his remarkable legacy.

ACKNOWLEDGEMENTS

We are grateful to Kenneth Blay, Bruce Jacobs, Stephen Lihn, John Mulvey, and several anonymous readers for their insightful comments and suggestions on an earlier version of this article.

MATHEMATICAL PROOFS

In this appendix, we provide the mathematical details of the article.

Representing Preferences through Utility Functions

This section serves to justify the assertions made in the main text, concerning rational investors and the existence of a utility function. Our favorite reference for this is Kreps [1988]; see also Jonathan Levin's Stanford lectures.[7] In the theory of rational choice, with or without uncertainty, a decision-maker (or agent) must decide between elements of a set X. The agent's choices are modeled by means of a relation $\succ$, where $x \succ y$ means that the agent definitely prefers x to y. The relation is called a *preference relation* if it satisfies $x \succ y \implies y \not\succ x$ (asymmetry) and $x \succ z \implies \forall y, x \succ y$ or $y \succ z$ (negative transitivity). If X is countable, then a binary relation $\succ$ is a preference relation if and only if there is a function $\tilde{p} : X \to \mathbb{R}$, which we call a *preference function* to distinguish it from the utility functions to come later, such that

$$x \succ y \iff \tilde{p}(x) > \tilde{p}(y). \tag{48}$$

Also, if X is any subset of a separable metric space, and $\succ$ is a continuous preference relation, then once again there exists a function $\tilde{p} : X \to \mathbb{R}$ such that (48) holds [Kreps, 1988, Prop. 3.3, Thm. 3.7].

This is about as abstract and as general as one can get; the elements of X could represent almost any conceivable kind of choice. To apply abstract rational choice theory to decision-making under uncertainty, we need to introduce probabilities in some way, and to think of an element $x \in X$ as an "uncertain prospect." One can think of probabilities as objective or subjective. In the approach of von Neumann et al. [1947], probabilities are objective and there is given some set Z whose elements are referred to as rewards or prizes. The choice set is effectively $X = \mathcal{P}$, a suitable space of probability distributions over Z, so when the agent chooses an uncertain prospect $x \in X$, they effectively lock themselves into a specific probability distribution p over the rewards. In the application to portfolio choice, elements of Z correspond to possible values for the investor's final wealth at some future horizon, and the probability distribution of asset returns is assumed to be objective and known to the agent. We shall discuss models in which Z is finite, or models in which Z is a continuous subset of $\mathbb{R}$, with the finite case being simpler mathematically.

First assume $Z \subset \mathbb{R}$ is finite, and $\mathcal{P}$ is the space of all probability distributions on Z, which is just the standard simplex. Consider the following three axioms (5.1 - 5.3 from Kreps [1988]):

Axiom 1 (Preference)**.** *The binary relation $\succ$ is a preference relation.*

[7] Available at: `https://web.stanford.edu/~jdlevin/teaching.html`.

Axiom 2 (Substitution). *For all $p, q, r \in \mathcal{P}$ and $a \in [0, 1)$,*

$$p \succ q \implies ap + (1-a)r \succ aq + (1-a)r\,, \tag{49}$$

Axiom 3 (Continuity). *For all $p, q, r \in \mathcal{P}$, if $p \succ q \succ r$, then there exist $a, b \in (0, 1)$ such that*

$$ap + (1-a)r \succ q \succ bp + (1-b)r\,. \tag{50}$$

The von Neumann-Morgenstern (NM) theorem states that a binary relation on $\mathcal{P}$ satisfies Preference, Substitution, and Continuity if and only if there exists a function $u : Z \to \mathbb{R}$ such that

$$p \succ q \iff \mathbb{E}_p[u(z)] > \mathbb{E}_q[u(z)]\,, \tag{51}$$

where $\mathbb{E}_p[u(z)]$ is defined as $\sum_{z \in Z} p(z)u(z)$. Hence, while Axiom 1 implies the existence of a preference function, the NM-theorem identifies the preference function as expected utility of the reward.

