# Direct Data-based Decision Making under Uncertainty

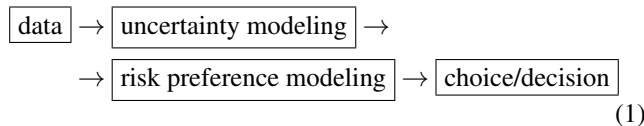Bogdan Grechuk[*]      Michael Zabarankin[†]

## Abstract

In a typical one-period decision making model under uncertainty, unknown consequences are modeled as random variables. However, accurately estimating probability distributions of the involved random variables from historical data is rarely possible. As a result, decisions made may be suboptimal or even unacceptable in the future. Also, an agent may not view data occurred at different time moments, e.g. yesterday and one year ago, as equally probable. The agent may apply a so-called "time" profile (weights) to historical data. To address these issues, an axiomatic framework for decision making based directly on historical time series is presented. It is used for constructing data-based analogues of mean-variance and maxmin utility approaches to optimal portfolio selection.

**Key Words:** time series, decision making under uncertainty, mean-variance analysis, portfolio optimization, utility theory.

## 1  Introduction

A typical process of decision making under uncertainty is as follows

$$\boxed{\text{data}} \rightarrow \boxed{\text{uncertainty modeling}} \rightarrow$$
$$\rightarrow \boxed{\text{risk preference modeling}} \rightarrow \boxed{\text{choice/decision}} \tag{1}$$

Let $\mathcal{X}$ be a set of available (feasible) actions. Scheme (1) can be formally stated as: (i) modeling unknown consequences of every action $X \in \mathcal{X}$ as a random variable (r.v.) $R(X)$, (ii) establishing a numerical representation $U' : \mathcal{R} \to \mathbb{R}$ for agent's preference relation, defined on a space $\mathcal{R}$ of all r.v.'s and (iii) finding best action by maximizing $U'$ with respect to $X \in \mathcal{X}$:

$$\max_{X \in \mathcal{X}} U'(R(X)). \tag{2}$$

What an agent has readily available is only historical/experimental data and his/her preferences towards risk and reward. *The rest is statistical inference from the data about corresponding uncertain outcomes based on various assumptions, which largely depend on the nature of data.* For example, measurements of the length of some object can be reliably assumed to be realizations of independent and identically distributed (i.i.d.) r.v.'s—timing of those measurements can be safely ignored. By the central limit theorem (CLT), the average of a large number of i.i.d. r.v.'s is approximately normally distributed, and consequently, confidence intervals for the true length can be readily obtained.

Merton's well-known model [40] of a financial market assumes that asset prices are stochastic processes with *stationary* and *independent* increments *without jumps*—the only such a stochastic process is a Brownian motion with drift. Cox and Ross [16] argued that asset prices are, in fact, not continuous processes—they may have jumps. Stochastic processes with stationary and independent increments (and with discontinuous sample paths in general), are called Lévy processes [48] and nowadays are widely-used in modeling of financial markets [31, 37, 41]. However, it is commonly acknowledged that

(a) The distributions of rates of return of financial assets are typically *non-symmetric* with left tails being much heavier than right tails [50].

(b) Increments of actual price processes are *not stationary*, and consequently, Lévy processes cannot be calibrated with real data [36].[1]

(c) "Periods of lower returns are systematically followed by compensating periods of higher returns" [51] ("mean reversion" phenomenon)—evidence that price increments are *not independent*.

In fact, the above issues with stochastic processes can be "fixed" by time-series models. For example, autoregressive models $\text{AR}(p)$ assume that asset's rate of return depends on $p$ previous ones, moving-average models $\text{MA}(q)$ involve last $q$ values of a stochastic error, autoregressive-moving-average models $\text{ARMA}(p,q)$ generalize $\text{AR}(p)$ and $\text{MA}(q)$, whereas ARIMA models generalize $\text{ARMA}(p,q)$, suitable to describe a wide range of non-stationary processes [8]. However, any time-series model is merely another inference from the historical data and its parameters are subject to estimation errors.

The discrepancy between a real-life phenomenon and its model is called *model error*—in contrast to approximation error, which can be resolved by simply increasing the sample size, the model error implies that *an increase of observations of asset rates does not directly translate into the accuracy/precision in estimation of the probability distributions of the rates.* There are various existing approaches that address model uncertainty. For example, bootstrapping [6] generates different scenarios for the variable of interest from a given time series, robust optimization [4] assumes that probabilities in question belong to certain intervals, whereas dual characterization of risk and deviation measures [2, 44] relies on risk envelopes, which can be viewed as sets of distortions of an underlying probability measure, see [32]. Notably, Pflug et al. [42]

---

[*]Department of Mathematics, University of Leicester, LE1 7RH, UK; e-mail: bg83@leicester.ac.uk

[†]Department of Mathematical Sciences, Stevens Institute of Technology, Hoboken, NJ 07030, USA; e-mail: mzabaran@stevens.edu

---

[1]Sato processes [47], whose increments are independent but not necessarily stationary, can be used instead.

showed that the naive $1/n$ investment strategy could be optimal in portfolio selection when model uncertainty is high. Savage [49] suggested to study decisions as functions from some state space $\Omega$ to a set of outcomes $\mathcal{Y} \subset \mathbb{R}$, which are now known as *Savage acts*. This approach involves no probability measure on $\Omega$—a critical feature that gave rise to various Savage-act versions of the expected utility theory (EUT) [11, 29]. For example, Gilboa and Schmeidler [21] proposed to study preference relations over acts, i.e., "functions from states of nature into finite-support distributions over a set of deterministic outcomes." In this case, the agent ends up with the same optimization problem (2), where $R$ is a functional from $\mathcal{X}$ to the set $\mathcal{A}$ of all acts, and $U' : \mathcal{A} \to \mathbb{R}$ is a numerical representation of Gilboa & Schmeidler's preference relation. Of course, the list of existing approaches goes far beyond these examples, see e.g. [3, 10, 13, 55] for alternative approaches and [38, 20] for recent surveys.

However, accurately modeling of outcomes of real-life actions in the context of any of these theories is difficult. For example, modeling of financial portfolio returns in terms of Gilboa-Schmeidler acts [21] includes forecasting of a set of finite-support distributions, and therefore, could, in fact, be harder than that in terms of r.v.'s. The main problem with uncertainty modeling is that, contemplating a choice among several alternatives, an agent ponders what alternative he/she would be most benefited from *in the future*, while the only available information is often the data representing their historical performances *in the past*.

*In view of failure of common statistical assumptions in application to a stock market [50, 36, 51] and in view of sensitivity of optimal decisions (portfolios) to errors in estimation of probability distributions of financial assets [28, 30], this work aims to identify intertemporal principles for comparing historical time series of asset rates of return and to develop an axiomatic framework for a rational decision making in portfolio theory on the space of historical time series.* For example, an agent may postulate that if $A$ always outperformed $B$ in the past, then $A \succ B$, even though better past performance does not guarantee better future performance.

The idea of making decisions based directly on historical data is not new,[2] but it has received relatively little attention in economic and financial literature. Gilboa and Schmeidler [22, 23] introduced a case-based decision theory, which makes decisions based on past experience in *similar* situations.[3] In a financial market setting, this theory would identify the moment in the past when the market behavior was most similar to the current one and would prescribe to invest all money into the financial asset which had the highest rate of return in that "similar" situation. However, it is not clear what "similarity measure" to use, and the resulting investment strategy may con-

tradict the diversification principle. There are other objections for the use of direct data-based decision making in portfolio selection:

(i) Information such as recent market trends and news about particular companies may provide valuable insights for selecting a financial portfolio.

(ii) The future may have little in common with the past, for instance, due to unique events such as BREXIT.

(iii) New financial assets lack historical data, but it is unlikely that agents would veiw, say, a new bank and a startup IT company similarly.

However, incorporating news and other non-quantitative information, e.g. a recent hire of a highly regarded CEO, into a mathematical model requires human participation and is, therefore, expensive and slow. In contrast, calibrating stochastic models based only on historical data can be fully automated and performed in milliseconds, which is particularly valuable for high-frequency trading. Thus, *if* the choice of optimal portfolio is based on some uncertainty modeling, which in turn uses historical data only, *then* the uncertainty modeling stage could be omitted, and decisions could be made based on data *directly*.

The contribution and organization of this work are as follows. Section 2 introduces the notion of time profile and discusses numerical representation of time series. Section 3 introduces intertemporal principles of rational choice. Section 4 reinterprets the mean-variance and maxmin utility analyses in the context of direct data-based decision making. Section 5 concludes the work. Appendix A contains proofs of key results in Section 3 and Appendix B provides an axiomatic foundation for a data-based analogue of the EUT.

## 2 Time Profiles and Numerical Representation of Time Series

Let $\mathcal{T} = \{s_1, \ldots, s_T\}$ be a finite set of discrete time moments $s_1 < \cdots < s_T$ in the past, and let $x_1, \ldots, x_T$ be corresponding rates of return of some financial asset. Since $x_1, \ldots, x_T$ encode a *time structure* and are not realizations of i.i.d. r.v.'s, the agent would unlikely view $x_1, \ldots, x_T$ as equally valuable data and may assign them corresponding weights $q_1, \ldots, q_T$ of historical data "*depreciation*" to be collectively referred to as *time profile* $Q$. For example, the agent may postulate that fraction $q_t/q_{t+1}$ is a constant $q \in (0,1]$ independent of $t$, which implies that

$$ q_t = q_T q^{T-t}, \quad t = 1, \ldots, T. \tag{3} $$

Alternatively, $q_1, \ldots, q_T$ can be chosen to be proportional to the (normalized) autocorrelation profile of the asset—if for some time periods (usually far in the past), the autocorrelation vanishes, then those past values play little role in predicting asset's behavior. Suppose, for example, the FTSE 100 index is such an asset. Figure 1 depicts the sample autocorrelation function (ACF) of daily prices of the index from 1-April-2015 to 1-April-2016 with the lag up to 80, taken from [1]. For a lag longer than 80 days, the autocorrelation is negligible, so

---

[2]In psychology, Gorban et al. [24] proposed a neural-network-based approach for making recommendations based on the questionnaire data *directly* that avoids intermediate stage of patient descriptions in terms of "measurements of an individuality."

[3]Gilboa and Schmeidler [22] argued that in a search for a nanny for their child, a couple faces a lot of uncertainty about how each candidate would perform and can hardly define "states of the world" that would adequately model the situation not to mention accurately forecasting of probabilities of each state.

that $T = 80$, and weights $q_1, \ldots, q_{80}$ are then proportional to the ACF in Figure 1 and satisfy $\sum_{t=1}^{T} q_t = 1$.
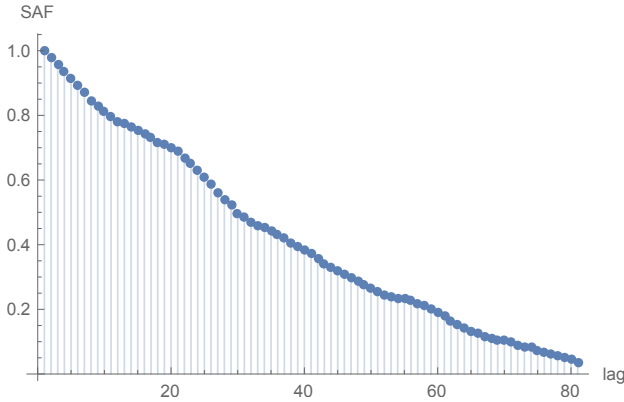


Figure 1: Sample autocorrelation function (ACF) for the FTSE 100 index for the period of 1-April-2015—1-April-2016.

Behavioral evidence supporting the notion of time profile includes, but is not limited to the following

(a) The effect of fading memory and emotions [18, Part 6]: an individual is much more likely to rely on and act upon recent experience rather than that occurred far in the past.

