Acta Mathematica Scientia, 2022, **42B**(6): 2377–2398
https://doi.org/10.1007/s10473-022-0611-0

Mathematica Acta Scientia
数学物理学报
http://actams.apm.ac.cn

UTILITY BASIS OF CONSUMPTION AND INVESTMENT DECISIONS IN A RISK ENVIRONMENT*

Dedicated to Professor Banghe LI on the occasion of his 80th birthday

Kangping WU (武康平)
School of Economics and Management, Tsinghua University, Beijing 100084, China
E-mail: *wukp@sem.tsinghua.edu.cn*

Abstract Using expectations regarding utilities to make decisions in a risk environment hides a paradox, which is called the expected utility enigma. Moreover, the mystery has not been solved yet; an imagined utility function on the risk-return plane has been applied to establish the mean-variance model, but this hypothetical utility function not only lacks foundation, it also holds an internal contradiction. This paper studies these basic problems. Through risk preference VNM condition is proposed to solve the expected utility enigma. How can a utility function satisfy the VNM condition? This is a basic problem that is hard to deal with. Fortunately, it is found in this paper that the VNM utility function can have some concrete forms when individuals have constant relative risk aversion. Furthermore, in order to explore the basis of mean-variance utility, an MV function is founded that is based on the VNM utility function and rooted in underlying investment activities. It is shown that the MV function is just the investor's utility function on the risk-return plane and that it has normal properties. Finally, the MV function is used to analyze the laws of investment activities in a systematic risk environment. In doing so, a tool, TRR, is used to measure risk aversion tendencies and to weigh risk and return.

Key words VNM condition; relative risk aversion tendency; mean-variance utility; systematic risk

2010 MR Subject Classification 91A30; 91B16

1 Introduction

As maximizing the expected return of an investment was negated by the St. Petersburg Paradox, it was proposed by D. Bernoulli to substitute expected utility for expected return. The utility index was further established by J. von Neumann. From this time awards, the maximization of expected utility has become the de facto practice in consumption and investment. Markovitz's portfolio selection (1952, see [13]) provides a larger application stage for the expected utility theory. In particular, the MV (mean-variance) approach has been applied

*Received July 15, 2022.

broadly in investment decision and financial analysis. Research results on this has been very numerous.

In practice, Markovitz's MV optimization may perform poorly. This phenomenon is referred to as the Markovitz optimization enigma. Lai et al. (2011, see [9]) studied the enigma and explained its root causes. They proposed a new approach of flexible modelling to resolve it. Van Staden et al. (2021, see [16]) found that the MV optimization can be remarkably robust for modelling misspecification errors in dynamic or multi-period portfolio optimization, in sharp contrast to single-period portfolio optimization settings. They explained the causes of this surprising robustness under both the pre-commitment MV and time-consistent MV approaches. Li et al. (2022, see [10]) employed model predictive control to construct a multi-period portfolio and to provide a comprehensive comparison of the models with regard to objective function choice, planning horizon and parameter estimation. They found that Markovitz portfolio optimization performs better in muti-period models than in single period ones.

In order to investigate the yield and price under uncertainty, Coyle (1999, see [5]) developed a duality model of production with risk aversion by using an MV approach. The model incorporates both mean-variance preferences and expected output supply, and is tractable for empirical research. Brown (2007, see [4]) gave a new application field for MV methodology. A mean-variance model is introduced in [4] to solve serial replacement problems with uncertain rewards. Liu (2022, see [11]) verified that Markowitz's asset portfolio theory is applicable for China's A-share market by randomly selecting its four stocks and constructing a portfolio of maximum Sharpe ratio.

Based on state-dependent risk aversion and efficient dynamic programming, Rainer (2022, see [15]) presented a heuristic mean-variance optimization in Markov decision processes to achieve a balance between maximizing expected rewards and minimizing risks. By using a CRRA utility function, Kassimatis (2021, see [7]) examined whether mean-variance is a good proxy for portfolios, and found that MV portfolios are a poor proxy for investors with CRRA preferences. Marianil et al. (2022, see [12]) proposed a measure for portfolio risk management by extending the Markowitz mean-variance approach to include the left-hand tail effects of asset returns. Two risk dimensions were captured: the asset covariance risk and the risk in left-hand tail similarity and volatility. From a simplified jump process, Khashanah et al (2022, see [8]) found that mean-variance portfolios need to be enhanced by incorporating higher-order components. Andrew et al. (2022, see [1]) investigated the impact of changes in the mean vector on mean-variance portfolio optimization. They found that the bounds of mean vector changes are unable to characterize portfolio sensitivity. Dai et al. (2020, see [6]) proposed a dynamic portfolio choice model with MV criteria for log-returns. Their consideration conform to investment common sense; for example, rich people should invest more in risky assets. The longer the investment period, the greater the proportion of investment in risky assets. For long-term investments, investors should not short sell major stock indices whose returns are higher than the risk-free rate.

Systematic risks and their impact on investment have attracted much attention. Berk and Tutarli (2020, see [2]) proposed two selection criteria for a mean variance optimization in a systematic risk environment: the beta coefficient and previous period return. In fact, the beta coefficient is a measure of systematic risk. Using these selection criteria, investors may obtain investable portfolios. Empirical analysis of the Istanbul Stock Exchange shows that the portfolio with the lowest beta coefficient is the best alternative. Bianconia et al. (2015, see [3]) introduced a measure of information dissemination for the determination of systemic risk. They found that VIX volatility has a significant direct impact upon the systemic risk of financial firms under distress. They also found that consumer pessimism can also predict systemic risk, and may be dominated by the VIX.

A common feature of all of the above research is the imagining of a utility function, which is used to derive an expectation regarding utilities. However, such a practice results in a paradox that is referred to as the expected utility enigma. Meanwhile, the mean-variance utility lacks foundation and hides its own contradictions. These problems are basic and intrinsic, and hide a danger which may lead to poor decisions regarding investment and consumption. In view of these problems, this paper starts by addressing the subject of risk preference, and makes an in-depth theoretical analysis. First, a VNM condition is proposed to explain the expected utility enigma. Second, the relative risk aversion tendency is used to explore the form of VNM utility functions and then to unveil the truth of the expected utility enigma. Third, a mean-variance utility function is constructed according to the VNM utility function and actual investment activities. Finally, using the MV utility function constructed, the law of investment decisions in systematic risks are revealed, and a new measure called the target rate of return is proposed in order to evaluate the risk aversion tendencies of investors.

Before starting the discussion, we explain some of the rules regarding the care of symbols in this paper. All vectors, matrices, and mappings whose values are vectors are expressed in italic bold letters. In addition, the word "function" refers to the mapping whose values are real numbers.

2 Expected Utility Enigma and VNM Condition

When individuals decide to consume or invest in an uncertain environment, they need first to know the expected utilities of their activities, and then to organize their activities according to the plan with maximal expected utility. Usually, the expected utility $Eu(\xi)$ of a random action ξ of an individual is given by the expectation that $Eu(\xi) = \int_\Omega u(\xi(\omega))\mathrm{d}P(\omega)$ is calculated from the utility function $u(x)$ of the individual. However, such an approach may lead to a contradiction, which is referred to as the expected utility enigma.

We can imagine a situation where a consumer chooses between two commodities X and Y. The amounts of X and Y are denoted by x and y. Suppose that $u(x,y) = (xy)^{0.25}$ and $v(x,y) = (xy)^{0.75}$ are the utility functions of the consumer. They display the same preference, i.e., the following fact holds for any $x_1 \geq 0, y_1 \geq 0$ and $x_2 \geq 0, y_2 \geq 0$:

$$(u(x_1,y_1) \leq u(x_2,y_2)) \Leftrightarrow (v(x_1,y_1) \leq v(x_2,y_2)).$$

The two functions obey the decreasing law of marginal utility: $u'_x(x,y) > 0, u'_y(x,y) > 0, u''_{xx}(x,y) < 0, u''_{yy}(x,y) < 0; v'_x(x,y) > 0, v'_y(x,y) > 0, v''_{xx}(x,y) < 0, v''_{yy}(x,y) < 0$. Thus $u(x,y)$ and $v(x,y)$ could be regarded as the cardinal utility functions used to calculate expected utilities.

Now suppose that the consumer is in an uncertain environment where their choice depends on the sides of a coin, while each side appears with a 50% probability. They are faced two options, A and B.

Option A: If the positive side appears, choose (1,1); otherwise, choose (3,3).

Option B: Always choose (2,2), no matter what side of the coin appears.

How does the consumer choose? Should they choose A or B? Facing this situation, the consumer needs first to calculate the expected utilities of the two options A and B. As $u(x,y)$ is a utility function of the consumer, they can obtain $Eu(A)$ and $Eu(B)$ as follows:

$$Eu(A) = 0.5u(1,1) + 0.5u(3,3) = 0.5 \times (1 + 9^{0.25}) \approx 1.366,$$
$$Eu(B) = 0.5u(2,2) + 0.5u(2,2) = u(2,2) = \sqrt{2} \approx 1.414.$$

Since $Eu(A) < Eu(B)$, the consumer should choose option B. However, $v(x,y)$ is also a utility function, so of course they can use $v(x,y)$ to calculate the expected utilities and obtain $Ev(A)$ and $Ev(B)$ as follows:

$$Ev(A) = 0.5v(1,1) + 0.5v(3,3) = 0.5 \times (1 + 9^{0.75}) \approx 3.098,$$
$$Ev(B) = 0.5v(2,2) + 0.5v(2,2) = v(2,2) = 4^{0.75} \approx 2.828.$$

Now $Ev(A) > Ev(B)$ means that the consumer should choose option A. The different answers here lead to a contradiction. Which option should be chosen? Is option A better than option B or is option B better than option A? This problem is the so-called expected utility enigma.

By this token, the underlying utility function used to calculate the expected utilities is a key factor that decides whether or not the decision is correct when individuals are in an uncertain environment. In order to solve the expected utilities enigma, the VNM condition can not be ignored; that is, the underlying utility function has to satisfy the VNM condition.

Before explaining the meaning of the VNM condition, let Ω denote the set of all natural states that affect the outcome of economic activities. The set Ω is called a state space. Let $\mathcal{F}$ denote the event field on Ω, which is a σ-field. Let $P : \mathcal{F} \to [0,1]$ denote the probability measure on $\mathcal{F}$. The probability space $(\Omega, \mathcal{F}, P)$ expresses that the economic environment is one of uncertainty.

Suppose that there are l kinds of commodities on the market. Then the commodity space becomes the l-dimensional Euclidean space R^l. The outcomes of economic activities are just vectors in R^l, called commodity vectors or selection schemes. However, not all vectors in R^l are available for individuals to choose, since activities are limited by some conditions. Let $\boldsymbol{S}$ denote the set of vectors from which individuals are allowed to choose. $\boldsymbol{S}$ is a subset of R^l, called outcome set. Usually, $\boldsymbol{S}$ is required to be a non-empty, convex and closed set.

For different outcomes, what is best? The answer depends on preferences. A rational individual's preference can be expressed as a binary relation $\preceq$ on the outcome set $\boldsymbol{S}$ obeying the following three axioms:

Reflexivity: $(\forall \boldsymbol{x} \in \boldsymbol{S})(\boldsymbol{x} \preceq \boldsymbol{x})$, i.e., every outcome is not better than itself.

Completeness: $(\forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{S})((\boldsymbol{x} \preceq \boldsymbol{y}) \vee (\boldsymbol{x} \succeq \boldsymbol{y}))$, so the individual knows good and bad.

Transitivity: $(\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \boldsymbol{S})(((\boldsymbol{x} \preceq \boldsymbol{y}) \wedge (\boldsymbol{y} \preceq \boldsymbol{z})) \Rightarrow (\boldsymbol{x} \preceq \boldsymbol{z}))$, so $\preceq$ is a good preordering.

For any $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{S}$, by $\boldsymbol{x} \sim \boldsymbol{y}$ we mean that $\boldsymbol{x} \preceq \boldsymbol{y}$ and $\boldsymbol{x} \succeq \boldsymbol{y}$; and by $\boldsymbol{x} \prec \boldsymbol{y}$ we mean that $\boldsymbol{x} \preceq \boldsymbol{y}$ and $\boldsymbol{x} \not\succeq \boldsymbol{y}$. When $\boldsymbol{x} \sim \boldsymbol{y}$, it is said that $\boldsymbol{x}$ and $\boldsymbol{y}$ are indifferent. When $\boldsymbol{x} \prec \boldsymbol{y}$, it is said that $\boldsymbol{x}$ is inferior to $\boldsymbol{y}$ (or $\boldsymbol{y}$ is preferred than $\boldsymbol{x}$).