To represent wealth levels, we are particularly interested in the case where $Z \subset \mathbb{R}$ is a bounded, closed interval which we consider to contain all practical portfolio selection problems. Let $\mathcal{P}$ be the space of all Borel probability measures on Z. Full mathematical details addressing general probability measures can be found in Fishburn [1979, Ch. 10]. However, for bounded intervals, the problem simplifies significantly. We only need to replace Axiom 3 with weak continuity, a notion which is covered in any standard graduate-level probability course.

Axiom 4 (Weak Continuity). *Relation $\succ$ is continuous in the weak topology.*

The version of the continuous NM-theorem that is arguably the most relevant to modeling of portfolio selection is the following.

Proposition 3. *Let $Z = [-M, N] \subset \mathbb{R}$ be a bounded, closed interval and $\mathcal{P}$ the space of Borel probability measures on Z. A binary relation $\succ$ on $\mathcal{P}$ satisfies Axiom 1, Axiom 2, Axiom 4 if and only if there exists a continuous function $u : Z \to \mathbb{R}$ such that*

$$p \succ q \iff \mathbb{E}_p[u(z)] > \mathbb{E}_q[u(z)]. \tag{52}$$

We emphasize that Proposition 3 is a direct consequence of the theorems in Fishburn [1979, Ch. 10] and Kreps [1988, Cor. 5.22].

In summary, it is perhaps not surprising that expected-utility theory is, paraphrasing Jonathan Levin, "the workhorse of modern economics." The role played by expected-utility as the preference-function for uncertain prospects is equivalent to Preference, Substitution, and Continuity in both continuous and discrete contexts.

Proof of Proposition 2

Before proving the proposition, as a sanity check, we verify that elliptical distributions satisfy the conditions of the proposition. Let $\mathbf{r}$ be an n-dimensional elliptically distributed random variable. Note that the class of MVE distributions, characterized by (19) and (20)–(23), is closed under changes of variables given by invertible linear transformations. Without loss of generality, then, we can assume that $\boldsymbol{\mu} = \mathbf{e}_1$ (the first standard basis vector) and $\boldsymbol{\Sigma} = \mathbf{I}$. Then

$$\mathbf{r} = \mathbf{e}_1 + R\mathbf{U}\,, \tag{53}$$

where $\mathbf{U}$ is uniformly distributed on the sphere S^{n-1}, and R is a positive random variable independent of $\mathbf{U}$. Setting

$$Z := \mathbf{e}_1'\mathbf{r}\,, \tag{54}$$

$$\boldsymbol{\varepsilon} := \mathbf{r} - Z\mathbf{e}_1\,, \tag{55}$$

we clearly have $\mathbb{E}[Z] = 1$ and $\mathbf{e}_1'\boldsymbol{\varepsilon} = 0$, so conditions (20) and (21) are satisfied. To verify condition (22), by writing $\mathbf{U} = (U_1, \ldots, U_n)'$, we have that

$$Z = 1 + RU_1\,, \tag{56}$$

$$\boldsymbol{\varepsilon} = (0, RU_2, \ldots, RU_n)'\,. \tag{57}$$

The conclusion now follows from

$$\mathbb{E}[U_j|Z] = \mathbb{E}[U_j|U_1] = 0\,, \tag{58}$$

for $j = 2, \ldots, n$.

Now, we turn to the proof of Proposition 2. In the following, we denote by π the probability distribution of random vectors $\mathbf{r}$ that satisfy (19)–(22) of the proposition.

Proof. ("if"): Let $\mathbf{h}^* := \operatorname{argmax}_{\mathbf{h}} \mathbb{E}\left[u(\mathbf{h}'\mathbf{r})\right]$. We can decompose $\mathbf{h}^*$ into a component along the Markowitz portfolio and an orthogonal component $\mathbf{q}$ (in the inner product given by $\boldsymbol{\Sigma}$) such that

$$\mathbf{h}^* = c\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \mathbf{q}, \tag{59}$$

with $\mathbf{q}'\boldsymbol{\mu} = 0$. Then, we have

$$\mathbb{E}[u(\mathbf{h}'\mathbf{r})] = \mathbb{E}\left[u\left(\left(c\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \mathbf{q}\right)'(\boldsymbol{\mu}Z + \boldsymbol{\varepsilon})\right)\right] \tag{60}$$