(b) A reversion of the *behavioral time discounting* principle stating that "money available at the present time is worth more than the same amount in the future" [18, Part 3].

(c) Only 21% of agents agree that historical data should be equally weighted [15].

Technical arguments in favor of time profiles include

(a) In time-series analysis, (4) is known as the weighted moving average, in which more weight is often given to the most recent data. In particular, (3) are weights in *exponential smoothing* [9].

(b) The ACF decreases with time and almost vanishes after 80 days (see Figure 1).

(c) In mean-variance portfolio selection, the optimal portfolios with time profiles based on geometric progression (3) with various $q$ outperform optimal portfolio in which asset rates of return are modeled by an ARIMA time-series model — to be discussed in Example 9 (Figure 5).

For a time series $X = (x_1, \ldots, x_T)$ and time profile $Q = (q_1, \ldots, q_T)$, the weighted average and the mean-square deviation of $x_1, \ldots, x_T$ are defined by

$$\mathbb{E}_Q[X] = \sum_{t=1}^{T} q_t x_t, \tag{4}$$

$$\sigma_Q(X) = \sqrt{\sum_{t=1}^{T} q_t (x_t - \mathbb{E}_Q[X])^2} = \sqrt{\mathbb{E}_Q[X^2] - (\mathbb{E}_Q[X])^2}, \tag{5}$$

respectively. Here, $\mathbb{E}_Q[X]$ and $\sigma_Q(X)$ are *not* assumed to be estimates of the future expected value and standard deviation,

respectively, they are just weighted average and standard deviation of the time series $X$. For two time series $X = (x_1, \ldots, x_T)$ and $Y = (y_1, \ldots, y_T)$ and time profile $Q = (q_1, \ldots, q_T)$, the covariance is defined by

$$\text{cov}_Q(X, Y) = \sum_{t=1}^{T} q_t (x_t - \mathbb{E}_Q[X])(y_t - \mathbb{E}_Q[Y]). \tag{6}$$

In fact, the agent may contemplate a whole set $\mathcal{Q}$ of various time profiles. First, the agent may postulate that $q_t$ are non-negative, impose normalization $\sum_{t=1}^{T} q_t = 1$ and define the set of time profiles to be

$$\mathcal{Q} = \mathcal{Q}^{\max}$$
$$= \left\{ (q_1, \ldots, q_T) \in \mathbb{R}^T \,\middle|\, q_1 \geqslant 0, \ldots, q_T \geqslant 0, \ \sum_{t=1}^{T} q_t = 1 \right\}. \tag{7}$$

Next, the agent may assume that $q_1 \leqslant \ldots \leqslant q_T$ for every $Q \in \mathcal{Q}$—more recent data is more valuable. A maximal "time averse" subset $\mathcal{Q} \subset \mathcal{Q}^{\max}$ is given by

$$\mathcal{Q} = \left\{ (q_1, \ldots, q_T) \in \mathbb{R}^T \,\middle|\, 0 \leqslant q_1 \leqslant \ldots \leqslant q_T, \ \sum_{t=1}^{T} q_t = 1 \right\}. \tag{8}$$

Particular cases of (8) include

$$\mathcal{Q}_q = \Big\{ (q_1, \ldots, q_T) \in \mathbb{R}^T \,\Big|\, 0 \leqslant q_1 \leqslant \ldots \leqslant q_T \leqslant q, \\ \sum_{t=1}^{T} q_t = 1 \Big\}, \ q \in [1/T, 1], \tag{9}$$

which represents the preferences of the "time averse" agent who, in addition, requires that no time period should have weight more than $q$, and

$$\mathcal{Q}'_q = \Big\{ (q_1, \ldots, q_T) \in \mathbb{R}^T \,\Big|\, q_t = q_T q^{T-t}, \ t = 1, \ldots, T, \\ \sum_{t=1}^{T} q_t = 1 \Big\}, \quad q \in (0, 1], \tag{10}$$

where weights $q_1, \ldots, q_T$ form a geometric progression (3).

With a chosen time profile set, $\mathcal{Q}$, the agent may define the utility of the time series $X = (x_1, \ldots, x_T)$ by

$$U(X) = \inf_{Q \in \mathcal{Q}} \mathbb{E}_Q[X], \tag{11}$$

and then can use (11) for comparing different time series. In fact, (11) is a data-based analogue of Gilboa & Schmeidler's maxmin model [21].

**Example 1** *(data-based version of coherent risk measure) If $\mathcal{Q}$ in (11) is a subset of $\mathcal{Q}^{\max}$ in (7), then $\mathcal{R}(X) = -U(X)$ is a* coherent risk measure *[2]. The difference is that $X = (x_1, \ldots, x_T)$ is interpreted as an r.v. taking values $x_1, \ldots, x_T$ and $\mathcal{Q}$ is the set of probabilistic scenarios $Q = (q_1, \ldots, q_T)$, i.e., $\mathbb{P}[X = x_t] = q_t$, $t = 1, \ldots, T$. The most popular coherent risk measure is arguably conditional value-at-risk $\text{CVaR}_\alpha(X)$ [43], which for $\alpha = k/T$ is defined as*

$$\text{CVaR}_\alpha(X) = -\frac{1}{k} \sum_{t=1}^{k} x'_t,$$

3

where $(x'_1,\ldots,x'_T)$ is a permutation of $(x_1,\ldots,x_T)$ such that $x'_1 \leqslant \ldots \leqslant x'_T$. In other words, $\mathrm{CVaR}_{k/T}(X)$ measures average losses of $X$ under $k$ worst-case scenarios. Note that $\mathrm{CVaR}_\alpha(X) = \mathrm{CVaR}_\alpha(Y)$ whenever $(y_1,\ldots,y_T)$ is a permutation of $(x_1,\ldots,x_T)$. However, in (11), $q_1,\ldots,q_T$ have a different interpretation: they are weights reflecting "depreciation" of the data at times $t = 1,\ldots,T$. The agent may value more recent data more: $q_1 \leqslant \ldots \leqslant q_T$, which can be viewed as "time aversion" (to be discussed in §3), and may further assume the ratio $q_t/q_{t+1}$ to be constant over $t$, i.e. (10).

**Example 2** *(data-based version of drawdown measure) Let $X = (x_1,\ldots,x_T)$ be a historical time series of the rate of return of some financial asset, and let $\widehat{x}_t = \sum_{j=1}^{t} x_j$ be uncompounded cumulative rate of return over period $[1,t]$. The drawdown of $X$ can be defined by $\xi_t = \max_{1 \leqslant k \leqslant t} \widehat{x}_k - \widehat{x}_t$ [14, 56]. Then the maximum drawdown $\max_{1 \leqslant t \leqslant T} \xi_t$ can be represented in the form of (11) with*

$$\mathcal{Q}'_q = \Big\{ (q_1,\ldots,q_T) \in \mathbb{R}^T \,\Big|\, q_j = 0, \ j \leqslant k, \ t+1 \leqslant j,$$
$$q_j = 1, \ k+1 \leqslant j \leqslant t, \ 0 \leqslant k < t \leqslant T \Big\},$$

*so that*

$$U(X) = \min_{k,t} \sum_{j=k+1}^{t} x_j = -\max_t \xi_t.$$

*Also, the average of the $k$ largest drawdowns can be defined as*

$$\mathrm{DD}_\alpha(X) = \frac{1}{k} \sum_{t=1}^{k} \xi'_t, \quad \alpha = (T-k)/T, \qquad (12)$$

*where $(\xi'_1,\ldots,\xi'_T)$ is a permutation of $(\xi_1,\ldots,\xi_T)$ such that $\xi'_1 \geqslant \ldots \geqslant \xi'_T$ [14, 56]. The functional $U(X) = -\mathrm{DD}_\alpha(X)$ is a particular case of (11). Remarkably, the time-series definition (12) of drawdown, which is a* dynamic *measure, is as simple as time-series analogue of a* one-period *risk measure, e.g. CVaR.*

Thus, while representation (11) is known, the time profile $\mathcal{Q}$ here has the meaning of historical data "depreciation." See Appendix B for a nonlinear generalization of (11).

Let $Q^* = (q_1^*,\ldots,q_T^*) \in \mathcal{Q}^{\max}$ with at least three of $q_1^*,\ldots,q_T^*$ being non-zero. The utility of the time series $X = (x_1,\ldots,x_T)$ can also be measured by the *mean-standard deviation functional*

$$U(X) = V(\mathbb{E}_{Q^*}[X], \sigma_{Q^*}(X)) \qquad (13)$$

with a continuous function $V$ strictly increasing in the first argument and strictly decreasing in the second one.

**Example 3** *(data-based version of the mean-standard deviation utility) The mean-standard deviation utility is defined by $V(m,\sigma) = m - \lambda\sigma$ with a specified "level of risk aversion" $\lambda > 0$ [25, Example 6]. With this $V$, (13) takes the form*

$$U(X) = \mathbb{E}_{Q^*}[X] - \lambda \sigma_{Q^*}(X). \qquad (14)$$

*Note that (14) is a particular case of (11) with [46, Example 1]*

$$\mathcal{Q} = \Big\{ (q_1,\ldots,q_T) \in \mathbb{R}^T \,\Big|\, \mathbb{E}_{Q^*}[Q] = 1, \ \sigma_{Q^*}(Q) \leqslant \lambda \Big\}.$$

**Example 4** *(data-based version of the mean-variance analysis) With a specified threshold $\mu$ on $\mathbb{E}_{Q^*}[X]$, a data-based analogue of the mean-variance analysis corresponds to the utility functional*

$$U(X) = \begin{cases} -\sigma_{Q^*}(X) & \text{when} \quad \mathbb{E}_{Q^*}[X] \geqslant \mu, \\ -\infty & \text{when} \quad \mathbb{E}_{Q^*}[X] < \mu. \end{cases} \qquad (15)$$

In contrast to the existing decision theories, the proposed direct-data based approach does not try to make any statistical inference from the historical data, but rather incorporates agent's perception of the historical data into a decision process through the time profiles, e.g. as in (11) and (13)–(15), and the goal of this work is to identify intertemporal principles of rational choice for constructing time profile sets. An axiomatic framework for intertemporal principles is laid out in §3, and then the direct data-based decision making approach is demonstrated in portfolio optimization with (11) (§4.1) and in mean-variance portfolio selection (§4.2).

# 3 Intertemporal Principles of Rational Choice

This section discusses intertemporal principles of rational choice for constructing time profile sets introduced in §2. For any asset $A$, let

$$x_A(t) = \frac{p_A(t+\Delta) - p_A(t)}{p_A(t)} - r_0(t), \quad t \in \mathcal{T}, \qquad (16)$$

be its historical excess rate of return over the risk-free rate. Here, $\mathcal{T} = \{s_1,\ldots,s_T\}$ is a set of time moments, $\Delta > 0$ is a fixed time period, $p_A(t)$ is the historical price of asset $A$ at time $t$ (specifically, the price of the last transaction before time $t$ [7, p. 132]), whereas $r_0(t)$ is the historical risk-free rate at $t$. If $A$ is the risk-free asset, then $x_A(t) = 0$ for all $t \in \mathcal{T}$. If $A$ is a risky asset, and either $p_A(t)$ or $p_A(t+\Delta)$ or $r_0(t)$ is unavailable for some $t$,[4] we define $x_A(t) = \mathfrak{k}$, where $\mathfrak{k}$ is a symbol indicating missing data (it plays the role similar to $\mathfrak{i} = \sqrt{-1}$ in complex numbers).