In the uncertain environment $(\Omega, \mathcal{F}, P)$, the activity of the individual is really a random vector $\boldsymbol{\xi} : \Omega \to \boldsymbol{S}$; its outcome $\boldsymbol{\xi}(\omega) \in \boldsymbol{S}$ is affected by the natural state $\omega \in \Omega$. Let $\mathcal{S}$ denote the set of all random vectors on Ω, i.e., $\mathcal{S} = \{\boldsymbol{\xi} | \boldsymbol{\xi} : \Omega \to \boldsymbol{S}$ is a random vector$\}$. The set $\mathcal{S}$ is called a risky selection set or a risk set. As any vector $\boldsymbol{x} \in \boldsymbol{S}$ can be regarded as a degenerate random vector, $\boldsymbol{S}$ is contained in $\mathcal{S}$, i.e., $\boldsymbol{S} \subseteq \mathcal{S}$.

Any two activities $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathcal{S}$ can be compounded into an activity $p\boldsymbol{\xi} \oplus (1-p)\boldsymbol{\eta}$ by probability p in such a way that the activity is $\boldsymbol{\xi}$ with probability p, and $\boldsymbol{\eta}$ with probability $1-p$. $p\boldsymbol{\xi} \oplus (1-p)\boldsymbol{\eta}$ is called a compound activity. Using random events to express things, $p\boldsymbol{\xi} \oplus (1-p)\boldsymbol{\eta}$ means that the individual takes $\boldsymbol{\xi}$ if A happens, and takes $\boldsymbol{\eta}$ if A doesn't happen, where $A \in \mathcal{F}$ is an event with probability p. The compounding operation obeys obvious the following laws:

Commutative law: $p\boldsymbol{\xi} \oplus (1-p)\boldsymbol{\eta} = (1-p)\boldsymbol{\eta} \oplus p\boldsymbol{\xi}$ holds for any $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathcal{S}$ and $p \in [0,1]$.

Associative law: The following formula holds for any $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathcal{S}$ and $\alpha, p, q \in [0,1]$:

$$\begin{aligned}&\alpha(p\boldsymbol{\xi} \oplus (1-p)\boldsymbol{\eta}) \oplus (1-\alpha)(q\boldsymbol{\xi} \oplus (1-q)\boldsymbol{\eta})\\ =\ &(\alpha p + (1-\alpha)q)\boldsymbol{\xi} \oplus (\alpha(1-p) + (1-\alpha)(1-q))\boldsymbol{\eta}.\end{aligned}$$

Meanwhile the individual can judge which is better for any two random activities, $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathcal{S}$, according to their preference. This means that there is a reflexive, complete and transitive binary relation $\preceq^r$ on the risk set $\mathcal{S}$ such that $\preceq^r$ expresses the individual's risk preference. Since $\boldsymbol{S} \subseteq \mathcal{S}$, $(\boldsymbol{x} \preceq \boldsymbol{y}) \Leftrightarrow (\boldsymbol{x} \preceq^r \boldsymbol{y})$ should hold for any $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{S}$; that is, the evaluation on $\boldsymbol{S}$ is consistent under $\preceq$ and $\preceq^r$. Hence the preference $\preceq$ is a confinement of $\preceq^r$ on $\boldsymbol{S}$, i.e., $\preceq = \preceq^r|_S$. In other words, the risk preference $\preceq^r$ is an expansion of the outcome preference $\preceq$ to the risk set $\mathcal{S}$. On this account, for convenience, we use the same symbol $\preceq$ to denote both the outcome preference $\preceq$ and the risk preference $\preceq^r$.

Some general rules should be complied for the expanding of preferences from $\boldsymbol{S}$ to $\mathcal{S}$. For example, when $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{S}$ and $\boldsymbol{x} \prec \boldsymbol{y}$, the evaluation $q\boldsymbol{x} \oplus (1-q)\boldsymbol{y} \prec p\boldsymbol{x} \oplus (1-p)\boldsymbol{y}$ should hold for any $p, q \in [0,1]$ with $p < q$. This means that choosing the worse outcome with greater probability is worse than choosing the worse outcome with smaller probability. More generally, when $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathcal{S}$ and $\boldsymbol{\xi} \prec \boldsymbol{\eta}, (p < q) \Leftrightarrow (q\boldsymbol{\xi} \oplus (1-q)\boldsymbol{\eta} \prec p\xi \oplus (1-p)\boldsymbol{\eta})$ should hold for any $p, q \in (0,1)$. Another example is that $(\boldsymbol{x} \prec \boldsymbol{y}) \Leftrightarrow (p\boldsymbol{x} \oplus (1-p)\boldsymbol{z} \prec p\boldsymbol{y} \oplus (1-p)\boldsymbol{z})$ should hold for any $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \boldsymbol{S}$ and $p \in (0,1)$. In other words, when the outcomes chosen with probability $1-p$ are the same, choosing the worse outcome with probability p is worse than choosing the better outcome with probability p. This property is referred to as the independence of evaluation. More generally, $(\boldsymbol{\xi} \prec \boldsymbol{\eta}) \Leftrightarrow (p\boldsymbol{\xi} \oplus (1-p)\boldsymbol{\gamma} \prec p\boldsymbol{\eta} \oplus (1-p)\boldsymbol{\gamma})$ should hold for any $\boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\gamma} \in \mathcal{S}$ and $p \in (0,1)$.

These general rules for preference expansion are recognized and admitted with the following two axioms:

Independence Axiom: $(\boldsymbol{\xi} \preceq \boldsymbol{\eta}) \Leftrightarrow (p\boldsymbol{\xi} \oplus (1-p)\boldsymbol{\gamma} \preceq p\boldsymbol{\eta} \oplus (1-p)\boldsymbol{\gamma})$ for any $\boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\gamma} \in \mathcal{S}$ and $p \in (0,1)$.

Continuity Axiom: Both $\{p \in [0,1] : p\boldsymbol{\xi} \oplus (1-p)\boldsymbol{\eta} \preceq \boldsymbol{\gamma}\}$ and $\{p \in [0,1] : p\boldsymbol{\xi} \oplus (1-p)\boldsymbol{\eta} \succeq \boldsymbol{\gamma}\}$ are closed subsets of interval $[0,1]$ for any $\boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\gamma} \in \mathcal{S}$.

The following theorem makes a clear and intuitive interpretation of the above two axioms, and shows that the preference expansion according the two axioms does conform to general rules of evaluation:

Theorem 2.1 The risk preference $\preceq$ satisfies the Independence and Continuity Axioms if and only if $\preceq$ conforms to the following five general rules:

Rule (1) $(\boldsymbol{\xi} \sim \boldsymbol{\eta}) \Leftrightarrow (p\boldsymbol{\xi} \oplus (1-p)\boldsymbol{\gamma} \sim p\boldsymbol{\eta} \oplus (1-p)\boldsymbol{\gamma})$ holds for any $\boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\gamma} \in \mathcal{S}$ and $p \in (0,1)$;

Rule (2) $(\boldsymbol{\xi} \prec \boldsymbol{\eta}) \Leftrightarrow (p\boldsymbol{\xi} \oplus (1-p)\boldsymbol{\gamma} \prec p\boldsymbol{\eta} \oplus (1-p)\boldsymbol{\gamma})$ holds for any $\boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\gamma} \in \mathcal{S}$ and $p \in (0,1)$;

Rule (3) $(\boldsymbol{\xi} \prec \boldsymbol{\eta}) \Leftrightarrow (\boldsymbol{\xi} \prec p\boldsymbol{\xi} \oplus (1-p)\boldsymbol{\eta} \prec \boldsymbol{\eta})$ holds for any $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathcal{S}$ and $p \in (0,1)$;

Rule (4) $(p < q) \Leftrightarrow ((1-p)\boldsymbol{\xi} \oplus p\boldsymbol{\eta} \prec (1-q)\boldsymbol{\xi} \oplus q\boldsymbol{\eta})$ holds for any $p, q \in (0,1)$ and $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathcal{S}$ with $\boldsymbol{\xi} \prec \boldsymbol{\eta}$;

Rule (5) For any $\boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\gamma} \in \mathcal{S}$ with $\boldsymbol{\xi} \prec \boldsymbol{\gamma} \prec \boldsymbol{\xi}$, there exists a real $c \in (0,1)$ such that $(1-c)\boldsymbol{\xi} \oplus \boldsymbol{\eta} \sim \boldsymbol{\gamma}$.

Furthermore, the rules (4) and (5) above imply that $(1-a)\boldsymbol{\xi} \oplus a\boldsymbol{\eta} \prec \boldsymbol{\gamma} \prec (1-b)\xi \oplus b\boldsymbol{\eta}$ holds for any $a \in (0,c)$ and $b \in (c,1)$.

Proof It is easy to verify that $\preceq$ satisfies the Independence and Continuity Axioms if $\preceq$ conforms to the five general rules listed in the theorem, so we only need to prove the necessity. For this purpose, suppose that $\preceq$ satisfies the Independence and Continuity Axioms. Let $\boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\gamma} \in \mathcal{S}$ and $p, q \in (0,1)$ be given arbitrarily. From the Independence Axiom, rules (1) and (2) are obviously satisfied. In the following we prove rules from (3) to (5).

Proof of rule (3). Suppose that $\boldsymbol{\xi} \prec \boldsymbol{\eta}$. Note that $\boldsymbol{\xi} = (1-p)\boldsymbol{\xi} \oplus p\boldsymbol{\xi}$ and $\boldsymbol{\eta} = p\boldsymbol{\eta} \oplus (1-p)\boldsymbol{\eta}$. Form rule (2), $\boldsymbol{\xi} = (1-p)\boldsymbol{\xi} \oplus p\boldsymbol{\xi} \prec (1-p)\boldsymbol{\eta} \oplus p\boldsymbol{\xi} = p\boldsymbol{\xi} \oplus (1-p)\boldsymbol{\eta}$ and $p\boldsymbol{\xi} \oplus (1-p)\boldsymbol{\eta} \prec p\boldsymbol{\eta} \oplus (1-p)\boldsymbol{\eta} = \boldsymbol{\eta}$ hold. Rule (3) is proven.

Proof of rule (4). Here we know that $\boldsymbol{\xi} \prec \boldsymbol{\eta}$. To show the necessity in rule (4), suppose that $p < q$. Let $t = p/q$ and $\boldsymbol{\gamma} = (1-q)\boldsymbol{\xi} \oplus q\boldsymbol{\eta}$. Obviously, $0 < t < 1$. Since $\boldsymbol{\xi} \prec \boldsymbol{\eta}$, it is immediately derived from rule (3) that $\boldsymbol{\xi} \prec \boldsymbol{\gamma}$. Again from rule (3) we have $(1-t)\boldsymbol{\xi} \oplus t\boldsymbol{\gamma} \prec \boldsymbol{\gamma}$. Note that the following fact is true:

$$(1-t)\boldsymbol{\xi} \oplus t\boldsymbol{\gamma} = (1-t)\boldsymbol{\xi} \oplus t((1-q)\boldsymbol{\xi} \oplus q\boldsymbol{\eta}) = (1-t+t(1-q))\boldsymbol{\xi} \oplus tq\boldsymbol{\eta} = (1-p)\boldsymbol{\xi} \oplus p\boldsymbol{\eta}.$$

Thus it can be seen that $(1-p)\boldsymbol{\xi} \oplus p\boldsymbol{\eta} = (1-t)\boldsymbol{\xi} \oplus t\boldsymbol{\gamma} \prec \boldsymbol{\gamma} = (1-q)\boldsymbol{\xi} \oplus q\boldsymbol{\eta}$ holds. The necessity in rule (4) is proven.

Now we prove the sufficiency in rule (4). Suppose that $(1-p)\boldsymbol{\xi} \oplus p\boldsymbol{\eta} \prec (1-q)\boldsymbol{\xi} \oplus q\boldsymbol{\eta}$. The reflexivity of $\preceq$ implies that $p \neq q$. If $p > q$, then from the necessity in rule (3) we have that $(1-p)\boldsymbol{\xi} \oplus p\boldsymbol{\eta} \succ (1-q)\boldsymbol{\xi} \oplus q\boldsymbol{\eta}$. This is a contradiction, so $p > q$ cannot hold. Therefore $p < q$. The sufficiency in rule (4) is proven, and rule (4) is proven.