$$= \mathbb{E}\left[u\left(c\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}Z + \mathbf{q}'\boldsymbol{\varepsilon}\right)\right] \tag{61}$$

$$= \mathbb{E}\left[\mathbb{E}\left[u\left(c\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}Z + \mathbf{q}'\boldsymbol{\varepsilon}\right)\middle|Z\right]\right] \tag{62}$$

$$\leq \mathbb{E}\left[u\left(\mathbb{E}\left[c\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}Z + \mathbf{q}'\boldsymbol{\varepsilon}\middle|Z\right]\right)\right] \tag{63}$$

$$= \mathbb{E}\left[u\left(c\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}Z\right)\right] \tag{64}$$

$$= \mathbb{E}\left[u\left((c\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})'\mathbf{r}\right)\right]\,, \tag{65}$$

where we used Jensen's inequality in (63). This shows that the expected utility of wealth of the portfolio $\mathbf{h}^*$ is at most that of the mean-variance portfolio, as desired.

("only if"): Let $\mathbf{h}^* := \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$. By assumption, for every utility function u there is a scalar c such that

$$c\,\mathbf{h}^* = \mathrm{argmax}_{\mathbf{h}}\,\mathbb{E}\left[u(\mathbf{h}'\mathbf{r})\right]\,. \tag{66}$$

The first-order conditions of (66) imply

$$\int \mathbf{r}u'\big(c\,(\mathbf{h}^*)'\mathbf{r}\big)d\pi(\mathbf{r}) = \mathbf{0}\,. \tag{67}$$

Applying this to the family of utility functions $-\exp(-x) + bx$, $b > 0$, we observe that for every b, there is a c such that

$$\int \mathbf{r}\exp\big(-c\,(\mathbf{h}^*)'\mathbf{r}\big)d\pi(\mathbf{r}) + b\boldsymbol{\mu} = \mathbf{0}\,. \tag{68}$$

By varying b, it follows that for uncountably many c, the vector

$$\int \mathbf{r}\exp\big(-c\,(\mathbf{h}^*)'\mathbf{r}\big)d\pi(\mathbf{r}) \tag{69}$$

is proportional to $\boldsymbol{\mu}$. In other words, for uncountably many c, for any $\mathbf{h}$ such that $\mathbf{h}'\boldsymbol{\mu} = 0$, we obtain

$$f_{\mathbf{h}}(c) := \int \mathbf{h}'\mathbf{r}\exp\big(-c\,(\mathbf{h}^*)'\mathbf{r}\big)d\pi(\mathbf{r}) = 0\,. \tag{70}$$

$f_{\mathbf{h}}(c)$ is an analytic function in c. Therefore, by the identity theorem for analytic functions, this implies that $f_{\mathbf{h}}(c) = 0$ for all c in its domain in $\mathbb{C}$, which includes the strip $\{-\delta < \mathrm{Re}(c) < \delta\}$, $\delta > 0$. In particular, this implies that for every $t \in \mathbb{R}$, and for any $\mathbf{h}$ such that $\mathbf{h}'\boldsymbol{\mu} = 0$,

$$\int \mathbf{h}'\mathbf{r}\exp\big(it(\mathbf{h}^*)'\mathbf{r}\big)d\pi(\mathbf{r}) = 0\,. \tag{71}$$