Let $X$ be a portfolio consisting of risky assets $A_1,\ldots,A_n$ with portfolio weights $(\alpha_1,\ldots,\alpha_n) \in \mathbb{R}^n$ (short selling is allowed) and of a risk-free asset $A_0$ with the weight $\alpha_0 = 1 - \sum_{i=1}^{n} \alpha_i \in \mathbb{R}$. Let

$$x(t) = \sum_{i=1}^{n} \alpha_i x_i(t) = \underbrace{\sum_{i \in I} \alpha_i x_i(t)}_{=a(t)} + \underbrace{\sum_{i \in J} |\alpha_i| \mathfrak{k}}_{=b(t)}$$
$$= a(t) + b(t)\mathfrak{k}, \quad t \in \mathcal{T},$$

where $x_i(t)$ is defined by (16) for asset $A_i$ and where $I$ and $J$ are sets of indices $i$ such that $x_i(t) \neq \mathfrak{k}$ and $x_i(t) = \mathfrak{k}$, respectively.

---

[4] This may happen when (a) asset $A$ might first appear on the market after time $t$; (b) values of $p_A(t)$ or $p_A(t+\Delta)$ are missing; (c) $t + \Delta > \tau^*$, i.e., *future* price is needed to calculate $x_A(t)$.

Thus, the portfolio $X$ corresponds to a function $x : \mathcal{T} \to F$, where $F = \{a + b\mathfrak{k}, \mid a \in \mathbb{R}, b \geqslant 0\}$ is a real vector space[5] with addition and multiplication by a constant defined by

$$(a_1 + b_1\mathfrak{k}) + (a_2 + b_2\mathfrak{k}) = (a_1 + a_2) + (b_1 + b_2)\mathfrak{k},$$
$$\alpha(a + b\mathfrak{k}) = \alpha a + (|\alpha|b)\mathfrak{k}.$$

The set $\mathcal{X}$ of all possible portfolios is identified with set $F^T$ of all vectors $(x_1, \ldots, x_T)$, where $x_t = x(s_t)$, $t = 1, \ldots, T$, or, equivalently, $\mathcal{X} \subset \mathbb{R}^{2T}$ is the set of vectors $(a_1, b_1, a_2, b_2, \ldots, a_T, b_T)$ with $b_t \geqslant 0$, $t = 1, \ldots, T$, where $x_t = a_t + b_t\mathfrak{k}$, $t = 1, \ldots, T$.

Let $\succeq$ be a preference relation on $\mathcal{X}$: $X \succ Y$ if $X$ is strictly preferred to $Y$, $X \sim Y$ if an agent is indifferent between $X$ and $Y$, and $X \succeq Y$ if either $X \succ Y$ or $X \sim Y$. One of the fundamental principles of rational choice is that $\succeq$ forms a complete weak order on $\mathcal{X}$.

**Axiom 1 (complete weak order)** $\succeq$ *is complete and transitive:*

*(i) $X \succeq Y$ or $Y \succeq X$ for all $X, Y \in \mathcal{X}$ (completeness).*

*(ii) $X \succeq Y$ and $Y \succeq Z$ imply that $X \succeq Z$ for all $X, Y, Z \in \mathcal{X}$ (transitivity).*

Axiom 1(i) asserts that decision making is based solely on historical data and *no other information is available* to the agent (important and non-trivial assumption). Under a mild technical assumption,[6] axiom 1 implies that $\succeq$ admits a numerical representation $U : \mathcal{X} \to \mathbb{R}$ such that $X \succeq Y \iff U(X) \geqslant U(Y)$, see Theorem 2.6 in [19].

Set $\mathcal{X} = F^T$ is a metric space with distance $\rho(X, Y)$:

$$\rho(X, Y) = \sqrt{\sum_{t=1}^{T} \rho_F^2(x_t, y_t)},$$

where $X = (x_1, \ldots, x_T)$, $Y = (y_1, \ldots, y_T)$, and $\rho_F(x, y) = \rho(a + bu, c + du) = \sqrt{(a - c)^2 + (b - d)^2}$. We can then define open and closed subsets of $\mathcal{X}$ with respect to topology induced by this metric.

**Axiom 2 (continuity)** *For every $X \in \mathcal{X}$, sets $\overline{\mathcal{B}}(X) = \{Y \in \mathcal{X} \mid Y \succeq X\}$ and $\underline{\mathcal{B}}(X) = \{Y \in \mathcal{X} \mid X \succeq Y\}$ are closed in $\mathcal{X}$.*

Axiom 2 states that $\succeq$ does not change when $X$ and $Y$ are slightly perturbed. With axiom 2, Theorem 2.15 in [19] implies that a numerical representation $U$ for $\succeq$ can be chosen as a continuous function on $\mathcal{X}$.

It is well known that rational agents *diversify* their portfolios rather than "keep all eggs in one basket."

**Axiom 3 (diversification principle)** *For any $X, Y \in \mathcal{X}$ and $\alpha \in [0, 1]$, mixture $Z_\alpha = \alpha X + (1 - \alpha)Y$ cannot be strictly worse than both $X$ and $Y$: either $Z_\alpha \succeq X$ or $Z_\alpha \succeq Y$.*

A numerical representation $U$ of $\succeq$ that follows this principle is a quasi-concave function: $U(\alpha X + (1 - \alpha)Y) \geqslant \min\{U(X), U(Y)\}$. See [12] for a recent study of quasi-concave utility functions. A trivial sufficient condition for axiom 3 to hold is the existence of *concave* numerical representation $U$ of $\succeq$: i.e., $U(\alpha X + (1 - \alpha)Y) \geqslant \alpha U(X) + (1 - \alpha)U(Y)$ for any $X, Y \in \mathcal{X}$ and $\alpha \in [0, 1]$.

Often investment decisions are made in two steps: (i) decide on portion $\alpha$ of the capital to be invested into risky assets (and keep the remaining money in a savings account), and (ii) select a risky portfolio for the portion $\alpha$. Next axiom states that the choice of the risky portfolio in step (ii) does not depend on $\alpha$.

**Axiom 4 (positive homogeneity)** *For any $X, Y \in \mathcal{X}$ and $\alpha > 0$, $X \succeq Y$ implies that $\alpha X \succeq \alpha Y$.*

**Proposition 1** *Let $\mathcal{X}_+ = \{X \mid X \succeq X_0\}$, where $X_0$ denotes investment into the risk-free asset only. $\succeq$ satisfies axioms 1–4 on $\mathcal{X}_+$ if and only if it has a continuous numerical representation $U$ on $\mathcal{X}_+$ satisfying*

*(i) Concavity[7]: $U(\alpha X + (1 - \alpha)Y) \geqslant \alpha U(X) + (1 - \alpha)U(Y)$ for all $X, Y \in \mathcal{X}_+$ and $\alpha \in [0, 1]$.*

*(ii) Positive homogeneity: $U(\alpha X) = \alpha U(X)$, for every $X \in \mathcal{X}_+$ and $\alpha \geqslant 0$.*

**Proof** See Appendix A. $\qquad\square$

For the agent who considers only portfolios strictly preferable to the risk-free investment $X_0$, restriction to the set $\mathcal{X}_+$ in Proposition 1 is inessential. Suppose $U$ satisfies (i) and (ii) on the whole set $\mathcal{X}$. Then, it can be represented by

$$U(X) = \inf_{Q \in \mathcal{Q}} \langle Q, X \rangle, \qquad (17)$$

where $X = (x_1, \ldots, x_n) = (a_1 + b_1\mathfrak{k}, \ldots, a_T + b_T\mathfrak{k})$ and $\mathcal{Q} \subset R^{2T}$ is a set of vectors $Q = (q_1, \widehat{q}_1, q_2, \widehat{q}_2, \ldots, q_T, \widehat{q}_T)$, which can be chosen convex, closed, and bounded, and where

$$\langle Q, X \rangle = \sum_{t=1}^{T} q_t a_t + \sum_{t=1}^{T} \widehat{q}_t b_t.$$

Elements $Q_X \in \mathcal{Q}$ for which infimum in (17) is attained will be called *identifiers* of $X$.

If $\mathcal{Q} = \{Q\}$ is a singleton, (17) simplifies to $U(X) = \langle Q, X \rangle$. In general, $q_t$ (or $\widehat{q}_t$) is interpreted as the weight/ importance that the agent assigns to the historical data (or absence of data) at time $t \in \mathcal{T}$. The agent may consider a set $\mathcal{Q}$ of possible weights and may select $Q \in \mathcal{Q}$ such that $\langle Q, X \rangle$ is the worst possible: this is an interpretation of (17). Portfolio optimization with (17) takes the form

$$\max_{X \in \mathcal{X}'} \inf_{Q \in \mathcal{Q}} \langle Q, X \rangle,$$

where $\mathcal{X}' \subset \mathcal{X}$ is the set of all feasible portfolios, which motivates calling (17) *maxmin utility theory* (MMUT).

Some agents prefer investing into portfolios which performed well in the past, while avoiding assets with poor or unknown performance, e.g. those which just appear on the market. This intuition is formalized in the following axiom.

---

[5]$F$ is similar to the set of complex numbers $z = a + \mathrm{i}b$, which are also considered as vectors $(a, b)$ with certain arithmetic operations.

[6]Namely, the existence of a countable order dense subset of $\mathcal{X}$. A subset $\mathcal{Z}$ of $\mathcal{X}$ is called order dense, if for any $X, Y \in \mathcal{X}$ with $X \succ Y$ there exists $Z \in \mathcal{Z}$ such that $X \succ Z \succ Y$.

[7]Under (ii), this is equivalent to super-additivity: $U(X + Y) \geqslant U(X) + U(Y)$ for all $X, Y \in \mathcal{X}_+$.

**Axiom 5 (monotonicity)** *Let* $X = (a_1 + b_1 \mathfrak{k}, \ldots, a_T + b_T \mathfrak{k})$ *and* $Y = (c_1 + d_1 \mathfrak{k}, \ldots, c_T + d_T \mathfrak{k}) \in \mathcal{X}$. *If*

$$a_t \geqslant c_t \quad \text{and} \quad b_t \leqslant d_t, \qquad t = 1, \ldots, T,$$

*then* $X \succeq Y$. *In other words, if the proportion of money invested into assets with no historical data in X is less than that in Y and if collectively the assets with known data in X always outperformed those in Y, then* $X \succeq Y$.

In the MMUT (17), axiom 5 holds if an only if $\mathcal{Q}$ in (17) is such that

$$q_t \geqslant 0 \quad \text{and} \quad \widehat{q}_t \leqslant 0, \quad t = 1, \ldots, T, \qquad \forall Q \in \mathcal{Q}.$$

If no data is missing and the agent uses direct historical simulation for forecasting, axiom 5 is equivalent to monotonicity axiom for r.v.'s.

**Corollary 1** *Direct historical simulation views* $x_1, \ldots, x_T$ *as equally probable outcomes of an r.v.* $R$: $\mathbb{P}[R = x_t] = 1/T$, $t = 1, \ldots, T$. *Let* $X = (x_1, \ldots, x_T)$ *and* $Y = (y_1, \ldots, y_T)$, *then with this assumption, the condition* $x_1 \geqslant y_1, \ldots, x_T \geqslant y_T$ *in axiom 5 is equivalent to* $\mathbb{P}[R(X) \geqslant R(Y)] = 1$, *and axiom 5 simplifies to*

$$\mathbb{P}[R(X) \geqslant R(Y)] = 1 \quad \Rightarrow \quad X \succeq Y, \tag{18}$$

*which is the well-known monotonicity axiom on the space of r.v.'s.*

However, financial interpretations of axiom 5 and (18) are completely different. With simple historical simulation, the agent interprets inequalities $x_1 \geqslant y_1, \ldots, x_T \geqslant y_T$ as evidence that a portfolio associated with $X = (x_1, \ldots, x_T)$ will outperform the one associated with $Y = (y_1, \ldots, y_T)$ in the future with probability 1 and, thus, strongly prefers $X$ to $Y$. This again shows that historical data can be easily misinterpreted (mis-modeled), which could result in poor decisions. In contrast, axiom 5 does not imply that the portfolio associated with $X$ will outperform the one associated with $Y$ in the future—it merely states historical facts.

