Portfolio choice*

Jonas Nygaard Eriksen**

Asset Pricing, Fall 2021

*This note provides a review of static portfolio choice for risk averse individuals with preferences that can be described constistently with the expected utility model of choice under uncertainty. The note outlines and discusses the mean-variance model and details its implications for portfolio choice. The note draws strong inspiration from the excellent textbook treatments provided in Huang and Litzenberger (1988), Pennacchi (2008), Gollier (2001), and Campbell (2017). The note is prepared for use only in the Master's course "Asset Pricing". Please do not cite, circulate, or use for purposes other than this course.

**CREATES, Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. E-mail: jeriksen@econ.au.dk.

Contents

1. Introduction

What determines investors' demand for risky assets? That is, how do investors decide on the composition of their portfolios? This note applies the theory of rational decision making under uncertainty, embodied in the expected utility theory, to the financial problem of choosing a portfolio in a one-period setting. We begin with a study of a classic problem in finance: how to allocate wealth between a risk-free and a risky asset. The analysis makes use of the expected utility framework and the Pratt (1964) and Arrow (1971) coefficients of risk aversion to demonstrate that the share of wealth allocated to the risky asset depends on the individual's level of risk aversion, and that even risk averse agents will always allocate some positive investment to the risky asset if its risk premium is positive. This important, and somewhat surprising, result is known as the *participation principle* because it implies that risk aversion alone is insufficient to explain non-participation in risky asset markets with positive risk premia. One could potentially object to this implication and instead conjecture that high uncertainty around the expected excess return could delay investment in the risky asset, but the model is clear in its prediction that even a small positive premium should induce investment in the risky asset. We continue by deriving the intuitive result that more risk averse individuals will allocate a smaller share of their initial wealth to the risky asset compared to their less risk averse peers.

We then turn our attention to the financial problem of combining multiply assets in a portfolio, where we first introduce the intuition using a collection of examples involving two assets before continuing with the more general cases of (i) combining N risky assets and (ii) N risky assets and a risk-free asset — which we will refer to colloquially as the $N + 1$ asset case. The problem was famously studied by Harry Markowitz during his graduate studies at the University of Chicago. His main insight – which earned him the Nobel Memorial Price in Economic Sciences in 1990 – was to recognize that a risk averse investor should focus on balancing the expected return and variance (a proxy for risk) of her combined portfolio return rather than focus on individual assets due to the benefits of diversification (Markowitz, 1952). In a nutshell, diversification is the observation that the attractiveness of a given asset can differ when evaluated on its own relative to when held in a portfolio. One should therefore not evaluate assets in solitude, but rather consider how they co-move (measured by their covariances or correlations) with the other assets in the portfolio, and how that affect diversification possibilities. The main implication of the model is that a rational investor wishes to form a portfolio that efficiently trades off higher expected return for lower variance. Interestingly, not all feasible portfolios are efficient, and the model provides a way to distinguish efficient from inefficient portfolios.

Intuitively, it makes sense that investors want to construct a portfolio with the lowest variance for a given target return. Real world investors, however, may care about things other than mean and variance, but the analysis of Markowitz (1952) relies on these moments alone and does not take, say, higher-order moments into account. Accordingly, we need to ensure that investors' preferences are consistent with this notion. We show and discuss how mean-variance preferences can be motivated for arbitrary return distributions if investors have quadratic utility or for arbitrary preferences if returns are multivariate normally distributed. Although both assumptions are somewhat restrictive in their nature, the model is still able to deliver important and relevant insight into the portfolio choice problem that continues to be of use for academics, practitioners, and households alike today.

The remaining part of the note considers the general problem of allocating wealth between $N \geq 2$ risky assets, and potentially a risk-free one as well. We solve the quadratic optimization problem of determining the portfolio with the lowest possible variance for a given required expected return, and use it to identity the entire mean-variance portfolio frontier. We discuss the properties of the minimum variance portfolio of risky asset and introduce the portfolio separation theorem of Tobin (1958). The latter demonstrates that any two distinct frontier portfolios can be combined to create a third frontier portfolio. That is, investors can trade in as little as two frontier portfolios to create a frontier portfolio that suits their risk preferences. Together, this delivers a theory of optimal portfolio choice in the mean-variance sense.