Therefore, by standard arguments, we have

$$\mathbb{E}[\mathbf{h}'\mathbf{r}|(\mathbf{h}^*)'\mathbf{r}] = 0\,. \tag{72}$$

Let us define

$$Z := \frac{(\mathbf{h}^*)'\mathbf{r}}{\mathbb{E}\left[(\mathbf{h}^*)'\mathbf{r}\right]}\,, \tag{73}$$

$$\boldsymbol{\varepsilon} := \mathbf{r} - \boldsymbol{\mu} Z\,, \tag{74}$$

$$\mathbf{h}_{\boldsymbol{\alpha}} := \boldsymbol{\alpha} - \frac{\boldsymbol{\alpha}'\boldsymbol{\mu}}{(\mathbf{h}^*)'\boldsymbol{\mu}}\mathbf{h}^*\,, \tag{75}$$

where $\boldsymbol{\alpha}$ is an arbitrary vector. Then, $(\mathbf{h}^*)'\boldsymbol{\varepsilon} = 0$ and $\boldsymbol{\alpha}'\boldsymbol{\varepsilon} = \mathbf{h}_{\boldsymbol{\alpha}}'\mathbf{r}$. Since $(\mathbf{h}_{\boldsymbol{\alpha}})'\boldsymbol{\mu} = 0$, we see that

$$\mathbb{E}[\boldsymbol{\alpha}'\mathbf{r}|Z] = 0\,. \tag{76}$$

Therefore, as (76) is true for any vector $\boldsymbol{\alpha}$, we obtain

$$\mathbb{E}[\boldsymbol{\varepsilon}|Z] = \mathbf{0}\,, \tag{77}$$

which completes the proof. $\square$

REFERENCES

Aliaga-Diaz, R., G. Renzi-Ricci, A. Daga, and H. Ahluwalia (2020). "Portfolio Optimization with Active, Passive, and Factors: Removing the Ad Hoc Step". In: *The Journal of Portfolio Management* 46.4, pp. 39–51.

Allen, D., C. Lizieri, and S. Satchell (2019). "In Defense of Portfolio Optimization: What If We Can Forecast?" In: *Financial Analysts Journal* 75.3, pp. 20–38.

Almgren, R. and N. Chriss (1999). "Value under Liquidation". In: *Risk* 12.12, pp. 61–63.

Almgren, R. and T. M. Li (2016). "Option Hedging with Smooth Market Impact". In: *Market Microstructure and Liquidity* 2.01, p. 1650002.

Arrow, K. J. (1971). *Essays in the Theory of Risk-Bearing.* North-Holland Amsterdam.

Auer, B. R., F. Schuhmacher, and H. Kohrs (2023). "Rehabilitating Mean-Variance Portfolio Selection: Theory and Evidence". In: *The Journal of Portfolio Management.*

Baldacci, B., J. Benveniste, and G. Ritter (2021). "Optimal Turnover, Liquidity, and Autocorrelation". In: *arXiv preprint arXiv:2110.03810.*

Bank, P., H. M. Soner, and M. Voß (2017). "Hedging with Temporary Price Impact". In: *Mathematics and Financial Economics* 11, pp. 215–239.

Benveniste, J., P. N. Kolm, and G. Ritter (2024). "Mean-Variance Equivalent Distributions and Expected Utility Maximization". In: *In preparation.*

Bergault, P., D. Evangelista, O. Guéant, and D. Vieira (2021). "Closed-Form Approximations in Multi-Asset Market Making". In: *Applied Mathematical Finance* 28.2, pp. 101–142.

Bernstein, P. L. (1993). *Capital Ideas: The Improbable Origins of Modern Wall Street.* Simon and Schuster.

Best, M. J. and R. R. Grauer (1991). "On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results". In: *The Review of Financial Studies* 4.2, pp. 315–342.

Blay, K., A. Gosh, S. Kusiak, H. Markowitz, N. Savoulides, and Q. Zheng (2020). "Multiperiod Portfolio Selection: A Practical Simulation-Based Framework". In: *Journal of Investment Management* 18.4, pp. 94–129.

Bloch, M., J. Guerard, H. Markowitz, P. Todd, and G. Xu (1993). "A Comparison of Some Aspects of the US and Japanese Equity Markets". In: *Japan and the World Economy* 5.1, pp. 3–26.

Blume, M. E. (1972). "On the Assessment of Risk". In: *Journal of Finance* 26.1.

Campbell, J. Y. and S. B. Thompson (2008). "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" In: *The Review of Financial Studies* 21.4, pp. 1509–1531.