In fact, when forecasting uses methods other than direct historical simulation, axiom 5 differs from (18) as the following example demonstrates.

**Example 5** *In a "Gaussian world," the future excess rate of return of a portfolio with past excess rates of return* $X = (x_1, \ldots, x_T)$ *can be modeled by a normally distributed r.v.* $R(X)$ *with mean* $\mu_X$ *and variance* $\sigma_X^2$ *estimated from X by*

$$\mu_X = \frac{1}{T} \sum_{t=1}^{T} x_t, \qquad \sigma_X^2 = \frac{1}{T-1} \sum_{t=1}^{T} (x_t - \mu_X)^2.$$

*Then* (18) *simplifies to*

$$X \succeq Y \quad \text{if} \quad \mu_X \geqslant \mu_Y \quad \text{and} \quad \sigma_X^2 \leqslant \sigma_Y^2,$$

*which neither implies axiom 5 nor follows from it.*

Axiom 5 assumes that $X$ outperformed $Y$ at every single time moment in the past. Some agents, however, may consider cumulative past performance.

**Axiom 6 (time aversion)** *Let* $X = (a_1 + b_1 \mathfrak{k}, \ldots, a_T + b_T \mathfrak{k})$ *and* $Y = (c_1 + d_1 \mathfrak{k}, \ldots, c_T + d_T \mathfrak{k}) \in \mathcal{X}$. *If*

$$\sum_{t=\tau}^{T} a_t \geqslant \sum_{t=\tau}^{T} c_t, \quad \sum_{t=\tau}^{T} b_t \leqslant \sum_{t=\tau}^{T} d_t, \quad \tau = 1, \ldots, T, \tag{19}$$

*then* $X \succeq Y$.

Let $\mathcal{T}_1, \mathcal{T}_2 \subset \mathcal{T}$ be subsets of $\mathcal{T}$ of equal cardinality, such that $\max_t \mathcal{T}_1 \leqslant \min_t \mathcal{T}_2$. Let $I_{\mathcal{T}_1, \mathcal{T}_2}(t)$ be a function taking values 1, $-1$, and 0, for $t \in \mathcal{T}_1$, $t \in \mathcal{T}_2$, $t \in \mathcal{T} \setminus (\mathcal{T}_1 \cup \mathcal{T}_2)$, respectively. Since the first condition in (19) holds whenever $c(t) = a(t) + \delta I_{\mathcal{T}_1, \mathcal{T}_2}(t)$, $\delta > 0$, axiom 6 implies that

$$x(t) \succeq x(t) + \delta I_{\mathcal{T}_1, \mathcal{T}_2}(t) \quad \forall \delta > 0. \tag{20}$$

Similarly, let $J_{\mathcal{T}_1, \mathcal{T}_2}(t)$ be a function taking values $-\mathfrak{k}$, $\mathfrak{k}$, and 0, for $t \in \mathcal{T}_1$, $t \in \mathcal{T}_2$, $t \in \mathcal{T} \setminus (\mathcal{T}_1 \cup \mathcal{T}_2)$, respectively. Axiom 6 implies that

$$x(t) \succeq x(t) + \delta J_{\mathcal{T}_1, \mathcal{T}_2}(t) \quad \forall \delta > 0. \tag{21}$$

For the MMUT (17), axiom 6 yields the following characterization of the set $\mathcal{Q}$.

**Proposition 2** *Let* (17) *be a numerical representation of* $\succeq$ *on* $\mathcal{X}$. *The following statements are equivalent.*
 (i) $\succeq$ *satisfies axiom 6.*
 (ii) $\succeq$ *satisfies* (20) *and* (21).
 (iii) *The set* $\mathcal{Q}$ *in* (17) *is such that*

$$0 \leqslant q_1 \leqslant \ldots \leqslant q_T, \quad 0 \geqslant \widehat{q}_1 \geqslant \ldots \geqslant \widehat{q}_T \quad \forall Q \in \mathcal{Q}.$$

**Proof** See Appendix A. □

Many models of financial market assume that observations of asset's rates of return at different times are *independent*. Black & Scholes [5] priced options based on the assumption that stock prices follow a Brownian motion, whereas Sato [48] replaced the Brownian motion by a Levy process, which also assumes price increments (and consequently returns) to be independent. In such models, discrete historical data $x_1, \ldots, x_T$ can be considered as realizations of *independent* and identically distributed r.v.'s. In this case, the *order* of data is not important, which corresponds to the following *no-time-structure* principle

$$X \sim Y \text{ whenever } (y_1, \ldots, y_T) \text{ is a permutation of } (x_1, \ldots, x_T), \tag{22}$$

which implies that $X = (a_1 + b_1 \mathfrak{k}, \ldots, a_T + b_T \mathfrak{k})$ with $a_1 \leqslant \ldots \leqslant a_T$ (positive trend) and $Y = (a_T + b_T \mathfrak{k}, \ldots, a_1 + b_1 \mathfrak{k})$ with the reverse order of data (negative trend), are equally preferable—this does not seem to be a rational behavior.

An axiomatic framework for data-based expected utility theory (36) is presented in Appendix B.

When there is no missing data or when the agent deliberately excludes assets with missing/incomplete data, every portfolio $X$ can be identified with its historical time series $(x_1, \ldots, x_T) \in \mathbb{R}^T$. In this case, axioms on a preference relation $\succeq$ are introduced and studied on $\mathcal{X} = \mathbb{R}^T$.

- Axiom 1 states that $\succeq$ defines a complete and transitive weak order on $\mathbb{R}^T$.
- Axiom 2, formulated in usual topology of $\mathbb{R}^T$, states that the sets $\{Y \in \mathcal{X} \mid Y \succeq X\}$ and $\{Y \in \mathcal{X} \mid X \succeq Y\}$ are closed in $\mathbb{R}^T$ and implies the existence of continuous function $U : \mathbb{R}^T \to \mathbb{R}$ such that $X \succeq Y \Leftrightarrow U(X) \geqslant U(Y)$.
- Axiom 3 states that $U$ is a quasi-concave function on $\mathbb{R}^T$.
- Axioms 3 and 4 imply that $U$ is concave and positive homogeneous, hence admit the form (11).
- Axiom 5 states that if $X = (x_1, \ldots, x_T)$ and $Y = (y_1, \ldots, y_T)$ are such that $x_1 \geqslant y_1, \ldots, x_T \geqslant y_T$, then $X \succeq Y$. It implies that $q_t$ in (11) are non-negative.
- Axiom 6 states that $X = (x_1, \ldots, x_T) \succeq Y = (y_1, \ldots, y_T)$ provided that $\sum_{t=\tau}^{T} x_t \geqslant \sum_{t=\tau}^{T} y_t$, $\tau = 1, \ldots, T$ (portfolios with better average *recent* performance are preferable). It implies that $q_1 \leqslant \ldots \leqslant q_T$ for every $Q \in \mathcal{Q}$ in (11).

Now, with what introduced axioms is the mean-standard deviation functional (13) consistent?

**Proposition 3** (13) *is a numerical representation U of $\succeq$ if and only if $\succeq$ satisfies axioms 1 and 2, and the following two additional axioms*
(a) $X + C \succ X$ for all $C > 0$, and
(b) if $X$ and $Y \neq 0$ are such that $\mathbb{E}_{Q^*}[Y] = \mathrm{cov}_{Q^*}(X, Y) = 0$, then $X \succ X + Y$.

**Proof** See [33, Theorem 1]. □

Next propositions address consistency of the mean-standard deviation utility (14) with axiom 5 (monotonicity).

**Proposition 4** *Let U be given by* (14). *The following statements are equivalent*
(a)
$$\lambda \leqslant \sqrt{\frac{z}{1-z}}, \qquad z = \min_{t \in \{t : q_t^* > 0\}} q_t^*. \tag{23}$$

(b) $q_1 \geqslant 0, \ldots, q_T \geqslant 0$ for every $Q \in \mathcal{Q}^*$.
(c) $U(X) \geqslant U(Y)$ for every $X, Y \in \mathbb{R}^T$ such that $X \geqslant Y$.

**Proof** See Appendix A. □

**Proposition 5** *Let $\succeq$ be a preference relation with a numerical representation* (14). *Then $\succeq$ satisfies axioms 1, 2, and 3. Moreover, $\succeq$ satisfies: (i) axiom 5 if* (23) *holds; (ii) axiom 6 if* (23) *holds and $q_1^* = \cdots = q_T^*$; (iii) axiom 8 in Appendix B if* (23) *holds and $q_1^* \leqslant \ldots \leqslant q_T^*$; and (iv)* (22) *if $q_1^* = \cdots = q_T^*$.*

**Proof** See Appendix A. □

What axioms are considered to be rational?—In Christoforou's questionnaire [15],[8] almost all respondents out of 215 agreed with axiom 1, 88% with axiom 5, 58% with axiom 2, and only 21% agreed with (22).

---

[8] The posed questions such as "Would you favor instrument that historically returned profit 10% during 50% of years (and no profit during other 50% of years) over instrument which returned profit 6% during 100% of years?" better reflect real investment alternatives, where only past data but not probabilities of future events are available.

# 4 Data-based Portfolio Optimization

## 4.1 Maxmin Portfolio Optimization

In the direct data-based decision approach, historical excess rates or return of the risk-free asset and $n$ risky assets during last $T$ time periods are given by $X_0 = (0, \ldots, 0)$, $X_1 = (x_{11}, \ldots, x_{1T}), \ldots, X_n = (x_{n1}, \ldots, x_{nT})$, respectively. With no short sales and with $U$ being the utility (11), a portfolio optimization problem is formulated by

$$u^* = \max_{v \in V} \min_{Q \in \mathcal{Q}} \sum_{t=1}^{T} \sum_{i=0}^{n} q_t v_i x_{it},$$
$$V = \left\{ v = (v_0, \ldots, v_n)^\top \in \mathbb{R}^{n+1} \,\middle|\, v \geqslant 0, \ \sum_{i=0}^{n} v_i = 1 \right\}. \tag{24}$$

**Example 6 (single time profile)** *For a singleton $\mathcal{Q} = \{Q^*\}$, problem* (24) *simplifies to*

$$u^* = \max_{v \in V} \sum_{i=0}^{n} v_i u_i, \quad u_i = \sum_{t=1}^{T} q_t^* x_{it}, \quad i = 0, \ldots, n.$$

*In this case, $u^* = \max_i u_i$, and an optimal strategy is $v_{i^*} = 1$, $v_i = 0, i \neq i^*$, where $i^* = \arg\max_i u_i$, i.e. investing the whole capital in the "best" asset—no diversification.*

**Example 7 ("ultimate" risk aversion)** *For $\mathcal{Q} = \mathcal{Q}^{\max}$,* (24) *is equivalent to the problem of finding mixed-strategy Nash equilibrium in a two-player zero-sum game [53] with $(n + 1) \times T$ payoff matrix $\mathbf{X}$ having elements $x_{it}$. In this case, $u^*$ is equal to the value of the game, and the optimal investment strategy $v^* \in V$ is a solution to the linear program*

$$u^* = \max_{v \in V, u \in \mathbb{R}} u$$
$$\text{subject to} \quad \sum_{i=0}^{n} v_i x_{it} \geqslant u, \quad t = 1, \ldots, T.$$