We then introduce a risk-free asset to the investment universe, and show that the portfolio choice problem simplifies significantly. With a risk-free rate, the efficient frontier becomes a straight line from the risk-free asset through the tangency portfolio — the only efficient portfolio consisting purely of risky assets. In this case, portfolio separation implies that *all* investors hold the tangency portfolio in combination with the risk-free assets, and that *all* investors hold risky assets in the same relative proportion as identified by the weights of the tangency portfolio.

This note unfolds as follows. Section 2 discusses the relation between risk aversion and portfolio choice, and demonstrates the participation principle. Section 3 describes the mean-variance model, its assumptions, and the implications for portfolio selection. We solve the investor's optimization problem for risky assets without and with a risk-free asset.

2. Risk aversion and portfolio choice

This section considers the relation between risk aversion (as introduced in the previous topic) and the portfolio choice problem of an individual in a single period setting. This serves as an intuitive introduction to portfolio selection and the role played by the absolute and relative coefficients of risk aversion developed by Pratt (1964) and Arrow (1971). Specifically, the aim is to demonstrate how absolute and relative risk aversion affects the individual's portfolio allocations relative to initial wealth. Moreover, we show that a risk averse individual will always allocate some positive investment (however small that stake may be) to the risky asset if it pays a positive risk premium. This is referred to as the *principle of participation* and it is an important implication of the expected utility model for portfolio choice. It tells us that non-participation in risky asset markets with positive risk premia cannot be justified by risk aversion alone.

2.1. The participation principle

Consider the so-called canonical portfolio problem in which a risk averse individual has to divide her initial wealth between a risk-free and a risky asset. Assume that a risk-free asset exists that pays a safe rate of return of r_f. For simplicity, suppose that there is just one risky asset that pays an uncertain return equal to $\widetilde{r}$ (e.g., the market portfolio). The risk averse individual is endowed with initial wealth Y_0 and the objective is to determine the optimal money amount ϕ to allocate to the risky asset. The investor's uncertain end-of-period wealth $\widetilde{Y}_1$ is characterized by

$$\begin{aligned} \widetilde{Y}_1 &= (Y_0 - \phi)(1 + r_f) + \phi(1 + \widetilde{r}) && (1) \\ &= Y_0(1 + r_f) + \phi(\widetilde{r} - r_f), && (2) \end{aligned}$$

where $(Y_0 - \phi)$ denotes the investment in the risk-free asset. We can interpret (2) as follows. The first term represents the individual's return on wealth when the entire portfolio is invested in the risk-free asset and the second term represents the difference in return gained by investing the money amount ϕ in the risky asset.

We assume that the individual cares only about end-of-single-period consumption so that maximizing end-of-period consumption is equivalent to maximizing end-of-period wealth. If the individual is a von Neumann and Morgenstern (1944) expected utility maximizer, she will then choose her portfolio allocations by maximizing the expected utility of end-of-period wealth

$$\max_{\phi} \mathbb{E}\left[U\left(\widetilde{Y}_1\right)\right] = \mathbb{E}\left[U\left(Y_0(1 + r_f) + \phi(\widetilde{r} - r_f)\right)\right], \tag{3}$$

where the decision variable is the money amount ϕ to allocate to the risky asset. When the individual is risk averse and non-satiated in consumption ($U'(Y) > 0$ and $U''(Y) < 0$), the necessary and sufficient first-order condition (FOC) for the solution to the individual's problem in (3) is

$$\mathbb{E}\left[U'\left(\widetilde{Y}_1\right)(\widetilde{r} - r_f)\right] = 0, \tag{4}$$

where we can note that the second-order condition for a maximum

$$\mathbb{E}\left[U''\left(\widetilde{Y}_1\right)(\widetilde{r} - r_f)^2\right] < 0 \tag{5}$$

is automatically satisfied by the concavity of the utility function, i.e., $U''\left(\widetilde{Y}_1\right)$ is negative by assumption. It turns out that the FOC in (4) can provide a set of intuitive relations for portfolio allocations

$$\begin{aligned} \phi > 0 \quad &\Leftrightarrow \quad \mathbb{E}[\widetilde{r}] > r_f && (6) \\ \phi = 0 \quad &\Leftrightarrow \quad \mathbb{E}[\widetilde{r}] = r_f && (7) \\ \phi < 0 \quad &\Leftrightarrow \quad \mathbb{E}[\widetilde{r}] < r_f. && (8) \end{aligned}$$