Cartea, Á., S. Jaimungal, and J. Ricci (2014). "Buy Low, Sell High: A High Frequency Trading Perspective". In: *SIAM Journal on Financial Mathematics* 5.1, pp. 415–444.

Chamberlain, G. (1983). “A Characterization of the Distributions That Imply Mean-Variance Utility Functions”. In: *Journal of Economic Theory* 29.1, pp. 185–201.

Collin-Dufresne, P., K. Daniel, and M. Sağlam (2020). “Liquidity Regimes and Optimal Dynamic Asset Allocation”. In: *Journal of Financial Economics* 136.2, pp. 379–406.

Connor, G., L. R. Goldberg, and R. A. Korajczyk (2010). *Portfolio Risk Analysis*. Princeton University Press.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). “Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?” In: *The Review of Financial Studies* 22.5, pp. 1915–1953.

Du, J., M. Jin, P. N. Kolm, G. Ritter, Y. Wang, and B. Zhang (2020). “Deep Reinforcement Learning for Option Replication and Hedging”. In: *The Journal of Financial Data Science* 2.4, pp. 44–57.

Edirisinghe, C., J. Chen, and J. Jeong (2023). “Optimal Leveraged Portfolio Selection Under Quasi-Elastic Market Impact”. In: *Operations Research* 71.5, pp. 1558–1576.

El Aoud, S. and F. Abergel (2015). “A Stochastic Control Approach to Option Market Making”. In: *Market Microstructure and Liquidity* 1.01, p. 1550006.

Fabozzi, F. J., P. N. Kolm, D. A. Pachamanova, and S. M. Focardi (2007). *Robust Portfolio Optimization and Management*. John Wiley & Sons.

Fama, E. F. and K. R. French (1992). “The Cross-Section of Expected Stock Returns”. In: *The Journal of Finance* 47.2, pp. 427–465.

— (2020). “Comparing Cross-Section and Time-Series Factor Models”. In: *The Review of Financial Studies* 33.5, pp. 1891–1926.

Fan, J., Y. Fan, and J. Lv (2008). “High Dimensional Covariance Matrix Estimation Using a Factor Model”. In: *Journal of Econometrics* 147.1, pp. 186–197.

Fan, J., J. Zhang, and K. Yu (2012). “Vast Portfolio Selection with Gross-Exposure Constraints”. In: *Journal of the American Statistical Association* 107.498, pp. 592–606.

Feldstein, M. S. (1969). “Mean-Variance Analysis in the Theory of Liquidity Preference and Portfolio Selection”. In: *The Review of Economic Studies* 36.1, pp. 5–12.

Fishburn, P. C. (1979). *Utility Theory for Decision Making*. Krieger.

Fu, H., B. Hientzsch, P. N. Kolm, J. Pan, and S. Xu (2024). “Dynamic Hedging of Multi-Stock Options under Price Impact: A Deep Learning Approach”. In: *Working Paper*.

Gârleanu, N. and L. H. Pedersen (2013). “Dynamic Trading with Predictable Returns and Transaction Costs”. In: *The Journal of Finance* 68.6, pp. 2309–2340.

— (2016). “Dynamic Portfolio Choice with Frictions”. In: *Journal of Economic Theory* 165, pp. 487–516.

Graham, B., K. Hirano, and G. Imbens (2023). “The ET Interview: Professor Gary Chamberlain”. In: *Econometric Theory* 39.1, pp. 1–26.

Gu, S., B. Kelly, and D. Xiu (2020). "Empirical Asset Pricing via Machine Learning". In: *The Review of Financial Studies* 33.5, pp. 2223–2273.

Guerard Jr., J. B. and M. Takano (1991). "Stock Selection and Composite Modeling in Japan". In: *Security Analysts Journal* 29, pp. 1–13.

Ingersoll, J. E. (1987). *Theory of financial decision making*. Vol. 3. Rowman & Littlefield.

Jacobs, B. I. and K. N. Levy (1988). "Disentangling Equity Return Regularities: New Insights and Investment Opportunities". In: *Financial Analysts Journal* 44.3, pp. 18–43.