Proof of rule (5). Let $A = \{p \in [0,1] : (1-p)\boldsymbol{\xi} \oplus p\boldsymbol{\eta} \preceq \boldsymbol{\gamma}\}$ and $B = \{p \in [0,1] : (1-p)\boldsymbol{\xi} \oplus p\boldsymbol{\eta} \succeq \boldsymbol{\gamma}\}$. Then $A \cup B = [0,1]$. The continuity Axiom tells us that both A and B are closed subsets of interval $[0,1]$.

As we know that $\boldsymbol{\xi} \prec \boldsymbol{\gamma} \prec \boldsymbol{\eta}$, it can be seen that $0 \in A$ and $1 \in B$. Thus A and B are non-empty subsets of $[0,1]$. Now the connectivity of interval $[0,1]$ implies that $A \cap B \neq \Phi$. Hence there exists a real number $c \in A \cap B$. Obviously, $(1-c)\boldsymbol{\xi} \oplus c\boldsymbol{\eta} \sim \boldsymbol{\gamma}$ and $0 < c < 1$. Rule (5) is proven.

Furthermore, from rule (4), it can be found that $(1-a)\boldsymbol{\xi} \oplus a\boldsymbol{\eta} \prec \boldsymbol{\gamma} \prec (1-b)\boldsymbol{\xi} \oplus b\boldsymbol{\eta}$ holds for any $a \in (0,c)$ and $b \in (c,1)$. Theorem 2.1 is proven. □

Based on the above preparations and analyses, the expected utility enigma can now be solved. Note that the expectation $Eu(\boldsymbol{\xi}) = \int_\Omega u(\boldsymbol{\xi}(\omega))\mathrm{d}P(\omega)(\boldsymbol{\xi} \in \mathcal{S})$ defines a function $Eu : \mathcal{S} \to R$, which is an expansion of the underlying function $u : \boldsymbol{S} \to R$, i.e., $Eu|_S = u$. So

$Eu(\boldsymbol{x}) = u(\boldsymbol{x})$ for any $\boldsymbol{x} \in \boldsymbol{S}$. $u : \boldsymbol{S} \to R$ is said to be a utility function of the outcome preference $\preceq$ if $(\boldsymbol{x} \preceq \boldsymbol{y}) \Leftrightarrow (u(\boldsymbol{x}) \leq u(\boldsymbol{y}))$ holds for any $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{S}$. Utility functions are invariant under strictly increasing transformations, which says that if $u : \boldsymbol{S} \to R$ is a utility function of $\preceq$, then $v(\boldsymbol{x}) = \varphi(u(\boldsymbol{x}))(\boldsymbol{x} \in \boldsymbol{S})$ is a utility function of $\preceq$ too, where $\varphi : R \to R$ is a strictly increasing function.

We know that in an uncertain environment $(\Omega, \mathcal{F}, P)$, the individual evaluates activities according to their risk preferences $\preceq$. If they want to use an expectation $Eu(\boldsymbol{\xi})$ to evaluate things, $Eu(\boldsymbol{\xi})$ must be a utility function of the risk preference $\preceq$. This requirement is called the VNM condition on the underlying function $u : \boldsymbol{S} \to R$. The specific expression of this condition is as follows:

VNM condition: $u : \boldsymbol{S} \to R$ satisfies that $(\forall \boldsymbol{\xi}, \boldsymbol{\eta} \in \mathcal{S})((\boldsymbol{\xi} \preceq \boldsymbol{\eta}) \Leftrightarrow (Eu(\boldsymbol{\xi}) \leq Eu(\boldsymbol{\eta})))$.

When $u : \boldsymbol{S} \to R$ satisfies the VNM condition, $u : \boldsymbol{S} \to R$ is said to be a VNM function. It can be verified that if $u : \boldsymbol{S} \to R$ is a VNM function, then $u : \boldsymbol{S} \to R$ is a utility function of the outcome preference $\preceq$, so a VNM function is also referred to as the VNM utility function. It can be verified further that VNM functions are invariant under affine transformations. That is, if $u : \boldsymbol{S} \to R$ is a VNM function, then $v(\boldsymbol{x}) = a + bu(\boldsymbol{x})$ is also a VNM function too for any $a, b \in R$ with $b > 0$.

Up until now, the expected utility enigma has been solved by asking the underlying function to be a VNM function. When individuals make decisions in an environment with uncertainty, they must use a VNM function to evaluate things. A lack of VNM functions will inevitably lead to incorrect decisions.

3 Relative Risk Aversion Tendency and the VNM Function

After solving the expected utility enigma, there are two important questions that follow. One is whether the VNM functions exist. The other is how to identify a function as a VNM function. Fortunately, economics gives an answer to the first question, and tells us that there exist VNM functions if the preference $\preceq$ satisfies the following three conditions:

(1) $\Omega = \boldsymbol{S}$ and $\{\boldsymbol{x}\} \in \mathcal{F}$ for any $\boldsymbol{x} \in \boldsymbol{S}$;

(2) $\preceq$ satisfies the Independence and Continuity Axioms;

(3) $\preceq$ is measurable and inherited.

The first condition means that every outcome can be viewed as a natural state and appears randomly. The meaning of the second condition has been explained in Theorem 2.1. Third, saying that $\preceq$ is measurable means that both $\{\boldsymbol{x} \in \boldsymbol{S} : \boldsymbol{x} \preceq \boldsymbol{z}\}$ and $\{\boldsymbol{x} \in \boldsymbol{S} : \boldsymbol{x} \succeq \boldsymbol{z}\}$ are measurable sets for any $\boldsymbol{z} \in \boldsymbol{S}$. Finally, saying that $\preceq$ is inherited means that for any $\boldsymbol{x} \in \boldsymbol{S}$ and $\boldsymbol{\xi} \in \mathcal{S}$, $\boldsymbol{\xi} \preceq \boldsymbol{x}$ if $P\{\boldsymbol{\xi}(\omega) \preceq \boldsymbol{x}\} = 1$, and $\boldsymbol{\xi} \succeq \boldsymbol{x}$ if $P\{\boldsymbol{\xi}(\omega) \succeq \boldsymbol{x}\} = 1$. Thus the judgement $\boldsymbol{\xi} \preceq \boldsymbol{x}$ is an inheritance from the judgement that $\boldsymbol{\xi}(\omega) \preceq \boldsymbol{x}$ holds almost everywhere. It is obvious that these conditions are common. Thus VNM functions exist in general.

In addition, VNM functions must be cardinal utility functions. If this were not so, the expectation of utilities could be meaningless. As a result of the cardinal meaning of VNM functions, there exist cardinal utility functions. The existence puzzle of cardinal utilities is now solved.

In the past, it was generally recognized that utilities are hard to measure with a ruler. When you consume a certain quantity of goods, there is no way to know how much utility you obtain. In light of this, economists abandoned the cardinal utility assumption, and used instead ordinal utility theory. However economists are contradictory. While abandoning cardinal utilities, they again used cardinal utility functions to build dynamic or multi-period models such as business cycle models, economic growth models and financial models. In particular, they use cardinal utilities but do not know whether or not cardinal utilities exist. Thus their models are built like castles in the clouds. Now, with the help of VNM utilities, a positive answer is obtained for the existence puzzle regarding cardinal utilities, and thus a foundation is added for dynamic or multi-period models.

The above discussions about the existence of VNM and of cardinal utilities can be summarized in the following theorem:

Theorem 3.1 Suppose that $\boldsymbol{S}$ is a non-empty convex closed subset of space R^l, that $\Omega = \boldsymbol{S}$ and that $\{\boldsymbol{x}\} \in \mathcal{F}$ for any $\boldsymbol{x} \in \boldsymbol{S}$. If the preference $\preceq$ satisfies the Independence and Continuity Axioms, and is measurable and inherited, then there exist VNM functions, and there also exist cardinal utility functions of $\preceq$.

Now we consider the second question raised at the beginning of this section. In order to identify VNM functions, we start for analyzing risk aversion tendencies. In general, there are three kinds of attitudes towards risk: risk averse, risk love and risk neutral. An individual with risk preference $\preceq$ is called a risk averter if $E\boldsymbol{\xi} \succ \boldsymbol{\xi}$, a risk lover if $\boldsymbol{\xi} \succ E\boldsymbol{\xi}$, and risk neutral if $\boldsymbol{\xi} \sim E\boldsymbol{\xi}$, for any degenerated $\boldsymbol{\xi} \in \mathcal{S}$. In terms of gambling, a fair gamble is one in which the sum that is bet is equal to the expected return. Facing a fair gamble, risk averters reject it, but risk lovers accept the gamble. Risk neutrals are indifferent to fair gambling. The reality is that most people are risk averse. Only a small portion are risk neutral. Very few are risk lovers.

Let $v : \boldsymbol{S} \to R$ be a VNM function of the individual. It can be shown that $v : \boldsymbol{S} \to R$ is concave for risk averters, and convex for risk lovers. If they are risk neutral, then $v : \boldsymbol{S} \to R$ is a linear or one-order function. The function $v(\boldsymbol{x}) = v(x_1, x_2, \cdots, x_l)$ could be assumed to be twice differentiable and have non-zero first order derivatives. Under this assumption, $v''_{ii}(x) = (\partial^2 v(\boldsymbol{x}))/(\partial x_i^2)$ $(i = 1, 2, \cdots, l)$ are negative for risk averters, positive for risk lovers, and zero for risk neutrals. As a result, the decreasing marginal utility is equivalent to risk aversion, and so is verified to be a prevailing phenomenon.

With the help of the Arrow-Pratt coefficient of risk aversion, a measurement vector $\boldsymbol{\theta}(\boldsymbol{x})$ is found and proposed here for multi-variate function $v(\boldsymbol{x})$. The vector $\boldsymbol{\theta}(\boldsymbol{x})$, called a relative risk aversion tendency, is defined as follows:

$$\theta_i(\boldsymbol{x}) = \theta_i(x_1, x_2, \cdots, x_l) = -\frac{v''_{ii}(\boldsymbol{x})x_i}{v'_i(\boldsymbol{x})} \quad (\boldsymbol{x} \in \boldsymbol{S}, i = 1, 2, \cdots, l).$$

The economic meaning of $\theta_i(\boldsymbol{x})$ can be explained by a gambling plane. We can imagine a gamble designed by an event F with probability p. The amount of commodity i becomes $x_i(1+a)$ if F happens, and becomes $x_i(1 + b)$ if F doesn't happen, where a and b are percentages of changes in quantity. The amount x_j of other commodities $j (j \neq i)$ remain unchanged. This gamble can be denoted by (a, b), which is a point on the plane R^2 called a gambling plane, as shown in Figure 1. The origin $\mathbf{0}$ of the coordinate means no gambling.

For convenience, let $u(a) = v(x_1, \cdots, x_{i-1}, x_i(1 + a), x_{i+1}, \cdots, x_l)(a \in R)$. $u(\cdot)$ is the underlying utility function of the gamble. The expected return ER and the expected utility

Eu are as follows:

$$ER = ER(a,b) = px_i(1+a) + (1-p)x_i(1+b) = x_i(1+pa+(1-p)b),$$
$$Eu = Eu(a,b) = pu(a) + (1-p)u(b).$$

Fair gambles are those (a,b) with expected return x_i, i.e., $ER(a,b) = x_i$, so a gamble (a,b) is fair if and only if $pa(1-p)b = 0$. The line J consisting of all fair gambles is called fair gambling line, as shown in Figure 1.