*If $\mathcal{Q}$ is an arbitrary closed convex subset of $\mathcal{Q}^{\max}$, von Neumann minimax theorem [54] implies that $u^* = \max_{v \in V} \min_{Q \in \mathcal{Q}} v^\top \mathbf{X} Q^\top = \min_{Q \in \mathcal{Q}} \max_{v \in V} v^\top \mathbf{X} Q^\top$, and $u^*$ can be found as*

$$u^* = \min_{Q \in \mathcal{Q},\, u \in \mathbb{R}} u$$
$$\text{subject to} \quad \sum_{t=1}^{T} q_t x_{it} \leqslant u, \quad i = 0, \ldots, n. \tag{25}$$

**Example 8** *For $\mathcal{Q}$ being the maximal "time averse" set* (8), *or for $\mathcal{Q}$ belonging to the one-parameter family* (9), *problem* (25) *is a linear program.*

The set $\mathcal{Q}$ in (11) can also be found by the inverse portfolio approach introduced in [26, 27] in terms of r.v.'s. The idea is that the agent recovers $\mathcal{Q}$ from the time series $X^* = (x_1^*, \ldots, x_T^*)$ of the rate of return of a portfolio that he/she is relatively satisfied with—such a portfolio should solve (24) with $U$ given by (11). Proposition 7 in [26] implies that the maximal possible (most robust) such $\mathcal{Q}$ is given by

$$\mathcal{Q} = \left\{ Q \in \mathcal{Q}^{\max} \,\middle|\, \sum_{t=1}^{T} q_t x_t^* \geqslant 0 \right\},$$

provided that the following *no-perfect-history assumption* holds: there is no time series $(x_1, \ldots, x_T)$ of the rate of return of a feasible portfolio such that $x_1 \geqslant 0, \ldots, x_T \geqslant 0$ with at least one inequality being strict. For example, for $T \gg n$, there is unlikely to be a portfolio that outperforms the risk-free asset for every time period in the past. Thus, $T$ can be chosen sufficiently large to guarantee that the no-perfect-history assumption holds, otherwise the historical time series can be perceived to be too short for making a reliable decision.

## 4.2 Mean-Variance Portfolio Selection

Suppose there is a risk-free asset with constant rate of return $r_0$, and there are $n$ risky assets. In the typical approach (1), rates of return of the risky assets are modeled as r.v.'s $r_1, \ldots, r_n$ on some probability space. With $v_i$ being the fraction of the capital invested into asset $i$, the rate of return of a portfolio is then $\sum_{i=0}^{n} v_i r_i$. The mean-variance portfolio selection problem [39]

$$\min_{v_0, v_1, \ldots, v_n} \sigma\left(\sum_{i=0}^{n} v_i r_i\right)$$
$$\text{subject to} \quad \sum_{i=0}^{n} v_i = 1, \quad \mathbb{E}\left[\sum_{i=0}^{n} v_i r_i\right] \geqslant r_0 + \Delta, \quad (26)$$

with desired premium $\Delta > 0$ over $r_0$, continues to be a cornerstone of the modern portfolio theory: *one-fund theorem, two-fund theorem, capital asset pricing model (CAPM), Sharpe ratio and asset beta* all stem from (26), see [34],—this is one of the reasons to consider its data-based analogue:

$$\min_{v_0, v_1, \ldots, v_n} \sigma_Q\left(\sum_{i=0}^{n} v_i X_i\right)$$
$$\text{subject to} \quad \sum_{i=0}^{n} v_i = 1, \quad \mathbb{E}_Q\left[\sum_{i=0}^{n} v_i X_i\right] \geqslant r_0 + \Delta, \quad (27)$$

where $X_i = (x_{i1}, \ldots, x_{iT}) \in \mathbb{R}^T$ is the time series of the rate of return for risky asset $i \in \{1, \ldots, n\}$, and $\sigma_Q(\cdot)$, $\mathbb{E}_Q[\cdot]$ and $\text{cov}_Q(\cdot, \cdot)$ are defined by (4), (5) and (6), respectively, with a time profile $Q = (q_1, \ldots, q_T) \in \mathcal{Q}^{\max}$. Note that (27) is equivalent to $\max_{v_0, v_1, \ldots, v_n} U\left(\sum_{i=0}^{n} v_i X_i\right)$ subject to $\sum_{i=0}^{n} v_i = 1$ with $U$ defined by (15).

Optimal portfolio weights $v_0^*$ and $v^* = (v_1^*, \ldots, v_n^*)^\top$ in (27) are found in a closed-form [57, (8.2.2)]

$$v^* = \frac{\Delta \Lambda_Q^{-1}(\mathbf{X}Q^\top - r_0 e)}{(\mathbf{X}Q^\top - r_0 e)^\top \Lambda_Q^{-1}(\mathbf{X}Q^\top - r_0 e)}, \quad (28)$$
$$v_0^* = 1 - \sum_{i=1}^{n} v_i^*,$$

where $e = (1, \ldots, 1)^\top$ is the $n$-dimensional unit vector and where $\mathbf{X}$ and $\Lambda_Q$ are matrices with entries $x_{it}$, $i = 1, \ldots, n$, $t = 1, \ldots, T$, and $\text{cov}_Q(X_i, X_j)$, $i, j = 1, \ldots, n$, respectively.

If $e^\top v^* = 1$, which implies that $v_0^* = 0$ (no investment into the risk-free asset), then the optimal portfolio is called a master fund of positive type (market portfolio) [45] with the weights [57, (8.2.4)]

$$v_M = \frac{\Lambda_Q^{-1}(\mathbf{X}Q^\top - r_0 e)}{e^\top \Lambda_Q^{-1}(\mathbf{X}Q^\top - r_0 e)}, \quad (29)$$

and the time series of the rate of return $X_M = v_M^\top \mathbf{X}$. The optimality conditions for the master fund can be restated as the capital asset pricing model (CAPM) [46]:

$$\mathbb{E}_Q[X_i] - r_0 = \beta_i(\mathbb{E}_Q[X_M] - r_0), \quad \beta_i = \frac{\text{cov}_Q(X_i, X_M)}{\sigma_Q^2(X_M)},$$
$$i = 1, \ldots, n. \quad (30)$$

Problem (26) requires knowing the expected values and variance-covariance matrix of asset rates of return. If historical rates of return of each asset are assumed to be realizations of i.i.d. r.v.'s, then (26) is a particular case of (27) with $q_1 = \cdots = q_T = 1/T$, which is what is solved in practice. However, (26) does not distinguish portfolios with different trends but having the same histogram of historical rates over the same period of time. Also, it is well-known that (26) is inconsistent with ordinary monotonicity axiom[9] (18). However, in the light of the historical data "depreciation," this could be a result of mis-interpretation of historical data as a "forecast" for future rates of return. In fact, what is violated is axiom 5 (monotonicity). Indeed, the agent may well believe that assets with good historical performance are now overpriced and may prefer a portfolio with acceptable past performance and least historical volatility.

**Example 9** *Let $q_1, \ldots, q_T$ be given by the geometric progression (3) for some $q \in (0, 1]$. We select $n = 10$ assets from the S & P 500 index[10] and use $T = 60$ assets' monthly rates of return from 1-May-2011 to 1-May-2016. Also, $r_0 = 0.1\%$. Tables 1 and 2 present master fund's weights[11] (29) and asset betas in (30) for $q = 1$, 0.99, 0.97 and 0.95, whereas Figure 2 shows proportions of the capital invested into Apple Inc. (AAPL), Amazon.com Inc (AMZN), Bank of America Corp (BAC). Figure 3 depicts three mean-variance efficient frontiers for $q = 1$, 0.99, and 0.95. Figure 4 shows Sharpe ratio $S = \frac{\mathbb{E}[X^*] - r_0}{\sigma(X^*)}$ for the optimal portfolio (28) as a function of $q$. The optimal portfolios (29) are then tested on the out-of-sample from 1-May-2016 to 1-May-2017. Figure 5 shows in-sample and out-of-sample performance for the master funds (29) with $q = 0.9$, 0.95, 0.99 and 1 and for the one in (26), in which $r_1, \ldots, r_n$ follow the ARIMA time-series model [8, Chapter 6] with parameters determined automatically in Mathematica.*

**Example 10** *Figure 6 depicts the mean-variance efficient frontier for the Markowitz's portfolio problem (27) with the time*

---

[9]The mean-variance approach may violate "ordinary" monotonicity axiom: if $R_X$ is the rate of return of a solution of (26), then there may exist another feasible portfolio with the rate of return $R_Y$, such that $\mathbb{P}[R_Y \geqslant R_X] = 1$ and $\mathbb{P}[R_Y > R_X] > 0$, see [25, Example 5]. The best monotone approximation of a mean-variance functional is obtained in [35]. Alternatively, one may obtain a monotone preference relation if the standard deviation in (26) is replaced by a general deviation measure [44, 45, 46].

[10]Namely, Apple Inc. (AAPL), Amazon.com Inc (AMZN), Bank of America Corp (BAC), Twenty-First Century Fox Inc Class B (FOX), IBM Common Stock (IBM), The Coca-Cola Co (KO), McDonald's Corporation (MCD), Microsoft Corporation (MSFT), Nike Inc (NKE), and Visa Inc (V).

[11]Negative portfolio weights correspond to short selling.

Table 1: Master fund's weights (29)

|  | $q = 1$ | $q = 0.99$ | $q = 0.95$ | ARIMA |
|---|---|---|---|---|
| AAPL | 0.1683 | 0.1203 | $-0.3572$ | 0.0326 |
| AMZN | 0.3107 | 0.3857 | 1.0597 | 0.0290 |
| BAC | $-0.1536$ | $-0.1742$ | $-0.7656$ | $-0.0756$ |
| FOX | $-0.1410$ | $-0.2690$ | $-0.9489$ | 0.3970 |
| IBM | $-0.3572$ | $-0.3863$ | $-0.1950$ | 0.0414 |
| KO | $-0.2479$ | $-0.1952$ | 0.0492 | 0.4038 |
| MCD | $-0.0865$ | $-0.0622$ | 0.2467 | $-0.3802$ |
| MSFT | 0.2222 | 0.2528 | 0.3238 | $-0.2077$ |
| NKE | 0.1107 | 0.1057 | 0.0746 | 0.4670 |
| V | 1.1742 | 1.2224 | 1.5128 | 0.2926 |

Table 2: Asset betas in CAPM (30)

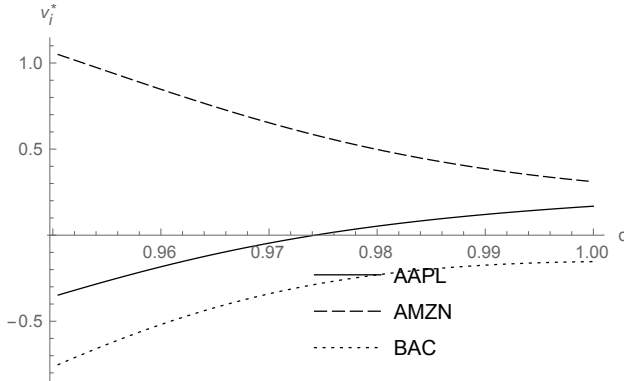| $q$ | 1 | 0.99 | 0.97 | 0.95 |
|---|---|---|---|---|
| AAPL | 0.9958 | 1.0870 | 1.2533 | 1.3757 |
| AMZN | 1.0596 | 1.1765 | 1.3839 | 1.5338 |
| BAC | 1.7657 | 1.6715 | 1.5341 | 1.4750 |
| FOX | 1.2788 | 1.3252 | 1.3946 | 1.4172 |
| IBM | 0.7261 | 0.7649 | 0.8457 | 0.9249 |
| KO | 0.5313 | 0.5847 | 0.6675 | 0.7102 |
| MCD | 0.5120 | 0.5470 | 0.6073 | 0.6485 |
| MSFT | 1.0042 | 1.0596 | 1.1766 | 1.2758 |
| NKE | $-0.1475$ | $-0.3406$ | $-0.7033$ | $-0.9638$ |
| V | 0.8441 | 0.8997 | 0.9983 | 1.0626 |



Figure 2: Proportions of the initial capital invested into the three assets: AAPL, AMZN and BAC; see Example 9.

weights chosen according to the ACF of the FTSE 100 index in Figure 1. We select $n = 70$ most actively traded[12] assets from the FTSE 100 index and use assets' daily rates of return from 1-April-2015 to 1-April-2016. It is assumed that $r_0 = 0.01\%$.