Consider first the simple special case in which the expected return on the risky asset equals the risk-free rate, i.e., $\mathbb{E}[\widetilde{r}] = r_f$. In that case, $\phi = 0$ represents a solution to the FOC, and so the individual allocates her entire wealth to the risk-free asset. This becomes immediately clear when we note that $\phi = 0$ implies that end-of-period wealth becomes $\widetilde{Y}_1 = Y_0(1 + r_f)$ so that $U'(Y_1) = U'(Y_0(1 + r_f))$ becomes deterministic. As a result, the FOC becomes $\mathbb{E}\left[U'\left(\widetilde{Y}_1\right)(\widetilde{r} - r_f)\right] = U'(Y_0(1 + r_f))\,\mathbb{E}[\widetilde{r} - r_f] = 0$. In many ways, this intuitive result is reminiscent of the finding that a risk averse individual would reject a fair lottery. In this context, one can think of the risky asset as representing the fair lottery in that the risky asset has an expected return exactly equal to the risk-free rate.

We then turn to the case in which $\mathbb{E}[\widetilde{r}] > r_f$. We first demonstrate the straightforward result that $\phi = 0$ will not satisfy the FOC as $U'(Y_0(1 + r_f))\,\mathbb{E}[\widetilde{r} - r_f] > 0$ for $\phi = 0$ due to $\mathbb{E}[\widetilde{r}] > r_f$ and marginal utility always being positive by assumption (the individual always prefer more to less). Next, we instead argue that $\phi > 0$ is needed to satisfy the FOC when $\mathbb{E}[\widetilde{r}] > r_f$. To appreciate this, consider a simplified setting in which the risky asset pays one of two possible return realizations: a low return $r_L < r_f$ with probability π or a high return $r_H > r_f$ with probability $1 - \pi$. The corresponding end-of-period wealth levels are denoted Y_L and Y_H, respectively. It follows immediately that $U'(Y_L)(r_L - r_f) < 0$ and $U'(Y_H)(r_H - r_f) > 0$. The objective is then to determine ϕ such that $U'\left(\widetilde{Y}_1\right)(\widetilde{r} - r_f)$ averages to zero (in expectation) for all realizations of $\widetilde{r}$. For that to be true, it must be the case that $Y_H > Y_L$ so that $U'(Y_H) < U'(Y_L)$ due to the concavity of the utility function. To ensure that $\mathbb{E}\left[U'\left(\widetilde{Y}_1\right)(\widetilde{r} - r_f)\right] = 0$, the high return realization should receive a weight, $U'(Y_H)$, that is smaller than the weight, $U'(Y_L)$, applied to the low return realization. Importantly, this can only occur if $\phi > 0$ so that $Y_H > Y_L$. This result delivers the important and key implication that a risk averse individual will always hold at least some positive amount of the risky asset if its expected rate of return exceeds that of the risk-free rate, i.e., the participation principle. A reverse argument can be made for the case in which $\mathbb{E}[\widetilde{r}] < r_f$, where the individual will always short the risky asset.

2.2. Investment behavior over wealth and risk aversion

This section turns to an investigation of how the portfolio choice of an individual is affected by his degree of risk aversion and initial wealth. As a natural starting point, we compare the portfolio allocation of two individuals with differing levels of risk aversion, but the same level of wealth. That is, we ask the question: if Kenneth is more risk averse than John, will he then invest a smaller amount of his wealth in the risky asset? The answer is yes, as we would intuitively suspect. The result is attributable to Arrow (1971) and can be stated as follows. Consider two investors indexed by $i = 1{,}2$. Suppose that for all wealth levels $Y > 0$ we have that $R_A^1(Y) > R_A^2(Y)$ ($R_R^1(Y) > R_R^2(Y)$), where $R_A^i(Y)$ ($R_R^i(Y)$) denotes the coefficient of absolute (relative) risk aversion for investor $i = 1{,}2$, then $\phi_1 < \phi_2$. Consequently, the less risk averse the individual is, the larger is the money amount ϕ that the individual allocates to the risky asset.