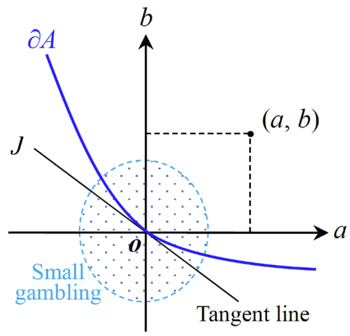

Figure 1 Risk averse acceptance set

The condition for the individual to accept a gamble (a,b) is that the expected utility $Eu(a,b)$ is not less than the utility $Eu(0,0) = u(0) = v(\boldsymbol{x})$ of no gambling. Let A be the set of all gambles accepted by the individual, i.e., $A = \{(a,b) \in R^2 : pu(a) + (1-p)u(b) \geq u(0)\}$, called the acceptance set. It can be shown that the acceptance set A of risk averters is convex. Figure 1 displays the shape of the acceptance set of a risk averter, where ∂A is the boundary of A. The boundary ∂A is determined by equation $pu(a) + (1-p)u(b) = u(0)$. Hence ∂A is the indifference curve through the origin $\boldsymbol{o}$; its slope at origin $\boldsymbol{o}$ is $-p/(1-p)$, which is just the slope of the fair gambling line J. Therefore J is just the tangent line of ∂A at the origin $\boldsymbol{o}$.

Let $b = \varphi(a)$ be the function determined by equation $pu(a)+(1-p)u(b) = u(0)$; i.e., $b = \varphi(a)$ describes ∂A. It can be shown that $\varphi'(0) = -p/(1-p)$ and $\varphi''(0) = -(u''(0)/u'(0))p/(1-p)^2$.

Assume that the individual is a risk averter, so that $u'(0) > 0$ and $u''(0) < 0$. Then $\varphi''(0) > 0$. Based on curvature theory, the larger the $\varphi''(0)$, the more curved the boundary ∂A, and so the more gambles near at origin $\boldsymbol{o}$ are rejected. As we know, gambles near at origin $\boldsymbol{o}$ are all small ones. Attitudes towards small gambles can reflect best the tendencies of risk aversion. Hence $\varphi''(0)$ measures the tendency of risk aversion. As $\varphi''(0)$ and $-u''(0)/u'(0)$ maintain a proportional relation, $-u''(0)/u'(0)$ can be regarded as a measure of the risk aversion tendency.

Calculating the derivatives of $u(a)$, it can be obtained that $u'(0) = v_i'(\boldsymbol{x})x_i$ and $u''(0) = v_{ii}''(\boldsymbol{x})x_i^2$, so $-u''(0)/u'(0) = -v_{ii}''(\boldsymbol{x})x_i/v_i'(\boldsymbol{x}) = \theta_i(\boldsymbol{x})$. This explains the significance of $\theta_i(\boldsymbol{x})$. Conforming with our expectation, $\theta_i(\boldsymbol{x})$ measures the risk aversion tendency on commodity i of the individual. Note that the changes of return in gambling are expressed by percentage changes, which are relative changes in quantity. For this reason, $\theta_i(\boldsymbol{x})$ is called a relative risk aversion tendency on commodity i of the individual.

Theorem 3.2 The relative risk aversion tendency $\boldsymbol{\theta}(\boldsymbol{x})$ depends only on the individual, not the forms of VNM functions of the individual. That is, if both $u : \boldsymbol{S} \to R$ and $v : \boldsymbol{S} \to R$ are

VNM functions of the individual, then $u''_{ii}(\boldsymbol{x})x_i/u'_i(\boldsymbol{x}) = v''_{ii}(\boldsymbol{x})x_i/v'_i(\boldsymbol{x})$ $(i = 1, 2, \cdots, l)$ holds for any $\boldsymbol{x} \in \boldsymbol{S}$.

Proof By cardinal utility, there is a rod scale or ruler in the mind of the individual used to measure the quantity of utilities. What is the scale of this ruler? This can be determined by a business plan. One foot can be defined as one meter long or 33.333 centimeters, etc., according to the plan. Once the length is determined, the ruler cannot be deformed or broken; otherwise, the measured quantity would be inaccurate. It does not matter whether long or short; what matters is that the ruler gives the reference unit of measurement, and then is used to mark the scale on a straight line. It does not matter either where the 0 mark is; as long as the 0 point is marked, the scales can be marked to the right or left.

One can use scales of different lengths, and adopt 0 points with different positions, to mark the scales on straight lines. The correspondence between any two different scales is really an affine transformation. Like the different scales with different positions of the 0 point, the correspondence between any two cardinal utility functions $u(\boldsymbol{x})$ and $v(\boldsymbol{x})$ is really an affine transformation too; that is, there exist real numbers $a > 0$ and b such that $v(\boldsymbol{x}) = au(\boldsymbol{x}) + b$.

Now suppose that $u : \boldsymbol{S} \to R$ and $v : \boldsymbol{S} \to R$ are VNM functions of the individual. They are cardinal utility functions, as VNM functions are cardinal ones. Thus there are real numbers $a > 0$ and b such that $v(\boldsymbol{x}) = au(\boldsymbol{x}) + b$ for any $\boldsymbol{x} \in \boldsymbol{S}$. This implies immediately that $u''_{ii}(\boldsymbol{x})x_i/u'_i(\boldsymbol{x}) = v''_{ii}(\boldsymbol{x})x_i/v'_i(\boldsymbol{x})$ holds for any $\boldsymbol{x} \in \boldsymbol{S}$ and $i = 1, 2, \cdots, l$.

For the above fact that $u(\cdot)$ can be affinely transformed to $v(\cdot)$, we can give a strict proof. As both $u(\cdot)$ and $v(\cdot)$ are VNM functions of the same individual, we have that for any $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathcal{S}$:

$$(\int_\Omega u(\boldsymbol{\xi}(\omega))\mathrm{d}P(\omega) \le \int_\Omega u(\boldsymbol{\eta}(\omega))\mathrm{d}P(\omega)) \Leftrightarrow (\int_\Omega v(\boldsymbol{\xi}(\omega))\mathrm{d}P(\omega) \le \int_\Omega v(\boldsymbol{\eta}(\omega))\mathrm{d}P(\omega)).$$

Let $A = \{Eu(\boldsymbol{\xi}) : \boldsymbol{\xi} \in \mathcal{S}\}$ and $B = \{Ev(\boldsymbol{\xi}) : \boldsymbol{\xi} \in \mathcal{S}\}$, where $Eu(\boldsymbol{\xi}) \triangleq \int_\Omega u(\boldsymbol{\xi}(\omega))\mathrm{d}P(\omega)$ and $Ev(\boldsymbol{\xi}) \triangleq \int_\Omega v(\boldsymbol{\xi}(\omega))\mathrm{d}P(\omega)$. It can be checked that for any $\alpha \in [0, 1]$, the distribution function of $(1-\alpha)\boldsymbol{\xi} \oplus \alpha\boldsymbol{\eta}$ is the weighted sum $(1-\alpha)f(\cdot) + \alpha g(\cdot)$ where $f(\cdot)$ and $g(\cdot)$ are the distribution functions of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ respectively. Thus $Eu((1-\alpha)\boldsymbol{\xi} \oplus \alpha\boldsymbol{\eta}) = (1-\alpha)Eu(\boldsymbol{\xi}) + \alpha Eu(\boldsymbol{\eta})$ for any $\alpha \in [0, 1]$. This implies that both A and B are convex subsets of the real line R.

Define $\varphi : A \to B$ as follows: $\varphi(Eu(\boldsymbol{\xi})) = Ev(\boldsymbol{\xi})(\boldsymbol{\xi} \in \mathcal{S})$. Obviously, $\varphi(\cdot)$ is increasing, and

$$\begin{aligned}\varphi((1-\alpha)Eu(\boldsymbol{\xi}) + \alpha Eu(\boldsymbol{\eta})) &= \varphi(Eu((1-\alpha)\boldsymbol{\xi} \oplus \alpha\boldsymbol{\eta})) \\ &= Ev((1-\alpha)\boldsymbol{\xi} \oplus \alpha\boldsymbol{\eta}) \\ &= (1-\alpha)Ev(\boldsymbol{\xi}) + \alpha Ev(\boldsymbol{\eta}) \\ &= (1-\alpha)\varphi(Eu(\boldsymbol{\xi})) + \alpha\varphi(Eu(\boldsymbol{\eta}))\ (\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathcal{S}).\end{aligned}$$

This shows that $\varphi : A \to B$ is convexly linear. Thus there exist $a, b \in R$ with $a > 0$ such that $\varphi(z) = az + b$ for any $z \in A$; i.e., $\varphi : A \to B$ is an affine transformation. Since $Eu(\boldsymbol{x}) = u(\boldsymbol{x})$ and $Ev(\boldsymbol{x}) = v(\boldsymbol{x})$ for any $\boldsymbol{x} \in \boldsymbol{S} \subseteq \mathcal{S}$, we have that $\varphi(u(\boldsymbol{x})) = \varphi(Eu(\boldsymbol{x})) = Ev(\boldsymbol{x}) = v(\boldsymbol{x})$ for any $\boldsymbol{x} \in \boldsymbol{S}$. Therefore $v(\boldsymbol{x}) = au(\boldsymbol{x}) + b$ for any $\boldsymbol{x} \in \boldsymbol{S}$; i.e., $u : \boldsymbol{S} \to R$ can be affinely transformed to $v : \boldsymbol{S} \to R$. Theorem 3.2 is proven. □

$\theta_i(\boldsymbol{x})$ changes with the change of x_i. What are the specific characteristics of such change? The specific answer to this is unknown, but because the stakes are in proportion, it seems that $\theta_i(\boldsymbol{x})$ has nothing to do with the size of x_i. In this case, it should be a good option to assume that $\theta_i(\boldsymbol{x})$ is constant. At least this option is acceptable and conforms to the principle of meeting change with constancy.

Theorem 3.3 Let $\boldsymbol{S} = R^l_+ = \{\boldsymbol{x} \in R^l : (x_1 \geq 0) \wedge (x_2 \geq 0) \wedge \cdots \wedge (x_l \geq 0))\}$ and $v : \boldsymbol{S} \to R$ be a twice differentiable VNM function of the individual with non-zero first order derivatives. Suppose that one unit of each commodity is a necessity, i.e., $(x_i = 1) \Rightarrow (v(\boldsymbol{x}) = 0)$ $(i = 1, 2, \cdots, l)$. Then the individual has a constant relative risk aversion tendency $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_l)$ (or $\boldsymbol{\theta}(\boldsymbol{x})$ is a constant vector $\boldsymbol{\theta}$) if and only if there is a constant real number $a > 0$ such that $v(\boldsymbol{x}) = a \prod\limits_{(i=1)}^{l} (x_i^{1-\theta_i} - 1)/(1-\theta_i)(\boldsymbol{x} \in \boldsymbol{S})$, where the value $(x_i^{1-\theta_i} - 1)/(1-\theta_i)$ at $\theta_i = 1$ can be defined supplementarily as $\ln(x_i)$ $(i = 1, 2, \cdots, l)$.

Proof We first explain why$(x_i^{1-\theta_i} - 1)/(1-\theta_i)$ can be viewed as $\ln(x_i)$ when $\theta_i = 1$. There are two reasons for this view. One is the $\lim\limits_{\theta_i \to 1} (x_i^{1-\theta_i} - 1)/(1-\theta_i) = \ln(x_i)$. Thus the supplementary definition gives the function continuity at $\theta_i = 1$. The other is the $\lim\limits_{\theta_i \to 1} \frac{\mathrm{d}}{\mathrm{d}x_i}[(x_i^{1-\theta_i} - 1)/(1-\theta_i)] = \frac{\mathrm{d}}{\mathrm{d}x_i} \ln(x_i)$. Hence the function with supplementary definition has continuous derivatives at $\theta_i = 1$.

Sufficiency From $v(\boldsymbol{x}) = a \prod\limits_{i=1}^{l} (x_i^{1-\theta_i} - 1)/(1-\theta_i)$ $(a > 0)$ and calculating the relative risk aversion tendency $\theta_i(\boldsymbol{x})$ on commodity i, it can be seen that $\theta_i(\boldsymbol{x}) = -v''_{ii}(\boldsymbol{x})x_i/v'_i(\boldsymbol{x}) \equiv \theta_i$. Thus the individual has a constant relative risk aversion tendency $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_l)$.