**Remark 1** *The weights $Q = (q_1, \ldots, q_T)$ can be viewed as parameters, and sensitivity of the optimal value $V(Q) = \sigma_Q(X^*(Q))$ in (27) with respect to changes in $q_1, \ldots, q_T$ can*

---

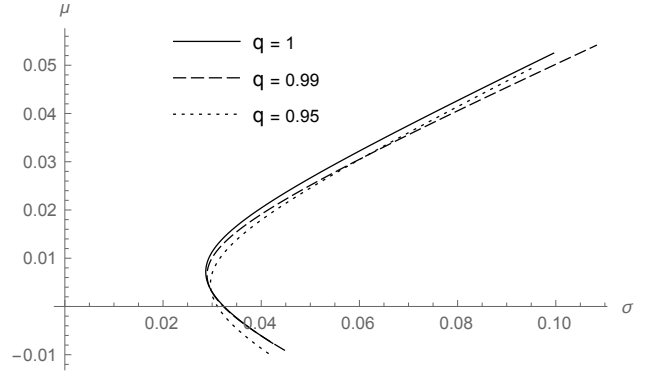[12]The comparison is made based on average trading volume.



Figure 3: Mean-variance efficient frontiers for three "time" profiles.
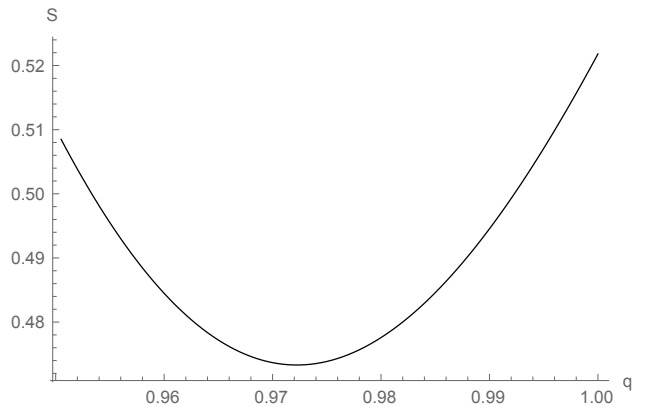


Figure 4: Dependence of the Sharpe ratio on the "time" profile.

*be assessed by the envelope theorem [52, Theorem 3.10.4] :*

$$\frac{\partial V}{\partial q_t} = \left. \frac{\partial \sigma_Q(X)}{\partial q_t} \right|_{X=X^*} - \lambda^*(Q) \left. \frac{\partial}{\partial q_t} (\mathbb{E}_Q[X] - r_0 - \Delta) \right|_{X=X^*},$$

*where $\lambda^*(Q) = \sigma_Q(X^*(Q))/(r_0 + \Delta)$ is an optimal Lagrange multiplier and*

$$\frac{\partial \mathbb{E}_Q[X]}{\partial q_t} = \frac{\partial}{\partial q_t} \sum_{t=1}^{T} q_t x_t = x_t,$$

$$\frac{\partial \sigma_Q(X)}{\partial q_t} = \frac{1}{2\sigma_Q(X)} \frac{\partial}{\partial q_t} (\mathbb{E}_Q\left[X^2\right] - (\mathbb{E}_Q[X])^2)$$

$$= \frac{x_t^2 - 2\mathbb{E}_Q[X]x_t}{2\sigma_Q(X)},$$

*so that*

$$\frac{\partial V}{\partial q_t} = \left( \frac{x_t^*}{2V} - \frac{r_0 + \Delta}{V} - \frac{V}{r_0 + \Delta} \right) x_t^*.$$

## 5 Conclusions

In a typical one-period decision making under uncertainty, outcomes of feasible actions are modeled as r.v.'s. As a result, optimal decisions depend on the accuracy of estimation of the
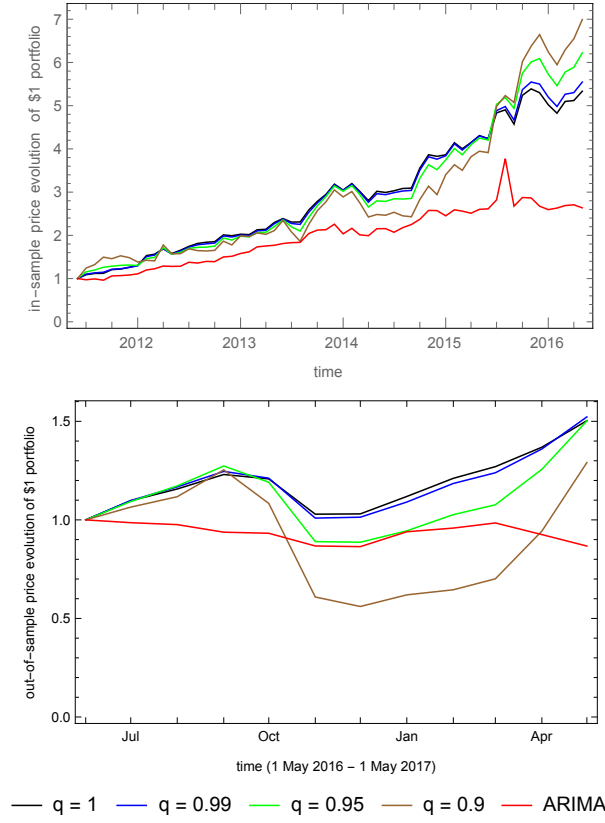
Figure 5: In-sample and out-of-sample price evolution of 1$ invested at the beginning of the corresponding periods into the master fund in (27) with $q = 1, 0.99, 0.97$, and $0.95$, and in (26) with the ARIMA model, respectively.
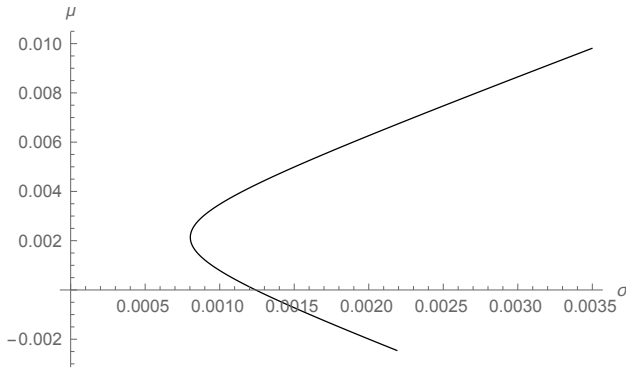


Figure 6: Mean-variance efficient frontier with the time weights from Figure 1.

corresponding probability distributions. Agents who believe that probability distributions of asset rates cannot be reliably estimated would unlikely use any random variable-based decision theory. They will also find it hard to apply Gilboa & Schmeidler's case-based decision theory [22, 23], which requires a similarity measure for market behavior over different time periods. As an alternative, this work has formulated "intertemporal" principles/axioms for a preference relation on the

space of historical time series to facilitate making a rational choice in portfolio selection. It does *not* suggest to dismiss existing decision theories which include uncertainty modeling. Instead, it shows how to *adapt* them to deal with historical time series. This adaptation, however, is not "mechanical": some of the proposed axioms, e.g. "time aversion," have no direct analogue in the existing theories. Example 5 demonstrates that the same axiom (in this case, monotonicity) may lead to completely different decisions when applied to r.v.'s and to time series. Thus, instead of making statistical inference from the historical data, an agent may incorporate his/her perception of the data through time profiles and make a decision based on the data directly.

Figure 5 (Example 9) shows that in mean-variance portfolio selection, the optimal portfolios with the exponential time profiles with various $q$ outperform optimal portfolio in which asset rates of return are modeled by the ARIMA model. However, *no matter what advantage on a particular dataset*, the direct data-based decision making approach demonstrates over approaches with uncertainty modeling, those agents who *believe* that asset prices (rates of return) can be reliably predicted by merely statistical means could hardly be discouraged—they will continue either relying on some "trusted" statistical model or searching for an ideal one. This work aims to provide an alternative decision making approach for those *who do not have such a belief*.

While the focus of this work is on direct data-based analogues of the Gilboa & Schmeidler's maxmin model, the mean-variance approach, and the EUT, other existing decision theories can be reinterpreted and analyzed in a similar fashion.

## Acknowledgment

# A   Appendix: Proofs

### Proof of Proposition 1

The "if" part is trivial—we proceed with the "only if" one. Let $\succeq$ satisfy axioms 1–4 on $\mathcal{X}_+$, and let $U'$ be a continuous numerical representation of $\succeq$ such that $U'(X_0) = 0$. If $U'(X) = 0$ for all $X \in \mathcal{X}_+$, then (i) and (ii) hold. Otherwise, fix some $X^* \in \mathcal{X}_+$ with $U'(X^*) > 0$, and let $f(\alpha) = U'(\alpha X^*)$. By axiom 4, $X^* \succ X_0$ yields $\alpha X^* \succ \alpha X_0$, hence $f(\alpha) > U'(\alpha X_0) = 0$ for all $\alpha > 0$. Axiom 2 implies that $f$ is a continuous function on $(0, \infty)$ with

$$\lim_{\alpha \to 0^+} f(\alpha) = 0. \tag{31}$$

Let $f(\alpha_2) = f(\alpha_1)$ for some $\alpha_2 < \alpha_1$. Then, $k(\alpha_1 X^*) \sim \alpha_1 X^*$, where $k = \alpha_2/\alpha_1 < 1$, and, by axiom 4, $k^2(\alpha_1 X) \sim k(\alpha_1 X) \sim \alpha_1 X$. Similarly, $k^n(\alpha_1 X^*) \sim \alpha_1 X^*$ for all $n$, hence $f(k^n \alpha_1) = f(\alpha_1)$ for all $n$, which contradicts (31). Thus, $f$ is a positive continuous function taking distinct values, which implies that $f$

is strictly monotone. With (31), $f$ cannot be decreasing, hence it is strictly increasing. Thus, $f$ has a strictly increasing inverse function $f^{-1}$, and $U(X) := f^{-1}(U'(X))$ is another numerical representation of $\succeq$. Then $U(\alpha X^*) = f^{-1}(U'(\alpha X^*)) = \alpha$ for all $\alpha > 0$. Then $X \sim U(X)X^*$ for all $X \in \mathcal{X}_+$, and, by axiom 4, $\alpha X \sim \alpha U(X)X^*$. Hence, $U(\alpha X) = U(\alpha U(X)X^*) = \alpha U(X)$, and (ii) follows.

For any $X, Y \in \mathcal{X}_+$ with $U(X) > 0$ and $U(Y) > 0$, $U(X/U(X)) = U(Y/U(Y)) = 1$ by (ii), and axiom 3 implies that $U(\alpha X/U(X) + (1-\alpha)U(Y/U(Y))) \geqslant 1$ for all $\alpha \in [0,1]$. With $\alpha = U(X)/(U(X) + U(Y))$ this yields $U(X + Y) \geqslant U(X) + U(Y)$, which with (ii), is equivalent to (i).