— (2000). *Equity Management: Quantitative Analysis for Stock Selection*. First Edition. McGraw-Hill.

— (2013). "Leverage Aversion, Efficient Frontiers, and the Efficient Region". In: *The Journal of Portfolio Management* 39.3, pp. 54–64.

— (2014). "Traditional Optimization is Not Optimal for Leverage-Averse Investors". In: *The Journal of Portfolio Management* 40.2, pp. 30–40.

— (2021). "Factor Modeling: The Benefits of Disentangling Cross-Sectionally for Explaining Stock Returns". In: *Journal of Portfolio Management* 47.6, pp. 33–50.

Jacobs, B. I., K. N. Levy, and H. M. Markowitz (2005). "Portfolio Optimization with Factors, Scenarios, and Realistic Short Positions". In: *Operations Research* 53.4, pp. 586–599.

Jobson, J. D. and R. M. Korkie (1981). "Putting Markowitz Theory to Work". In: *The Journal of Portfolio Management* 7.4, pp. 70–74.

Jorion, P. (1985). "International portfolio diversification with estimation risk". In: *Journal of Business*, pp. 259–278.

Kelly, B. and S. Pruitt (2013). "Market Expectations in the Cross-Section of Present Values". In: *The Journal of Finance* 68.5, pp. 1721–1756.

King, B. F. (1966). "Market and Industry Factors in Stock Price Behavior". In: *The Journal of Business* 39.1, pp. 139–190.

Kinlaw, W., M. P. Kritzman, and D. Turkington (2017). *A Practitioner's Guide to Asset Allocation*. John Wiley & Sons.

Kolm, P. N. and G. Ritter (2015). "Multiperiod Portfolio Selection and Bayesian Dynamic Models". In: *Risk* 28.3, pp. 50–54.

— (2019). "Dynamic Replication and Hedging: A Reinforcement Learning Approach". In: *The Journal of Financial Data Science* 1.1, pp. 159–171.

Kolm, P. N., R. Tütüncü, and F. J. Fabozzi (2014). "60 Years of Portfolio Optimization: Practical Challenges and Current Trends". In: *European Journal of Operational Research* 234.2, pp. 356–371.

Kreps, D. (1988). *Notes on the Theory of Choice*. Westview Press.

Kritzman, M., S. Page, and D. Turkington (2010). "In Defense of Optimization: The Fallacy of 1/N". In: *Financial Analysts Journal* 66.2, pp. 31–39.

Kroll, Y., H. Levy, and H. M. Markowitz (1984). "Mean-Variance versus Direct Utility Maximization". In: *The Journal of Finance* 39.1, pp. 47–61.

Laloux, L., P. Cizeau, J.-P. Bouchaud, and M. Potters (1999). "Noise Dressing of Financial Correlation Matrices". In: *Physical Review Letters* 83.7, p. 1467.

Levy, H. and H. M. Markowitz (1979). "Approximating Expected Utility by Function of Mean and Variance". In: *The American Economic Review* 69.3, pp. 308–317.

Ma, G., C. C. Siu, and S.-P. Zhu (2019). "Dynamic Portfolio Choice with Return Predictability and Transaction Costs". In: *European Journal of Operational Research* 278.3, pp. 976–988.

Markowitz, H. M. (1952). "Portfolio Selection". In: *The Journal of Finance* 7.1, pp. 77–91.

— (1959). *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press.

— (1987). *Mean-Variance Analysis in Portfolio Choice and Capital Markets*. Cambridge, MA: Basil Blackwell.

— (1990). *Foundations of Portfolio Theory: Nobel Lecture*. URL: `https://www.nobelprize.org/uploads/2018/06/markowitz-lecture.pdf`.

— (1991). "Foundations of Portfolio Theory". In: *The Journal of Finance* 46.2, pp. 469–477.

— (2010). "Portfolio Theory: As I Still See It". In: *Annual Review of Financial Economics* 2.1, pp. 1–23.

— (2014). "Mean-Variance Approximations to Expected Utility". In: *European Journal of Operational Research* 234.2, pp. 346–355.