Necessity Suppose that the individual has a constant relative risk aversion tendency $\boldsymbol{\theta} \in R^l$. Then the following formulas hold for all $\boldsymbol{x} \in \boldsymbol{S}$ and $i = 1, 2, \cdots, l$:

$$\theta_i \equiv -\frac{v''_{ii}(\boldsymbol{x})x_i}{v'_i(\boldsymbol{x})} = -\frac{x_i}{v'_i(\boldsymbol{x})} \frac{\partial v'_i(\boldsymbol{x})}{\partial x_i}.$$

Note that the partial derivative $\partial v_i(\boldsymbol{x})/\partial x_i$ requires that the other $x_j (j \neq i)$ are unchanged. Hence $\partial(\cdot)/\partial x_i$ and $d(\cdot)/dx_i$ have the same meaning, so above formula can be written as

$$\theta_i = -\frac{x_i}{v'_i} \frac{\mathrm{d}v'_i}{\mathrm{d}x_i} = -\frac{\mathrm{d}v'_i/v'_i}{\mathrm{d}x_i/x_i} = -\frac{\mathrm{d}\ln|v'_i|}{\mathrm{d}\ln x_i}, \text{ or } \mathrm{d}\ln|v'_i| = \mathrm{d}\ln x_i^{-\theta_i}.$$

Starting from commodity 1, we recursively deduce things from above formula.

For commodity 1, the amounts $x_2, x_3, \cdots, x_l$ are arbitrary but settled, and $\mathrm{d}\ln|v'_1| = \mathrm{d}\ln x_1^{-\theta_1}$ holds for all $x_1 > 0$. Solving this equation, we have that $\ln|v'_1(\boldsymbol{x})| = \ln x_1^{-\theta_1} + C_{11}$, where C_{11} is a constant relative to x_1 but dependent on $x_2, x_3, \cdots, x_l$, i.e., $C_{11} = C_{11}(x_2, x_3, \cdots, x_l)$. Thus $|v'_1(\boldsymbol{x})| = e^{C_{11}} x_1^{-\theta_1}$. Solving this equation, we have that $|v(\boldsymbol{x})| = e^{C_{11}} x_1^{1-\theta_1}/(1-\theta_1) + C_{12}$, where C_{12} is still a constant relative to x_1 but dependent on $x_2, x_3, \cdots, x_l$, i.e., $C_{12} = C_{12}(x_2, x_3, \cdots, x_l)$. Note that $v(\boldsymbol{x}) = 0$ when $x_1 = 1$. This implies that $C_{12} = -e^{C_{11}}/(1-\theta_1)$ and that $v(\boldsymbol{x}) = \pm e^{C_{11}}(x_1^{1-\theta_1} - 1)/(1-\theta_1)$. Let $A_2 = \pm e^{C_{11}} = A_2(x_2, \cdots, x_l)$. Then $v(\boldsymbol{x}) = A_2(x_1^{1-\theta_1} - 1)/(1-\theta_1)$. We get the first conclusion as follows:

Conclusion 1 There is a constant A_2 relative to x_1 such that

$$v(\boldsymbol{x}) = \frac{x_1^{1-\theta_1}}{1-\theta_1} A_2(x_2, x_3, \cdots, x_l).$$

For commodity 2, the amounts are arbitrary but settled, and $\mathrm{d}\ln|v'_2(\boldsymbol{x})| = \mathrm{d}\ln x_2^{-\theta_2}$ holds for all $x_2 > 0$. From Conclusion 1, we have that $v'_2(\boldsymbol{x}) = [(x_1^{1-\theta_1} - 1)/(1-\theta_1)]A'_{22}$, where

$A'_{22} = \partial A_2/\partial x_2$. Thus $\mathrm{d}\ln x_2^{-\theta_2} = \mathrm{d}\ln|v'_2(\boldsymbol{x})| = \mathrm{d}\ln\left|A'_{22}(x_1^{1-\theta_1}-1)/(1-\theta_1)\right| = \mathrm{d}\ln|A'_{22}|$ holds for all $x_2 > 0$.

Using reasoning similar as to that for the commodity 1, there are constants C_{21} and C_{22} relative to x_1 and x_2, but depending on $x_3, x_4, \cdots, x_l$, such that $|A_2| = |A_2(x_2, x_3, \cdots, x_l)| = e^{C_{21}}x_2^{1-\theta_2}/(1-\theta_2) + C_{22}$, where $C_{21} = C_{21}(x_3, x_4, \cdots, x_l)$ and $C_{22} = C_{22}(x_3, x_4, \cdots, x_l)$. Note that $v(\boldsymbol{x}) = 0$ when $x_2 = 1$. Substituting $x_2 = 1$ into $v(\boldsymbol{x}) = A_2(x_1^{1-\theta_1}-1)/(1-\theta_1)$, we obtain that $0 = \left[e^{C_{21}}/(1-\theta_2) + C_{22}\right](x_1^{1-\theta_1}-1)/(1-\theta_1)$. Hence $C_{22} = -e^{C_{21}}/(1-\theta_2)$, and $|A_2| = e^{C_{21}}x_2^{1-\theta_2}/(1-\theta_2) + C_{22} = e^{C_{21}}(x_2^{1-\theta_2}-1)/(1-\theta_2)$. Let $A_3 = \pm e^{C_{21}}$. Then $A_2 = A_3(x_2^{1-\theta_2}-1)/(1-\theta_2)$, and we get the second conclusion as follows:

Conclusion 2 There is a relative constant A_3 such that

$$v(x) = \prod_{j=1}^{2}\frac{x_j^{1-\theta_j}-1}{1-\theta_j}A_3(x_3, \cdots, x_l).$$

Recursively using similar reasoning, we can get the conclusion for commodity i as follows:

Conclusion i There is a relative constant A_i such that

$$v(\boldsymbol{x}) = \prod_{j=1}^{i}\frac{x_j^{1-\theta_j}-1}{1-\theta_j}A_{i+1}(x_{i+1}, \cdots, x_l) \quad (i < l).$$

Finally, for commodity l, $\mathrm{d}\ln|v'_l(\boldsymbol{x})| = \mathrm{d}\ln x_l^{-\theta_l}$ holds for all $x_l > 0$, where $x_1, x_2, \cdots, x_{l-1}$ are settled. From conclusion $l-1$, we have $v'_l(\boldsymbol{x}) = \prod\limits_{i=1}^{l-1}\left[(x_i^{1-\theta_i}-1)/(1-\theta_i)\right]A'_{ll}(x_l)$, where $A'_{ll} = A'_{ll}(x_l) = \mathrm{d}A_l(x_l)/\mathrm{d}x_l$. Thus $\mathrm{d}\ln x_l^{-\theta_l} = \mathrm{d}\ln|v'_l(\boldsymbol{x})| = \mathrm{d}\ln|A'_{ll}|$ holds for all $x_l > 0$. Solving this equation, we obtain that $\ln|A'_{ll}| = \ln x_l^{-\theta_l} + C_{l1}$, so $|A'_{ll}| = e^{C_{l1}}x_l^{-\theta_l}$, where C_{l1} is constant relative to $x_1, x_2, \cdots, x_l$. Solving $|A'_{ll}| = e^{C_{l1}}x_l^{-\theta_l}$ there is a constant C_{l2} relative to $x_1, x_2, \cdots, x_l$ such that $|A_l| = e^{C_{l1}}x_l^{1-\theta_l}/(1-\theta_l) + C_{l2}$. Note that $v(\boldsymbol{x}) = 0$ when $x_l = 1$. Substituting $x_l = 1$ into $v(\boldsymbol{x}) = A_l\prod\limits_{i=1}^{l-1}(x_i^{1-\theta_i}-1)/(1-\theta_i)$, we obtain that $C_{l2} = -e^{C_{l1}}/(1-\theta_l)$ and that $|A_l| = e^{C_{l1}}(x_l^{1-\theta_l}-1)/(1-\theta_l)$. Sine $(1, 1, \cdots, 1)$ is the necessity vector, the constant $a = e^{C_{l1}} > 0$ is just one desired such that $v(\boldsymbol{x}) = A_l\prod\limits_{i=1}^{l-1}(x_i^{1-\theta_i}-1)/(1-\theta_i) = a\prod\limits_{i=1}^{l}(x_i^{1-\theta_i}-1)/(1-\theta_i)$. Theorem 3.3 is proven. □

Imitating the above proof, Theorem 3.3 can be generalized into a more general form, which are described in Theorem 3.4. As a preparation, here we explain the semi-orderings $\leq$ and $\ll$ on R^l. For any $\boldsymbol{x}, \boldsymbol{y} \in R^l, \boldsymbol{x} \leq \boldsymbol{y}$ means that $x_i \leq y_i$ $(i = 1, 2, \cdots, l)$; $\boldsymbol{x} < \boldsymbol{y}$ means that $\boldsymbol{x} \leq \boldsymbol{y}$ but that $\boldsymbol{x} \neq \boldsymbol{y}; \boldsymbol{x} \ll \boldsymbol{y}$ means that $x_i < y_i$ $(i = 1, 2, \cdots, l)$. The zero vector is denoted by the bold letter $\mathbf{0} = (0, 0, \cdots, 0)$.

Theorem 3.4 Let $\boldsymbol{S} = R_+^l, \boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_l) \gg \mathbf{0}$ and $v : \boldsymbol{S} \to R$, and be a twice differentiable VNM function of the individual with non-zero first order derivatives. Suppose that $(x_i = \mu_i) \Rightarrow (v(\boldsymbol{x}) = 0)$ holds for each i. Then the individual has a constant relative risk aversion tendency $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_l)$ if and only if there is a number $a > 0$ such that $v(\boldsymbol{x}) = a\prod\limits_{i=1}^{l}(x_i^{1-\theta_i} - \mu_i^{1-\theta_i})/(1-\theta_i)(\boldsymbol{x} \in \boldsymbol{S})$, where the value $(x_i^{1-\theta_i} - \mu_i^{1-\theta_i})/(1-\theta_i)$ at $\theta_i = 1$ can be defined supplementarily as $\ln(x_i/\mu_i)$ $(i = 1, 2, \cdots, l)$.

The proof of this theorem is similar to the one of Theorem 3.3. It will not be repeated here.

However the vector $\boldsymbol{\mu}$ is meaningful. For an individual with a constant relative risk aversion tendency, $\boldsymbol{\mu}$ satisfies that $v(\boldsymbol{\mu}) = 0$ and $v(\boldsymbol{x}) > 0$ when $\boldsymbol{x} \gg \boldsymbol{\mu}$. This means that $\boldsymbol{\mu}$ is a package of necessities, and a starting point for lives. $\boldsymbol{\mu} \gg \mathbf{0}$ indicates a high living standard. The individual has passed through the difficult stage of robbing Peter to pay Paul, so $\boldsymbol{\mu}$ signifies entering a well-off life. It is because of this well-off life that the individual could have a relative constant risk aversion tendency.

The relative constant risk aversion tendency $\theta_i(\boldsymbol{x})$ has another significant meaning. From its definition, it is easy to see that $\theta_i(\boldsymbol{x})$ is the elasticity of marginal utility, which represents the ratio of the marginal utility decline to the consumption increase. As usual, it is said that $\theta_i(\boldsymbol{x})$ is small if $\theta_i(\boldsymbol{x}) < 1$, large if $\theta_i(\boldsymbol{x}) > 1$, and appropriate if $\theta_i(\boldsymbol{x}) = 1$. Thus $\theta_i(\boldsymbol{x})$ is a measure for the sensitivity of marginal utility to consumption.

As marginal utility represents scarcity, the elasticity of marginal utility $\theta_i(\boldsymbol{x})$ can be referred to as scarcity elasticity. The smaller the scarcity elasticity, the less the impact of consumption on marginal utility, and the more necessary the commodity. Therefore, a small $\theta_i(\boldsymbol{x})$ implies that the commodity i is a necessity.

$\theta_i(\boldsymbol{x})$ also denotes the ratio of instantaneous speed to the average speed of diminishing marginal utility. When the instantaneous speed is less than the average speed, $\theta_i(\boldsymbol{x})$ is small. When the former is greater than the latter, $\theta_i(\boldsymbol{x})$ is large. When both are equal, $\theta_i(\boldsymbol{x})$ is appropriate. Therefore $\theta_i(\boldsymbol{x})$ is also a measure for the declining intensity of marginal utility.