## Proof of Proposition 2

For $(iii) \Rightarrow (i)$ part, select any $X, Y$ satisfying (19) and choose any $Q \in \mathcal{Q}$. Let $\delta_1 = q_1 \geqslant 0$ and $\delta_t = q_t - q_{t-1} \geqslant 0$, $t = 2, \ldots, T$. Then

$$\sum_{t=1}^{T} q_t a_t = \sum_{t=1}^{T} \left( \sum_{\tau=1}^{t} \delta_\tau \right) a_t = \sum_{\tau=1}^{T} \delta_\tau \sum_{t=\tau}^{T} a_t$$
$$\geqslant \sum_{\tau=1}^{T} \delta_\tau \sum_{t=\tau}^{T} c_t = \sum_{t=1}^{T} q_t c_t,$$

where the inequality follows from (19). By similar argument, $\sum_{t=1}^{T} \widehat{q}_t b_t \geqslant \sum_{t=1}^{T} \widehat{q}_t d_t$, so that $U(X) \geqslant U(Y)$, and, consequently, $X \succeq Y$.

$(i) \Rightarrow (ii)$ is straightforward. For $(ii) \Rightarrow (iii)$, let $Q^Y = (q_1^Y, \widehat{q}_1^Y, \ldots, q_T^Y, \widehat{q}_T^Y) \in \mathcal{Q}$ be an identifier of any $Y = (a_1 + b_1\mathfrak{k}, \ldots, a_T + b_T\mathfrak{k}) \in \mathcal{X}$ and let $q_i^Y > q_j^Y$ for some $i < j$. Then

$$U(X) \leqslant \langle Q^Y, X \rangle = \sum_{t=1}^{T} q_t^Y a_t - \delta(q_i^Y - q_j^Y)$$
$$+ \sum_{i=1}^{T} \widehat{q}_t^Y b_t < \langle Q^Y, y(t) \rangle = U(Y),$$

for $X$ defined by $x(t) = y(t) - \delta I_{\mathcal{T}_1, \mathcal{T}_2}(t)$ with $\mathcal{T}_1 = \{i\}$, $\mathcal{T}_2 = \{j\}$. This contradicts (20), and, consequently, $0 \leqslant q_1^Y \leqslant \ldots \leqslant q_T^Y$ for every $Y \in \mathcal{X}$. Similarly, (21) yields $0 \geqslant \widehat{q}_1^Y \geqslant \ldots \geqslant \widehat{q}_T^Y$.

Let $\mathcal{Q}' \subseteq \mathcal{Q}$ be the closure of the convex hull of all $Q \in \mathcal{Q}$, which are identifiers of some $Y \in \mathcal{X}$, and let $U'$ be given by (17) with $\mathcal{Q}'$. Then $U(Y) = \langle Q^Y, Y \rangle \geqslant U'(Y)$ for every $Y \in \mathcal{X}$, so that $\mathcal{Q} \subseteq \mathcal{Q}'$, which yields $\mathcal{Q}' = \mathcal{Q}$. Consequently, $\mathcal{Q}$ is the closure of the convex hull of some vectors satisfying (ii), and thus, this condition holds for every $Q \in \mathcal{Q}$.

## Proof of Proposition 4

$(a) \rightarrow (b)$: (23) implies that

$$\frac{z}{1-z} \geqslant \lambda^2 \geqslant \sigma_{Q^*}^2(Q)$$
$$= q_1^*(q_1 - 1)^2 + \sum_{t=2}^{T} q_t^*(q_t - 1)^2$$

for every $Q \in \mathcal{Q}^*$. By convexity of $f(x) = x^2$,

$$\sum_{t=2}^{T} q_t^*(q_t - 1)^2 \geqslant \sum_{t=2}^{T} q_t^* \left( \frac{\sum_{t=2}^{T} q_t^*(q_t - 1)}{\sum_{t=2}^{T} q_t^*} \right)^2$$
$$= \frac{[q_1^*(q_1 - 1)]^2}{1 - q_1^*},$$

hence

$$\frac{z}{1-z} \geqslant (q_1 - 1)^2 \left( q_1^* + \frac{(q_1^*)^2}{1 - q_1^*} \right) = (q_1 - 1)^2 \left( \frac{q_1^*}{1 - q_1^*} \right).$$

Since $0 < z \leqslant q_1^*$, this implies that $1 \geqslant (q_1 - 1)^2$, hence $q_1 \geqslant 0$. By a similar argument, $q_1 \geqslant 0, \ldots, q_T \geqslant 0$.

$(b) \rightarrow (c)$:

$$U(X) = \inf_{Q \in \mathcal{Q}^*} \mathbb{E}_Q[X] = \mathbb{E}_{Q^X}[X] \geqslant \mathbb{E}_{Q^X}[Y]$$
$$\geqslant \inf_{Q \in \mathcal{Q}^*} \mathbb{E}_Q[Y] = U(Y),$$

where $Q^X$ is an identifier of $X$.

$(c) \rightarrow (a)$: Fix $j$ such that $q_j^* = z$, and let $X_x$, $x \in \mathbb{R}$, be a one-parameter family in $\mathbb{R}^T$ such that $x_j = x$, and $x_t = -\frac{z}{1-z}$, $t \neq j$. Then $\mathbb{E}_{Q^*}[X_x] = q_j^* x_j + \sum_{t \neq j} q_t^* x_t = zx - \frac{z}{1-z}(1 - q_j^*) = z(x - 1)$, $\mathbb{E}_{Q^*}[X_x^2] = q_j^* x_j^2 + \sum_{t \neq j} q_t^* x_t^2 = x^2 z + \frac{z^2}{1-z}$, and

$$U(X_x) = z(x - 1) - \lambda \sqrt{x^2 z + \frac{z^2}{1-z} - z^2(x-1)^2}.$$

Consequently,

$$\frac{\partial}{\partial x} U(X_x) \bigg|_{x=1} = z - \lambda \sqrt{z(1-z)}.$$

Since $U(X_x)$ is non-decreasing in $x$ by (c), (23) follows.

## Proof of Proposition 5

Axioms 1, 2, and 3 follow from existence, continuity, and concavity of $U$ in (14), respectively. Under (23), axiom 5 follows from Proposition 4. Let also $q_1^* \leqslant \ldots \leqslant q_T^*$, non-constant $X, Y \in \mathbb{R}^T$ be as in axiom 8, and $Q^X = (q_1^X, \ldots, q_T^X)$ be an identifier of $X$. Then $Q^X = Q^* \left( 1 - \lambda \frac{X - \mathbb{E}_{Q^*}(X)}{\sigma_{Q^*}(X)} \right)$ (see [46, Example 1]), which implies that $q_t^X \leqslant q_j^X$ whenever $x_t \geqslant x_j$ and $q_t^* \leqslant q_j^*$. Consequently,

$$U(X) - U(Y) \geqslant \mathbb{E}_{Q^X}[X - Y] = \epsilon \left( q_j^X - q_t^X \right) \geqslant 0,$$

which proves axiom 8. Similarly, with $q_1^* = \cdots = q_T^*$, (23) implies axiom 6. Finally, with $q_1^* = \cdots = q_T^*$, (22) follows from (14).

# B Appendix: Data-Based Expected Utility Theory (EUT)

A data-based analogue of the independence axiom for r.v.'s is stated as follows.

**Axiom 7 (independence)** *Let $A$ and $B$ be any disjoint sets such that $A \cup B = \mathcal{T}$. For any $X, Y \in \mathcal{X}$ denote $X_A \oplus Y_B$ a function $z : \mathcal{T} \rightarrow F$ such that $z(t) = x(t)$ for $t \in A$ and $z(t) = y(t)$ for $t \in B$. Then $X_A \oplus Z_B \succeq Y_A \oplus Z_B$ implies that $X_A \oplus W_B \succeq Y_A \oplus W_B$ for any $X, Y, Z, W \in \mathcal{X}$.*

Axiom 7 implies that portfolios are compared only on the set where the corresponding data differs, whereas time periods with identical data are ignored.

A subset $\mathcal{T}' \subset \mathcal{T}$ is called inessential for $\succeq$ if $X \sim Y$ whenever $x(t) = y(t)$, $t \notin \mathcal{T}'$ (even if $x(t) \neq y(t)$, $t \in \mathcal{T}'$), and essential otherwise. In other words, decision making ignores data for inessential time periods.

**Proposition 6** *Assume that there is no inessential $\mathcal{T}'$ such that $\mathcal{T} \setminus \mathcal{T}'$ is a set of cardinality of at most 2. Then $\succeq$ satisfies axioms 1, 2, and 7 if and only if it has a continuous numerical representation $U$ in the form*

$$U(X) = U(x_1, \ldots, x_T) = \sum_{t=1}^{T} u_t(x_t) \qquad (32)$$

*for some utility functions $u_t : F \to \mathbb{R}$, $t = 1, \ldots, T$.*

**Proof** The "if" part is trivial, since axiom 7 follows directly from (32). The "only if" part follows from Debreu [17], who proved mathematically identical result with different economic interpretation. □

The condition in Proposition 6 excludes the scenario in which an agent selects at most two special time moments $t_1$ and $t_2$ in the past, and make all decisions based on the rates of return at these moments only. With $x_t = a_t + b_t \mathfrak{k}$, (32) can be equivalently written as

$$U(X) = \sum_{t=1}^{T} \widehat{u}_t(a_t, b_t), \qquad (33)$$

where $\widehat{u}_t : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$, $t = 1, \ldots, T$, are functions such that $\widehat{u}_t(a, b) \equiv u_t(a + b\mathfrak{k})$.

If all functions $u_t$ in (32) are concave, axiom 3 holds. Axiom 4 is equivalent to $\widehat{u}_t$ in (33) being linear in both argument, while axiom 5 is equivalent to $\widehat{u}_t$ being non-decreasing in the first argument, and non-increasing in a second argument, $t = 1, \ldots, T$.

*Risk aversion* is yet another cornerstone of decision making under uncertainty—given an r.v. $R_X$ and its expected payoff $\mathbb{E}[R_X]$, a risk averse agent always prefers the latter: $\mathbb{E}[R_X] \succeq'$ $R_X$, where $\succeq'$ is a preference relation defined on the space of r.v.'s. In particular, the risk averse agent rejects lotteries due to their negative expected payoff. In the context of the suggested approach, a more volatile time series may be considered to be more risky.

This intuition is not supported by axiom 6. If in (20) $x(t) = a(t) + b(t)\mathfrak{k}$ with $a(t_1) < a(t_2)$ for some $t_i \in \mathcal{T}_i$, $i = 1, 2$, then with $\delta = (a(t_2) - a(t_1))/2$ we have $a(t_1) + \delta I_{\mathcal{T}_1, \mathcal{T}_2}(t_1) = a(t_2) + \delta I_{\mathcal{T}_1, \mathcal{T}_2}(t_2)$, i.e., $x(t) + \delta I_{\mathcal{T}_1, \mathcal{T}_2}(t)$ might be "less risky" than $x(t)$. Axiom 6 suggests to ignore this fact and still to select $x(t)$ on the basis of better *recent* performance.

Principles of risk aversion and time aversion can be combined into a single axiom: if $X$ is "less risky" than $Y$ and also outperforms $Y$ more recently then $X \succeq Y$. Or, formally,

**Axiom 8 (time consistent risk aversion)** *Let $x(t) = a(t) + b(t)\mathfrak{k}$. If $\inf_{t \in \mathcal{T}_1} a(t) \geq \sup_{t \in \mathcal{T}_2} a(t)$ as well as $\sup_{t \in \mathcal{T}_1} b(t) \leq \inf_{t \in \mathcal{T}_2} b(t)$, then (20) and (21) hold.*

In the EUT (32)–(33), axiom 8 yields the following result.