— (2016). *Risk-Return Analysis: The Theory and Practice Of Rational Investing (Volume Two)*. McGraw-Hill.

— (2020). *Risk-Return Analysis: The Theory and Practice Of Rational Investing (Volume Three)*. McGraw-Hill.

Markowitz, H. M. and K. Blay (2013). *Risk-Return Analysis: The Theory and Practice Of Rational Investing (Volume One)*. McGraw-Hill.

Markowitz, H. M. and E. L. van Dijk (2011). "Single-Period Mean-Variance Analysis in a Changing World". In: *Stochastic Programming: The State of the Art In Honor of George B. Dantzig*, pp. 213–237.

Merton, R. C. (1980). "On Estimating the Expected Return on the Market: An Exploratory Investigation". In: *Journal of Financial Economics* 8.4, pp. 323–361.

Miller, J., J. B. Guerard Jr, and M. Takano (1992). "Bridging the Gap Between Theory and Practice in Equity Selection Modeling: Case Studies of US and Japanese Models". In: *Unpublished Article*.

Moallemi, C. C. and M. Sağlam (2017). "Dynamic Portfolio Choice with Linear Rebalancing Rules". In: *Journal of Financial and Quantitative Analysis* 52.3, pp. 1247–1278.

Palczewski, A. and J. Palczewski (2014). “Theoretical And Empirical Estimates of Mean-Variance Portfolio Sensitivity”. In: *European Journal of Operational Research* 234.2, pp. 402–410.

Peierls, R. E. (1960). “Wolfgang Ernst Pauli, 1900-1958”. In: *Biographical Memoirs of Fellows of the Royal Society* 5, pp. 174–192.

Pettenuzzo, D., A. Timmermann, and R. Valkanov (2014). “Forecasting Stock Returns Under Economic Constraints”. In: *Journal of Financial Economics* 114.3, pp. 517–553.

Powell, R. (2023). *Lessons in Life and Investing from the Late Harry Markowitz*. URL: `https://www.evidenceinvestor.com/lessons-in-life-and-investing-from-harry-markowitz/`.

Pratt, J. W. (1964). “Risk Aversion in the Small and in the Large”. In: *Econometrica: Journal of the Econometric Society*, pp. 122–136.

Raponi, V., R. Uppal, and P. Zaffaroni (2023). “Robust Portfolio Choice”. In: *Available at SSRN 3933063*.

Ritter, G. (2016). “Stable Linear-Time Optimization in Arbitrage Pricing Theory Models”. In: *Risk Magazine*.

— (2017). “Machine Learning for Trading”. In: *Risk* 30.10, pp. 84–89.

Roll, R. and S. A. Ross (1980). “An Empirical Investigation of the Arbitrage Pricing Theory”. In: *The Journal of Finance* 35.5, pp. 1073–1103.

Rosenberg, B. (1974). “Extra-Market Components of Covariance in Security Returns”. In: *Journal of Financial and Quantitative Analysis* 9.2, pp. 263–274.

Ross, S. A. (1976). “The Arbitrage Theory of Capital Asset Pricing”. In: *Journal of Economic Theory* 13.3, pp. 341–360.

Rubinstein, M. (2011). *A History of the Theory of Investments: My Annotated Bibliography*. Vol. 335. John Wiley & Sons.

Schuhmacher, F., H. Kohrs, and B. R. Auer (2021). “Justifying Mean-Variance Portfolio Selection When Asset Returns Are Skewed”. In: *Management Science* 67.12, pp. 7812–7824.

Sexauer, S. C. and L. B. Siegel (2024). “Harry Markowitz and the Philosopher’s Stone”. In: *Financial Analysts Journal* 80.1, pp. 1–11.

Sharpe, W. F. (1963). “A Simplified Model for Portfolio Analysis”. In: *Management Science* 9.2, pp. 277–293.

Tobin, J. (1958). “Liquidity Preference as Behavior Towards Risk”. In: *The Review of Economic Studies* 25.2, pp. 65–86.

von Neumann, J. and O. Morgenstern (1947). *Theory of Games and Economic Behavior*. Princeton University Press.