In a word, the relative risk aversion tendency $\theta_i(\boldsymbol{x})$ has very significant meaning. It reflects both the scarcity elasticity and the declining intensity of marginal utility. The smaller $\theta_i(\boldsymbol{x})$, the weaker the tendency of relative risk aversion, the more necessary the commodity, and the weaker the declining intensity of marginal utility. The following theorem interprets the form of VNM utility functions of individuals with a weak tendency of relative risk aversion:

Theorem 3.5 Let $\boldsymbol{S} = R_+^l$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_l) \ll (1, 1, \cdots, 1)$. Suppose that $v : \boldsymbol{S} \to R_+$ is a twice differentiable VNM function of an individual with non-zero first order derivatives, and that for each i, $v(\boldsymbol{x}) = 0$ if $x_i = 0$. Then the individual has constant relative risk aversion tendency $\boldsymbol{\theta}$ if and only if there exists $a > 0$ such that $v(\boldsymbol{x}) = a \prod_{i=1}^{l} x_i^{1-\theta_i}$ $(\boldsymbol{x} \in \boldsymbol{S})$. Furthermore, when the individual has constant relative risk aversion tendency $\boldsymbol{\theta}$, $(v(\boldsymbol{x}) > 0) \Leftrightarrow (x \gg \mathbf{0})$ holds for any $\boldsymbol{x} \in \boldsymbol{S}$, and $v_i'(\boldsymbol{x}) > 0$ at $\boldsymbol{x} \gg \mathbf{0}$ for each $i = 1, 2, \cdots, l$.

Proof The sufficiency is easy to show by calculating derivatives; one only needs to show the necessity. For this purpose, let $-v_{ii}''(\boldsymbol{x})x_i/v_i'(\boldsymbol{x}) = \theta_i$ hold for all $\boldsymbol{x} \in \boldsymbol{S}$ and $i = 1, 2, \cdots, l$, so $\mathrm{d}\ln|v_i'| = \mathrm{d}\ln x_i^{-\theta_i}$ holds for all $x_i > 0$ with other $x_j's$ settled.

For $i = 1$, solving $\mathrm{d}\ln|v_i'| = \mathrm{d}\ln x_i^{-\theta_i}$ and using the fact that $v(\boldsymbol{x}) = 0$ when $x_1 = 0$ there exists a relative constant$A_2 = A_2(x_2, \cdots, x_l)$ such that $v(\boldsymbol{x}) = A_2 x_1^{1-\theta_1}$. For $i = 2$, from $\mathrm{d}\ln|v_2'| = \mathrm{d}\ln x_2^{-\theta_2}$ it can be obtained that $\mathrm{d}\ln|v_2'(\boldsymbol{x})| = \mathrm{d}\ln|A_{22}'| = \mathrm{d}\ln x_2^{-\theta_2}$, where $A_2' = \partial A_2/\partial x_2$. With similar reasoning, there is a relative constant $A_3 = A_3(x_3, \cdots, x_l)$ such that $A_2 = A_3 x_2^{1-\theta_2}$, so $v(\boldsymbol{x}) = A_3 x_1^{1-\theta_1} x_2^{1-\theta_2}$.

Recursively deducing from $i = 2$ to $i = l-1$, there is a relative constant $A_l = A_l(x_l)$ such that $v(\boldsymbol{x}) = A_l \prod_{i=1}^{l-1} x_i^{1-\theta_i}$. Now from $\mathrm{d}\ln|v_l'(\boldsymbol{x})| = \mathrm{d}\ln x_l^{-\theta_l}$ with the other x_i's settled, we have that $\mathrm{d}\ln|v_{ll}'| = \mathrm{d}\ln x_l^{-\theta_l}$. Solving this equation, there is a constant $a > 0$ such that $A_l = a x_l^{1-\theta_l}$.

Thus $v(\boldsymbol{x})=a\prod_{i=1}^{l}x_i^{1-\theta_i}$. The necessity and the theorem are proven. □

Theorems 3.3 to 3.5 have explained some relations between constant risk aversion tendency and forms of VNM functions, and have answered to certain extent the question of how to identify a function as a VNM function.

Coming back to the example of the expected utility enigma in Section 2, the truth can be eventually revealed. Two utility functions, $u(x,y)=(xy)^{0.25}$ and $v(x,y)=(xy)^{0.75}$ are given in the example for the individual to evaluate, but they get two contradict answers. Now, applying Theorem 3.5, it turns out that the truth is that if the relative risk aversion tendency is $\boldsymbol{\theta}=(0.75,0.75)$, then the VNM function is $u(x,y)$ rather than $v(x,y)$, and the individual should choose option B. If the relative risk aversion tendency is $\boldsymbol{\theta}=(0.25,0.25)$, then the VNM function is $v(x,y)$ rather than $u(x,y)$, and the individual should choose option A. If neither $(0.75,0.75)$ nor $(0.25,0.25)$ is the relative risk aversion tendency, then neither of the evaluations from the two functions is correct.

4 Re-establishing Mean-variance Utility

With the help of utility theory, Markovitz's mean-variance approach has been greatly developed. A utility function for mean and variance has been imagined, called the mean-variance utility function, or MV utility function for short. However this imagination hides two basic problems. One is similar to the enigma of expected utility without consideration of the VNM condition. The other is the MV utility function decoupling from underlying economic behavior. The value of the MV function is confused, because different behaviors can have the same mean-variance but different utilities. This may lead to paradoxes. Now we use the VNM condition to rebuild the MV utility function on the base of underlying behavior to solve the two problems. For this purpose, we first reexamine the mean-variance model for investment.

Let $(\Omega,\mathcal{F},P)$ denote a risky environment. The outcome of an investment is its return, usually expressed in terms of monetary revenue, which is random. Let $v:R\to R$ be the VNM utility function of an investor. Its value $v(x)$ denotes the utility amount of x units of revenue. Usually, the investor may face two of options. One is risk-free investment, such as in monetary assets, which are safe in terms of returns. The other is risky investment, such as securities, which are usually with uncertain returns. Generally, talking about investments means dealing with risks. Let ξ denote the return of an investment. It is a random variable. The mathematical expectation $r=E\xi$ expresses its expected return. The standard deviation $\sigma=\sqrt{\mathrm{Var}(\xi)}=\sqrt{E[(\xi-E\xi)^2]}$ expresses the risk of the investment. Whether the investor takes the investment activity ξ, it depends on weighing up of risk σ and return r; after all, high (or low) risk accompanies high (or low) expected return. The mean-variance model expresses how an investor weighs risks against expected returns.

Now assume R_f and R_m to be two options of an investor, where R_f is the rate of return of a risk-free investment, and R_m is the rate of a risky investment. We can view the sum to invest as one unit. Note that R_f is constant, but the R_m is random. $R_f=ER_f=r_f$ and $\sigma_f=\sqrt{\mathrm{Var}(R_f)}=0$; $R_m\neq ER_m=r_m$ and $\sigma_m=\sqrt{\mathrm{Var}(R_m)}>0$. High risk accompanying high return implies $r_m>r_f$. How much should the investor invest on the risky item R_m? It has long been said that one should not put all eggs in one basket. The investor might consider putting a proportion of their money in R_m, and putting another part in R_f. This proportion is the well-known β coefficient.

By R_β we denote the rate of the return of a portfolio with the coefficient $\beta : R_\beta = \beta R_m + (1-\beta)R_f$. The expected return is $r_\beta = ER_\beta = \beta r_m + (1-\beta)r_f$, and the standard deviation is $\sigma_\beta = \sqrt{E[(R_\beta - r_\beta)^2]} = \beta\sigma_m$. We have that $\beta = \sigma_\beta/\sigma_m$ and $r_\beta = r_f + ((r_m - r_f)/\sigma_m)\sigma_\beta$. This equation expresses the constraints on the risks and returns of portfolios, called the budget constraint of portfolios. This is displayed in Figure 2 as a straight line, called the budget line.

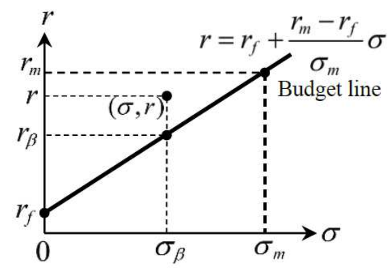

Figure 2 Risk-return plane

In the prevailing MV model, a utility function $U(\sigma, r)$ is imagined directly on the risk-return plane to be used to make a portfolio decision, but such an approach leaves the following basic questions unsolved:

(1) Where is the function from? What is its base? Can it be set up arbitrarily? Does it truly express the investor's objective function?

(2) Different investment activities may have different utility levels but the same mean-variance. This means that the value $U(\sigma, r)$ is not unique, and so is confused and leads to wrong decisions.

Now these questions can be solved by using the VNM condition. A solid theoretical foundation will be established for mean-variance modelling.

For question (1), it can be seen from the VNM condition that $U(\sigma, r)$ can not be set casually. Otherwise, wrong decisions could result. $U(\sigma, r)$ must be established on the VNM utility function $v : R \to R$ of the investor. Meanwhile this can not be decoupled from underlying investment activities. In order to bring out $U(\sigma, r)$, it is right to calculate the expected utility $Ev(\xi)$ of investment behavior ξ. Only in this manner of defining $U(\sigma, r)$ it becomes the objective function of the investor.

Question (2) is an extremely important issue arising inevitably from the process of solving question (1). To overcome it, the behavior considered is confined within the extent of normal random variables. However, as we know, it is hard for the investment return to obey normal distributions; the problem still exists, and up until now has not been fundamentally solved.

Here we shall put forward a method to solve the question (2) satisfactorily. The idea is based on the budget line to give every point (σ, r) of the risk-return plane a portfolio investment behavior such that the behavior corresponding to (σ, r) is decided uniquely, and has mean r and variance σ^2. The specific practices are as follows:

For any $\sigma \geq 0$ and $r \geq 0$, let $\xi(\sigma, r) = r + ((R_m - r_m)/\sigma_m)\sigma$. It is easy to verify that $\xi(\sigma, r)$ is a random variable and represents an investment activity with risk σ and expected return r; that is, we have that $E[\xi(\sigma, r)] = r$ and $\mathrm{Var}(\xi(\sigma, r)) = \sigma^2$.

The behavior $\xi(\sigma, r)$ has intuitive meanings. Let $R_\beta = \beta R_m + (1-\beta)R_f = r_f + \beta(R_m - R_f)$, where $\beta = \sigma/\sigma_m$. Then the following facts are true:

$$\begin{aligned}
r_\beta &= ER_\beta = r_f + \beta(r_m - r_f), \text{ i.e., } \beta(r_m - r_f) = r_\beta - r_f, \\
\sigma_\beta &= \sqrt{E[(R_\beta - r_\beta)^2]} = \beta\sigma_m = \sigma, \\
\xi(\sigma, r) &= r + [(R_m - r_m)/\sigma_m]\sigma = r + \beta(R_m - r_m) = \beta R_m + r - \beta r_m \\
&= \beta(R_m + r - r_\beta) - \beta(r - r_\beta) + r - \beta r_m \\
&= \beta(R_m + r - r_\beta) + (1-\beta)r - \beta(r_m - r_\beta) \\
&= \beta(R_m + r - r_\beta) + (1-\beta)r - \beta(r_m - (\beta r_m + (1-\beta)r_f)) \\
&= \beta(R_m + r - r_\beta) + (1-\beta)r - \beta((1-\beta)r_m - (1-\beta)r_f) \\
&= \beta(R_m + r - r_\beta) + (1-\beta)(r - \beta(r_m - r_f)) \\
&= \beta(R_m + r - r_\beta) + (1-\beta)(r - (r_\beta - r_f)) \\
&= \beta(R_m + r - r_\beta) + (1-\beta)(R_f + r - r_\beta).
\end{aligned}$$

Therefore, $\xi(\sigma, r) = \beta(R_m + r - r_\beta) + (1-\beta)(R_f + r - r_\beta)$. This shows that $\xi(\sigma, r)$ is a portfolio of the safe item $R_f + r - r_\beta$ and the risky item $R_m + r - r_\beta$ with the coefficient β. On the risk-return plane, as $\sigma_\beta = \sigma$, the corresponding point (σ, r) of $\xi(\sigma, r)$ is just the position to which the point (σ_β, r_β) representing portfolio R_β move upward, as shown in Figure 2. It can be proven that the correspondence between $\xi(\sigma, r)$ and (σ, r) is one-to-one, i.e., for any (σ_1, r_1) and (σ_2, r_2), we have that $((\sigma_1, r_1) = (\sigma_2, r_2)) \Leftrightarrow (\xi(\sigma_1, r_1) = \xi(\sigma_2, r_2))$. In addition, it can be seen that $\xi(\sigma_\beta, r_\beta) = R_\beta$ holds for any proportion β.