**Proposition 7** *Let (33) with all concave $\widehat{u}_t$ be a continuous numerical representation of $\succeq$ on $\mathcal{X}$. Then $\succeq$ satisfies axiom 8 if and only if*

$$\frac{\partial \widehat{u}_t^+}{\partial a}(a, b) \leq \frac{\partial \widehat{u}_j^-}{\partial a}(a, b), \quad \frac{\partial \widehat{u}_t^-}{\partial b}(a, b) \geq \frac{\partial \widehat{u}_j^+}{\partial b}(a, b), \qquad (34)$$
$$\forall x = a + b\mathfrak{k} \in F \quad if \quad t < j,$$

*where superscripts "+"/"−" indicate right/left partial derivatives, respectively. In particular, (34) simplifies to*

$$\frac{\partial \widehat{u}_t}{\partial a}(a, b) \leq \frac{\partial \widehat{u}_j}{\partial a}(a, b) \quad and \quad \frac{\partial \widehat{u}_t}{\partial b}(a, b) \geq \frac{\partial \widehat{u}_j}{\partial b}(a, b), \qquad (35)$$
$$\forall x = a + b\mathfrak{k} \in F \quad if \quad t < j,$$

*if $\widehat{u}_1, \ldots, \widehat{u}_T$ are differentiable.*

**Proof** We will prove only the equivalence of (20) in axiom 8 and the first inequality in (34), the equivalence of (21) in the axiom and of the second inequality in (34) can be shown similarly. For any $t < j$, $a \in \mathbb{R}$, $b > 0$, and $\epsilon > 0$, let $X, Y \in \mathcal{X}$ be such that $x_t = x_j = a + b\mathfrak{k}$, $x_k = b\mathfrak{k}$, $k \notin \{t, j\}$, and $y_t = a + \epsilon$, $y_j = a - \epsilon$, $y_k = x_k$, $k \notin \{t, j\}$. By axiom 8, $X \succeq Y$, so that $U(X) \geq U(Y)$. By (33), this is equivalent to $\widehat{u}_t(a, b) + \widehat{u}_j(a, b) \geq \widehat{u}_t(a + \epsilon, b) + \widehat{u}_j(a - \epsilon, b)$, which implies that

$$\frac{\partial \widehat{u}_t^+}{\partial a}(a, b) = \lim_{\epsilon \to 0+} \frac{\widehat{u}_t(a + \epsilon, b) - \widehat{u}_t(a, b)}{\epsilon}$$
$$\leq \lim_{\epsilon \to 0+} \frac{\widehat{u}_j(a, b) - \widehat{u}_j(a - \epsilon, b)}{\epsilon} = \frac{\partial \widehat{u}_j}{\partial a}(a, b),$$

where both equalities and the existence of limits follow from concavity of $u_t$ and $u_j$.

Conversely, let (34) hold. In fact, we can assume that $|\mathcal{T}_1| = |\mathcal{T}_2| = 1$, and the first statement in axiom 8 can be formulated as: $X \succeq Y$, where $y_t = x_t + \delta$, $y_j = x_j - \delta$, and $y_k = x_k$, $k \notin \{t, j\}$ for some $\delta \geq 0$, $a_t \geq a_j$, $b_t \leq b_j$, and $t < j$. This statement follows from

$$\widehat{u}_t(a_t + \epsilon, b_t) - \widehat{u}_t(a_t, b_t) \leq \epsilon \frac{\partial \widehat{u}_t^+}{\partial a}(a_t, b_t) \leq \epsilon \frac{\partial \widehat{u}_t^+}{\partial a}(a_j, b_t)$$
$$\leq \epsilon \frac{\partial \widehat{u}_j^-}{\partial a}(a_j, b_t) \leq \epsilon \frac{\partial \widehat{u}_j^-}{\partial a}(a_j, b_j)$$
$$\leq \widehat{u}_t(a_j, b_j) - \widehat{u}_t(a_j - \epsilon, b_j).$$

Indeed, the first two inequalities follow from concavity of $u_t$, the third one from (34), and the last two from concavity of $u_j$. □

In the case with no missing data, (32) and (35) simplify to

$$U(X) = \sum_{t=1}^{T} u_t(x_t), \quad u_t : \mathbb{R} \to \mathbb{R}, \quad t = 1, \ldots, T, \qquad (36)$$

and

$$u_1'(x) \leq \ldots \leq u_T'(x) \quad \forall x \in \mathbb{R}, \qquad (37)$$

respectively. Utility (36) is similar to classical EUT [53], with the crucial difference that utilities are applied to the data directly and that there may be *different* utility functions for different time moments. The interpretation of (36) is that $u_t(x)$

is a numerical equivalent of the appeal for the historical rate of return $x$ of a given portfolio at time $t$. If $u_t$ is differentiable, $u_t'(x)$ measures sensitivity of investor's utility to changes in data at time $t$, and (37) represents the principle "recent observations are more important than past ones." If $u_t(x) = q_t x$, $t = 1, \ldots, T$, (36) and (37) simplify to (11) with a singleton $\mathcal{Q} = \{Q\}$ with $q_1 \leqslant \ldots \leqslant q_T$.

In the study of Christoforou [15], 62% of respondents agreed with axiom 7 and 58% with axiom 8.

# References

[1] https://uk.finance.yahoo.com/ (accessed on November 9, 2017).

[2] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–227, 1999.

[3] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton University Press, Princeton, NJ, 2009.

[4] D. Bertsimas, D. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM review*, 53(3):464–501, 2011.

[5] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[6] E. Bradley and R.J. Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

[7] S. Bragg. *IFRS made easy*. John Wiley & Sons, 2010.

[8] P. Brockwell and R. Davis. *Introduction to time series and forecasting*. springer, 2016.

[9] R. Brown. *Smoothing, forecasting and prediction of discrete time series*. Courier Corporation, 2004.

[10] G. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18(3):853–877, 2007.

[11] R. Casadesus-Masanell, P. Klibanoff, and E. Ozdenoren. Maxmin expected utility over savage acts with a set of priors. *Journal of Economic Theory*, 92(1):35–65, 2000.

[12] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and L. Montrucchio. Complete monotone quasiconcave duality. *Mathematics of Operations Research*, 36(2):321–339, 2011.

[13] L. Chan, J. Karceski, and J. Lakonishok. On portfolio optimization: Forecasting covariances and choosing the risk model. *Review of Financial Studies*, 12(5):937–974, 1999.

[14] A. Chekhlov, S. Uryasev, and M. Zabarankin. Portfolio optimization with drawdown constraints. *Asset and Liability Management Tools*, pages 263–278, 2000.

[15] A. Christoforou. Modelling investor's preferences in the direct data based model. Master's thesis, University of Leicester, 2015.

[16] J. Cox and S. Ross. The valuation of options for alternative stochastic processes. *Journal of financial economics*, 3(1-2):145–166, 1976.

[17] G. Debreu. Topological methods in cardinal utility theory. *Mathematical methods in the social sciences*, 1959:16–26, 1960.

[18] S. Dhami. *The Foundations of Behavioral Economic Analysis*. Oxford University Press, 2016.

[19] H. Föllmer and A. Schied. *Stochastic Finance*. de Gruyter, Berlin New York, 3 edition, 2011.

[20] I. Gilboa and M. Marinacci. Ambiguity and the bayesian paradigm. In *Readings in Formal Epistemology*, pages 385–439. Springer, 2016.

[21] I. Gilboa and D. Schmeidler. Maxmin expected utility with nonunique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

[22] I. Gilboa and D. Schmeidler. Case-based decision theory. *The Quarterly Journal of Economics*, 110(3):605–639, 1995.

[23] I. Gilboa and D. Schmeidler. *A theory of case-based decisions*. Cambridge University Press, 2001.

[24] A. Gorban, D. Rossiyev, and M. Dorrer. Multineuron-neural networks simulator for medical, physiological, and psychological applications. *World Congress on Neural Networks*, 1995.

[25] B. Grechuk, A. Molyboha, and M. Zabarankin. Mean-deviation analysis in the theory of choice. *Risk Analysis: An International Journal*, 32(8):1277–1292, 2012.

[26] B. Grechuk and M. Zabarankin. Inverse portfolio problem with coherent risk measures. *European Journal of Operational Research*, 249(2):740–750, 2016.

[27] B. Grechuk and M. Zabarankin. Inverse portfolio problem with mean-deviation model. *European Journal of Operational Research*, 234(2):481–490, 2014.

[28] B. Grechuk and M. Zabarankin. Sensitivity analysis in applications with deviation, risk, regret, and error measures. *SIAM Journal on Optimization*, to appear.

[29] F. Gul and W. Pesendorfer. Expected uncertain utility theory. *Econometrica*, 82(1):1–39, 2014.

[30] I. Kondor, S. Pafka, and G. Nagy. Noise sensitivity of portfolio selection under various risk measures. *Journal of Banking & Finance*, 31(5):1545–1573, 2007.

[31] S. Kou. A jump-diffusion model for option pricing. *Management science*, 48(8):1086–1101, 2002.

[32] V. Lesnevski, B. L. Nelson, and J. C. Staum. Simulation of coherent risk measures based on generalized scenarios. *Management Science*, 53(11):1756–1769, 2007.

[33] A. Löffler. Variance aversion implies $\mu$- $\sigma^2$-criterion. *Journal of Economic Theory*, 69(2):532–539, 1996.

[34] D. G. Luenberger. Investment science. *OUP Catalogue*, 1997.

[35] F. Maccheroni, M. Marinacci, A. Rustichini, and M. Taboga. Portfolio selection with monotone mean-variance preferences. *Mathematical Finance*, 19(3):487–521, 2009.

[36] D. Madan. Stochastic processes in finance. *Annual Review of Financial Economics*, 2(1):277–314, 2010.

[37] D. Madan and E. Seneta. The variance gamma (vg) model for share market returns. *Journal of business*, pages 511–524, 1990.

[38] M. Marinacci. Model uncertainty. *Journal of European Economic Association*, 13:998–1076, 2015.

[39] H. M. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[40] R. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics*, pages 247–257, 1969.

[41] R. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1-2):125–144, 1976.

[42] G. Pflug, A. Pichler, and D. Wozabal. The 1/n investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance*, 36(2):410–417, 2012.

[43] R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26(7):1443–1471, 2002.

[44] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Generalized deviations in risk analysis. *Finance and Stochastics*, 10(1):51–74, 2006.

[45] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Master funds in portfolio analysis with general deviation measures. *The Journal of Banking and Finance*, 30(2):743–778, 2006.

[46] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Optimality conditions in portfolio analysis with general deviation measures. *Mathematical Programming*, 108(2-3):515–540, 2006.

[47] K. Sato. Self-similar processes with independent increments. *Probability Theory and Related Fields*, 89(3):285–300, 1991.

[48] K. Sato. *Lévy processes and infinitely divisible distributions*. Cambridge university press, 1999.

[49] L. J. Savage. *The foundations of statistics*. Courier Corporation, 1972.

[50] A. Sheikh and H. Qiao. Non-normality of market returns: A framework for asset allocation decision making (digest summary). *Journal of Alternative Investments*, 12(3):8–35, 2010.

[51] J. Siegel. *Stocks for the long run*, volume 4th edition. McGraw-Hill New York, 2007.

[52] K. Sydsaeter, P. Hammond, and A. Seierstad. *Further mathematics for economic analysis*. Pearson Education, 2008.

[53] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 3 edition, 1953.

[54] John Von Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320, 1928.

[55] D. Wozabal. A framework for optimization under ambiguity. *Annals of Operations Research*, 193(1):21–47, 2012.

[56] M. Zabarankin, K. Pavlikov, and S. Uryasev. Capital asset pricing model (CAPM) with drawdown measure. *European Journal of Operational Research*, (234):508–517, 2014.

[57] M. Zabarankin and S. Uryasev. *Statistical Decision Problems: Selected Concepts and Portfolio Safeguard Case Studies*. Springer, Berlin, 2014.