Note that $\xi(\sigma, r)$ represents the rate of return of the investment behavior (σ, r). Its return is $1 + \xi(\sigma, r)$, as the sum to invest is viewed as one unit. The expected utility is $Ev(1 + \xi(\sigma, r))$. Based on $Ev(1 + \xi(\sigma, r))$, a utility function $U(\sigma, r)$ can be defined on the risk-return plane R^2_+ as follows: $U(\sigma, r) \triangleq Ev(1 + \xi(\sigma, r))$ for any $(\sigma, r) \in R^2_+$.

Such $U(\sigma, r)$ are not only based on the VNM utility function $v : R \to R$, but also on the underling investment activity $\xi(\sigma, r)$. Therefore, this is truly the most objective function in which the investor can make investment decisions. Because of this, $U(\sigma, r)$ is called the mean-variance utility function, briefly, the MV utility function or the MV function. The following two theorems explain the characteristics of the MV functions:

Theorem 4.1 The MV function $U(\sigma, r)$ defined above is strictly concave for risk averters, strictly convex for risk lovers, and linear or of one order for risk neutrals.

Proof Let $(\sigma_1, r_1), (\sigma_2, r_2) \in R^2_+$ and $\alpha \in (0, 1)$ be arbitrarily given with $(\sigma_1, r_1) \neq (\sigma_2, r_2)$. We have that

$$\begin{aligned}
\alpha U(\sigma_1, r_1) + (1-\alpha)U(\sigma_2, r_2) &= E[\alpha v(1 + \xi(\sigma_1, r_1)) + (1-\alpha)v(1 + \xi(\sigma_2, r_2))], \\
U(\alpha(\sigma_1, r_1) + (1-\alpha)(\sigma_2, r_2)) &= Ev(1 + \xi(\alpha(\sigma_1, r_1) + (1-\alpha)(\sigma_2, r_2))).
\end{aligned}$$

When the investor is risk averse, their VNM utility function $v : R \to R$ is strictly concave, so we have that $E[\alpha v(1+\xi(\sigma_1, r_1)) + (1-\alpha)v(1+\xi(\sigma_2, r_2))] < Ev(1+\xi(\alpha(\sigma_1, r_1) + (1-\alpha)(\sigma_2, r_2)))$, and hence $\alpha U(\sigma_1, r_1) + (1-\alpha)U(\sigma_2, r_2) < U(\alpha(\sigma_1, r_1) + (1-\alpha)(\sigma_2, r_2))$; i.e., $U(\sigma, r)$ is strictly concave.

When the investor is a risk lever, their VNM utility function $v : R \to R$ is strictly convex, so we have that $E[\alpha v(1+\xi(\sigma_1, r_1)) + (1-\alpha)v(1+\xi(\sigma_2, r_2))] > Ev(1+\xi(\alpha(\sigma_1, r_1) + (1-\alpha)(\sigma_2, r_2)))$, and hence $\alpha U(\sigma_1, r_1) + (1-\alpha)U(\sigma_2, r_2) > U(\alpha(\sigma_1, r_1) + (1-\alpha)(\sigma_2, r_2))$; i.e., $U(\sigma, r)$ is strictly convex.

When the investor is risk neutral, their VNM utility function $v : R \to R$ is of one order or linear, so we have that $E[\alpha v(1+\xi(\sigma_1, r_1)) + (1-\alpha)v(1+\xi(\sigma_2, r_2))] = Ev(1+\xi(\alpha(\sigma_1, r_1) + (1-\alpha)(\sigma_2, r_2)))$, and hence $\alpha U(\sigma_1, r_1) + (1-\alpha)U(\sigma_2, r_2) = U(\alpha(\sigma_1, r_1) + (1-\alpha)(\sigma_2, r_2))$; and thus i.e., $U(\sigma, r)$ is of one order or linear.

The theorem is proven. □

The next theorem reveals the characteristics for $U(\sigma, r)$ to reflect the phenomenon of high risks accompanying high returns. This is important for investors to weigh risks against returns.

Theorem 4.2 Let the VNM function $v : R \to R$ be twice differentiable and $v'(x) > 0$ for all $x \in R$.

(1) If the investor is risk averse, then $U'_\sigma(\sigma, r) < 0$ and $U'_r(\sigma, r) > 0$ for all $(\sigma, r) \in R^2_+$, so the portfolio with low risk and high return is better than that with high risk and low return.

(2) If the investor is a risk lover, then $U'_\sigma(\sigma, r) > 0$ and $U'_r(\sigma, r) > 0$ for all $(\sigma, r) \in R^2_+$, so the portfolio with high risk and high return is better than that with low risk and low return.

(3) If the investor is risk neutral, then $U'_\sigma(\sigma, r) = 0$ and $U'_r(\sigma, r) > 0$ for all $(\sigma, r) \in R^2_+$, so the portfolio with high return is better than that with low return without regard to risk.

Proof $\xi(\sigma, r) = r + [(R_m - r_m)/\sigma_m]\sigma$ and $U(\sigma, r) = Ev(1 + \xi(\sigma, r))$. Calculating the partial derivatives, we can obtain that

$$
\begin{aligned}
U'_r(\sigma, r) &= \frac{\partial U(\sigma, r)}{\partial r} = E\left[v'(1+\xi(\sigma, r))\frac{\partial \xi(\sigma, r)}{\partial r}\right] = Ev'(1+\xi(\sigma, r)) > 0, \\
U'_\sigma(\sigma, r) &= \frac{\partial U(\sigma, r)}{\partial \sigma} = E\left[v'(1+\xi(\sigma, r))\frac{\partial \xi(\sigma, r)}{\partial \sigma}\right] = E\left[v'(1+\xi(\sigma, r))\frac{R_m - r_m}{\sigma_m}\right] \\
&= \mathrm{Cov}\left(v'(1+\xi(\sigma, r))\frac{R_m - r_m}{\sigma_m}\right) + Ev'(1+\xi(\sigma, r))E\left[\frac{R_m - r_m}{\sigma_m}\right] \\
&= \mathrm{Cov}\left(v'(1+\xi(\sigma, r))\frac{R_m - r_m}{\sigma_m}\right) \quad \left(\text{Because } E\left[\frac{R_m - r_m}{\sigma_m}\right] = 0\right).
\end{aligned}
$$

The symbol "Cov" above means covariance. Note that $\xi(\sigma, r)$ and $(R_m - r_m)/\sigma_m$ are positively related random variables; both become larger or smaller at the same time.

If the investor is risk averse, then $v''(x) < 0$ or $v'(x)$ changes inversely with x, so $v'(1 + \xi(\sigma, r))$ changes inversely with $(R_m - r_m)/\sigma_m$, and thus $\mathrm{Cov}(v'(1+\xi(\sigma, r)), (R_m - r_m)/\sigma_m) < 0$, and we obtain that $U'_\sigma(\sigma, r) < 0$.

If the investor is a risk lover, then $v''(x) > 0$, so $v'(1 + \xi(\sigma, r))$ changes in the same direction as $(R_m - r_m)/\sigma_m$, and thus $\mathrm{Cov}(v'(1 + \xi(\sigma, r)), (R_m - r_m)/\sigma_m) > 0$, and we obtain that $U'_\sigma(\sigma, r) > 0$.

If the investor is risk neutral, then $v''(x) = 0$ or $v'(x)$ is constant, so $v'(1 + \xi(\sigma, r))$ is independent on $(R_m - r_m)/\sigma_m$, and thus $\mathrm{Cov}(v'(1 + \xi(\sigma, r)), (R_m - r_m)/\sigma_m) = 0$, and we obtain that $U'_\sigma(\sigma, r) = 0$.

The theorem is proven. □

The characteristics of mean-variance utility revealed by Theorems 4.1 and 4.2 are displayed in Figure 3, where the shapes of the indifference curves are depicted separately for the risk averter, risk the lover and the risk neutral.

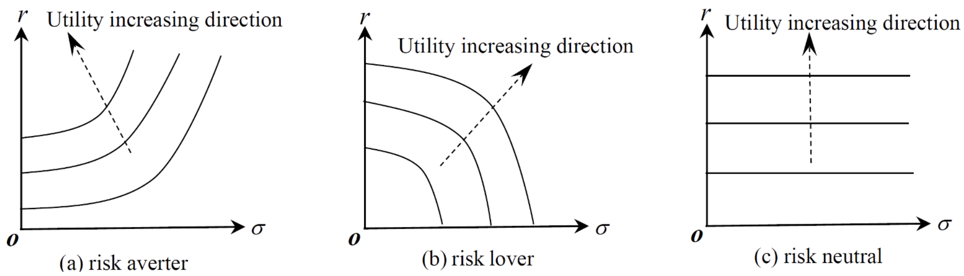

Figure 3 Indifference curves under mean-variance utility function

For an investor with constant relative risk aversion tendency θ, it can be shown that they have an VNM utility function with the form $v(x) = a(x^{1-\theta} - 1)/(1-\theta) + b$, where $a > 0$ and b are constants. This function may be further taken as $v(x) = (x^{1-\theta} - 1)/(1-\theta)$, since VNM functions have "invariance" under affine transformations.

Let $\varphi_m(x)$ be the density function of the distribution of the risky return R_m. For the investor with constant relative risk aversion tendency θ, the MV utility function $U(\sigma, r) = Ev(1 + \xi(\sigma, r))$ can be written as follows:

$$U(\sigma, r) = \frac{1}{1-\theta}\left(\int_{-\infty}^{+\infty}(1 + r + \frac{\sigma}{\sigma_m}(x - r_m))^{1-\theta}\varphi_m(x)\mathrm{d}x - 1\right).$$

If $\theta < 1$, then $U(\sigma, r)$ can be written as $U(\sigma, r) = \int_{-\infty}^{+\infty}(1 + r + \frac{\sigma}{\sigma_m}(x - r_m))^{1-\theta}\varphi_m(x)\mathrm{d}x$. If $\theta = 1$, then $U(\sigma, r)$ can be written as $U(\sigma, r) = \int_{-\infty}^{+\infty}\ln(1 + r + \frac{\sigma}{\sigma_m}(x - r_m))\varphi_m(x)\mathrm{d}x$.

These kinds of concrete forms of mean-variance utility functions might be useful in econometrics. In particular, they are convenient for quantitative analyses.

5 Investment Decision under Systematic Risks

Systematic risk refers to the fluctuation of the whole economic and financial system due to external or internal factors, which cause a series of continuous losses. No individuals are spared and anyone can suffer losses. This kind of risk can not be dispersed and can not be eliminated by investment diversification. Hence Markowitz's method of maximizing profits and minimizing risks is invalid here.

There are many factors that cause systematic risks, including political factors, policy factors, economic factors, social factors, environmental factors, and large-scale natural or man-made disasters (such as the COVID-19 epidemic), etc.. Once systematic risk occurs, all kinds of investment activities will be seriously affected. This means that in the systematic risk environment, the returns of various investment activities will show the same characteristics of rise and fall. Combining different risk options can not solve the risk. Only by following the logic of the mean variance model and choosing the optimal β coefficient can the loss be minimized. As such, it seems important to hold a certain percentage of safe assets.

Assume that the investor is in a systematic risk environment, and their VNM function is $v(x)$, where $x \in R$. They are faced with two options for investment activities. One option is to hold safe asset R_f. This could be viewed as holding a monetary asset, whose return is constant and not affected by the systematic risks. The other option is to hold risky asset R_m, which could be viewed as holding securities. The return of the risky asset is affected by systematic risks, and so is a random variable. Although risky assets are various, their return rates rise or fall simultaneously. Any one of them could be chosen as a representor. Let R_m be the chosen representor of all risky investment activities, and let R_f be the representor of safe assets. As we did in last section, R_m and R_f express the rates of returns. R_f is constant, but R_m is random. The sum for the investor to invest is viewed as one unit. Thus R_m and R_f represent both rates and net returns. The investor's MV function is then $U(\sigma,r) = Ev(1+\xi(\sigma,r))((\sigma,r) \in R_+^2)$, where $\xi(\sigma,r)$ is defined as in last section: $\xi(\sigma,r) = r + [(R_m - r_m)/\sigma_m]\,\sigma$.

No matter what happens, the object of the investor is to maximize his MV utility within the budget constraint of portfolios. Let (σ^*,r^*) be the optimal combination of risk and return; that is, (σ^*,r^*) is the solution of the following maximization problem of mean-variance utilities:

$$\begin{cases} \max\limits_{\sigma,r} U(\sigma,r), \\ r = r_f + \dfrac{r_m - r_f}{\sigma_m}\sigma. \end{cases}$$

Assume that the VNM function $v(x)(x \in R)$ is twice differentiable with positive first order derivatives. Then there is a Lagrange multiplier λ such that (σ^*,r^*) satisfies the condition that

$$\begin{cases} U_r'(\sigma^*,r^*) = \lambda, \\ U_\sigma'(\sigma^*,r^*) = -\lambda(r_m - r_f)/\sigma_m, \\ r^* = r_f + \sigma^*(r_m - r_f)/\sigma_m. \end{cases}$$

This condition is called the first order MV condition, and it can be also written as follows:

$$\begin{cases} \dfrac{r_m - r_f}{\sigma_m} = -\dfrac{U_\sigma'(\sigma^*,r^*)}{U_r'(\sigma^*,r^*)}, \\ r^* = r_f + \sigma^*(r_m - r_f)/\sigma_m. \end{cases}$$

The first order MV condition involves two tools. One is the slope $\pi = (r_m - r_f)/\sigma_m$ of the budget line. The other is the slope $\rho = \rho(\sigma,r) = -U_\sigma'(\sigma,r)/U_r'(\sigma,r)$ of the indifference curve. The two slopes have very significant implications for investment decision making.

(1) Actual Rate of Return (ARR) $\pi = (r_m - r_f)/\sigma_m$. As the slope of the budget line, π denotes the amount by which the expected return rate of the portfolio can increase, when the risk of portfolio is increased by one unit. This amount is determined by the budget line, so is a real amount that cannot be artificially changed. Therefore, it called the ARR of risk.

(2) Target Rate of Return (TRR) $\rho = \rho(\sigma,r) = -U_\sigma'(\sigma,r)/U_r'(\sigma,r)$. As the slope of the indifference curve, ρ denotes the amount by which the expected rate of return should be increased to keep the utility level constant when the risk at (σ,r) is increased by one unit. This amount is the goal that investors hope to achieve for the risk-return rate. Therefore, it is called the TRR of risk.

(3) Decision Principle (DP). After increasing the risk, if the ARR reaches the TRR, the utility level will remain unchanged; if the ARR exceeds the TRR, the utility level will rise;

if the ARR fails to reach the TRR, the utility level will drop. This implies the principle of investment decision which tell us that at the current point (σ, r), we should increase the risk if ARR>TRR, reduce the risk if ARR<TRR, and remain unchanged if ARR=TRR. The utility level reaches the highest point when ARR=TRR.

In what follows, we use the ARR and the TRR to explain the laws of investment activities in a systematic risk environment. Before the discussion, we classify systematic risks according to their degree of influence. In fact, systematic risks mainly affect investor's expectation of returns. When $ER_m < R_f$, it is viewed as serious influence. When $ER_m > R_f$, it is viewed as heavy influence. When $ER_m = R_f$, it is viewed as light influence.

Case 1 Serious influence of systematic risk, $r_m = ER_m < R_f = r_f$.

In this case, the ARR is negative. For risk averters and risk neutrals, TRR>ARR at every point (σ, r) of the budget line; thus they do not invest in risky item R_m and will put all of their money in safe item R_f. As they make up the overwhelming majority, social investment activities will be extremely depressed.

Even for risk lovers, if TRR≥ARR at (σ_m, r_m), they do not invest in R_m either. Instead, they will put all of their money in safe item R_f. This makes social investment situation more severe.

Case 2 Heavy influence of systematic risk, $r_m = ER_m = R_f = r_f$.

In this case, the ARR is zero. For risk averters, TRR>ARR at every point (σ, r) of the budget line, they will put all of their money in safe item R_f. Because risk averters are the majority, social investment activities are heavily depressed and very low.

For risk neutrals, TRR=ARR at every point (σ, r) of the budget line, and they do not care what choice they make. Some risk neutrals may invest in R_m, some may not.

For risk lovers, TRR<0=ARR at every point (σ, r) of the budget line, so they will invest all of their money in R_m. However, as they are very few, their investment activities can not improve the grim investment situation.

Case 3 Light influence of systematic risk, $r_m = ER_m > R_f = r_f$.

In this case, although systematic risk has caused adverse effects leading to a decline of expected rate of return r_m, the influence is light, so the expected rate of return r_m is still higher than r_f. Thus the ARR is still positive.

For risk neutrals and risk lovers, their TRR at every point (σ, r) of the budget line is non-positive, and so is less than the ARR. Thus they will definitely invest all of their money in R_m.

For risk averters, so long as TRR<ARR at initial point $(0, r_f)$ of the budget line, they will certainly invest some or all of their money in R_m.

In summary, those risk averters whose TRR at $(0, r_f)$ are less than the ARR, along with all risk lovers and risk neutrals, make up a quite large part of those who have an investment in R_m. Therefore, in the case that the influence of systematic risk is light, the situation for social investment activities is not so bad; quite a few people are still engaged in investment activities.

Why do those risk averters whose TRR at $(0, r_f)$ are equal to or greater than the ARR not invest in item R_m? In order to analyze this, let us reveal another profound implication of $\mathrm{TRR} = \rho(\sigma, r)$ for risk averters. From the calculation of $U'_r(\sigma, r)$ and $U'_\sigma(\sigma, r)$ in the proof of

Theorem 4.2, we have that

$$U'_r(\sigma,r) = Ev'(1+\xi(\sigma,r)),$$
$$U'_\sigma(\sigma,r) = \mathrm{Cov}(v'(1+\xi(\sigma,r)),(R_m-r_m)/\sigma_m) = \mathrm{Cov}(v'(1+\xi(\sigma,r)),R_m)/\sigma_m,$$
$$TRR = \rho(\sigma,r) = -\frac{U_\sigma(\sigma,r)}{U'_r(\sigma,r)} = -\frac{\mathrm{Cov}(v'(1+\xi(\sigma,r)),R_m)}{\sigma_m Ev'(1+\xi(\sigma,r))}.$$

Since $v''(x)<0$ and $\xi(\sigma,r)$ is positively related to R_m, $v'(1+\xi(\sigma,r))$ is negatively related to R_m. From the Pratt Theorem (1964, see [14]), we know that the stronger the risk aversion tendency, the more concave the utility function $v(x)$. Obviously, the more concave the $v(x)$ is, the faster the marginal utility $v'(x)$ diminishes. The faster the $v'(x)$ diminishes, the stronger the negative correlation between $v'(1+\xi(\sigma,r))$ and R_m, the greater the $-\mathrm{Cov}(v'(1+\xi(\sigma,r)),R_m)$, the higher the $\rho(\sigma,r)$. Therefore, the stronger the risk aversion tendency, the higher the $\rho(\sigma,r)$; that is, $\rho(\sigma,r)$ and the risk aversion tendency change uniformly or in the same direction. This shows that $\rho(\sigma,r)$ or TRR reflects the strength of the risk aversion tendency. Hence the TRR becomes a new tool to measure the risk aversion tendency of investors.

This profound aspect of TRR implies that the stronger the investor's risk aversion tendency, the higher the investor's requirements for risk return rate. If you do not meet the requirements, you will not take risks, but rather consider risk-free options.

Now we can find the reason that those risk averters with TRR$\geq$ARR at point $(0,r_f)$ have no investment in R_m. At the initial point $(0,r_f)$ of the budget line, $\rho(0,r_f)\geq\pi$ implies that the utility at any other point of budget line is less than the utility at $(0,r_f)$, so R_f is the optimal choice. In other words, the investor has a strong risk aversion tendency from the beginning so as to never invest in risky assets.

We can also find that when $\rho(\sigma_m,r_m)\leq\pi$, the investor never puts money into R_f, instead, they invest all of their money in R_m. Thus $\rho(\sigma_m,r_m)\leq\pi$ expresses the fact that the investor has a weak risk aversion tendency at last, so as to eventually invest all of their money in R_m. This kind of behavior makes the risk averter look like a risk lover, giving others the illusion that they loves risks.

In a word, for a risk averter, $\rho(0,r_f)\geq\pi$ means a strong risk aversion tendency for them to put all of their money in safe item R_f, and $\rho(\sigma_m,r_m)\leq\pi$ means a weak risk aversion tendency for them to put all of their money in risky item R_m. They will put their money into both R_f and R_m if and only if their risk aversion tendency is neither too strong ($\rho(0,r_f)<\pi$) nor too weak ($\rho(\sigma_m,r_m)>\pi$).

The conclusions drawn from the above analyses can be summarized into the following theorem:

Theorem 5.1 Investment decisions depend on the comparison between target rate of return (TRR) and actual rate of return (ARR). The TRR not only reflects the investor's required rate of return for taking risks, but also reflects the risk aversion tendency. In the case that the influence of systematic risk is light (ARR>0), all risk lovers, risk neutrals and those risk averters whose risk aversion tendency is not too strong ($\rho(0,r_f)<\pi$) will have investments in risky items; thus the situation for social investment activities is not so bad. Only in the cases where the influence of systematic risks is heavy or serious (ARR$\leq$0) will social investment activities be depressed to a great extent, and most people will not make investments.

This theorem has important policy implications. It tells us that governments should focus on those systematic risks that are expected to have serious or heavy influences. Priority should be placed upon maintaining stability in politics, the economy, the environment and in people's lives, in order to prevent macro-systemic risks from occurring. Maintaining currency stability to prevent currency itself from becoming a systematic risk factor is also important, as is maintaining the stability of the foreign exchange markets to prevent large fluctuations in exchange rates. Governments must maintain the stability of financial systems and markets to prevent the capital chain from breaking. Maintaining the continuity of policies to prevent long supply chains from breaking is crucial too. In summary, society should always pay attention to guard against the occurrence of those factors that induce heavy or serious systematic risks.

References

[1] Andrew P, Matthew W. An empirical evaluation of sensitivity bounds for mean-variance Portfolio optimization. Finance Research Letters, 2022, **44**: 102065

[2] Berk C, Tutari B. The effect of systematic risk and previous period returns on portfolio selection. Applied Finance and Accounting, 2020, **7**(1): 1–9

[3] Bianconia M, Hua X X, Tan C M. Determinants of systemic risk and information dissemination. International Review of Economics & Finance, 2015, **38**: 352–368

[4] Brown M J. A mean-variance serial replacement decision model: the correlated case. The Engineering Economist, 2007, **38**(3): 237–247

[5] Coyle B T. Risk aversion and yield uncertainty in duality models of production: a mean-variance approach. American Journal of Agricultural Economics, 1999, **81**(3): 553–567

[6] Dai M, Jin H Q, Kou S, Xu Y H. A dynamic mean-variance analysis for log returns. Management Science Volume, 2020, **67**(2): 1093–1108

[7] Kassimatis K. Mean-variance versus utility maximization revisited: The case of constant relative risk aversion. International Review of Financial Analysis, 2021, **78**: Art 101932

[8] Khashanah K, Simaan M, Simaan Y. Do we need higher-order comoments to enhance mean-variance portfolios? Evidence from a simplified jump process. International Review of Financial Analysis, 2022, **81**: 102068

[9] Lai T, Xing H, Chen Z. Mean-variance portfolio optimization when means and covariances are unknown. The Annals of Applied Statistics, 2011, **5**(2): 798–823

[10] Li, X, Sinem U A, John M M. Multi-period portfolio optimization using model predictive control with mean-variance and risk parity frameworks. European Journal of Operational Research, 2022, **299**(3): 1158–1176

[11] Liu Y. The validity of mean-variance model in A-share market. World Scientific Research Journal, 2022, **8**(3): 309–315

[12] Marianil F, Polinesil G, Recchioni M C. A tail-revisited Markowitz mean-variance approach and a portfolio network centrality. Computational Management Science, 2022, **19**: 425–455

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

[14] Pratt J. Risk aversion in the small and in the large. Econometrica, 1964, **32**(1): 22–36

[15] Rainer S. Heuristic mean-variance optimization in Markov decision processes using state-dependent risk aversion. IMA Journal of Management Mathematics, 2022, **33**(2): 181–199

[16] Van Staden P M, Dang D M, Forsyth P A. The surprising robustness of dynamic mean-variance portfolio optimization to model misspecification errors. European Journal of Operational Research, 2021, **289**(2): 774–792