The more interesting question to study is how the interplay between risk aversion *and* the level of the individual's initial wealth Y_0 influence portfolio allocations (ϕ). Consider a risk averse individual who prefers more to less that lives in an economy with one risk-free and one risky asset with a strictly positive expected excess return (risk premium). From the participation principle, we know that the individual will allocate part of her initial wealth to the risky asset. One may naturally wonder what the minimum risk premium is for the individual to invest all of her wealth in the risky asset. Let $\widetilde{r}$ and ϕ denote the risky asset return and the money amount invested in the risky asset, respectively. For an individual to invest all her wealth $\phi = Y_0$ in the risky asset it must hold that (using that (1) reduces to $\widetilde{Y}_1 = Y_0 (1 + \widetilde{r})$ for $\phi = Y_0$)

$$\mathbb{E}\left[U'\left(Y_0\left(1+\widetilde{r}\right)\right)\left(\widetilde{r}-r_f\right)\right] \geq 0. \tag{9}$$

Taking a first-order Taylor expansion of $U'\left(Y_0\left(1+\widetilde{r}\right)\right)$ around the point $U'\left(Y_0\left(1+r_f\right)\right)$, multiplying both sides with the risk premium $\widetilde{r} - r_f$, and taking expectations yields

$$\begin{aligned}\mathbb{E}\left[U'\left(Y_0\left(1+\widetilde{r}\right)\right)\left(\widetilde{r}-r_f\right)\right] \approx\ & U'\left(Y_0\left(1+r_f\right)\right)\mathbb{E}\left[\widetilde{r}-r_f\right] \\ & + U''\left(Y_0\left(1+r_f\right)\right)\mathbb{E}\left[\left(\widetilde{r}-r_f\right)^2\right]Y_0,\end{aligned} \tag{10}$$

where we say that risk is *small* if $\mathbb{E}\left[\left(\widetilde{r}-r_f\right)^2\right]$ is small so that we can safely ignore any higher order terms. Ignoring higher-order terms in the approximation, we can obtain an expression for the minimum risk premium required to induce full investment in the risky asset by setting (10) equal to zero and solving for the risk premium

$$\mathbb{E}\left[\widetilde{r}-r_f\right] \approx R_A\left(Y_0\left(1+r_f\right)\right)Y_0\mathbb{E}\left[\left(\widetilde{r}-r_f\right)^2\right], \tag{11}$$

where $R_A\left(Y_0\left(1+r_f\right)\right) = -\frac{U''\left(Y_0\left(1+r_f\right)\right)}{U'\left(Y_0\left(1+r_f\right)\right)}$ is the coefficient of absolute risk aversion of Pratt (1964) and Arrow (1971). From (11), we immediately observe that the higher an individual's absolute risk aversion, the higher is the minimum risk premium required to induce full investment in the risky asset. Intuitively, the curvature of an individual's utility function would be related to the minimum risk premium required to induce full investment in the risky asset because it indicates risk aversion, albeit that the second derivative $U''(\cdot)$ alone is insufficient to characterize the intensity of risk averse behavior — similar to its role in Pratt's approximation for the risk premium and certainty equivalent for a lottery asset in the expected utility model.

We now examine the relation between the amount invested in the risky asset ϕ, initial wealth Y_0, and the individual's Pratt-Arrow coefficient of absolute risk aversion $R_A(Y) = -\frac{U''(Y)}{U'(Y)}$. This coefficient can be increasing, decreasing, or constant as a function of initial wealth. If $R_A(Y)$ decreases with wealth, i.e., $\frac{dR_A(Y)}{dY_0} = R'_A(Y) < 0$, then the individual is said to display decreasing absolute risk aversion and will increase the money amount ϕ invested in the risky asset as initial wealth increases. If the individual displays increasing absolute risk aversion, then risk aversion increases with wealth and the money amount ϕ decreases accordingly with wealth. If the individual displays constant absolute risk aversion, then the money amount ϕ will remain unaffected by changes in wealth. These results, due to Arrow (1971), are summarized below.

Risk aversion	Coefficient	Investment behavior
Decreasing absolute	$R'_A(Y) < 0$	$\phi'(Y_0) > 0$
Constant absolute	$R'_A(Y) = 0$	$\phi'(Y_0) = 0$
Increasing absolute	$R'_A(Y) > 0$	$\phi'(Y_0) < 0$

Our aim is to prove that these claims are indeed true by focusing on the case of decreasing absolute risk aversion (DARA); the case most realistic when looking at the real-world behavior of individuals. Proofs for the remaining cases follow an analogous structure and can be proven by applying similar arguments. From the FOC in (4), one can determine the derivative of ϕ with respect to initial wealth Y_0 to be

$$\phi'(Y_0) = \frac{d\phi}{dY_0} = \frac{(1+r_f)\,\mathbb{E}\left[U''\left(\widetilde{Y}_1\right)(\widetilde{r}-r_f)\right]}{-\mathbb{E}\left[U''\left(\widetilde{Y}_1\right)(\widetilde{r}-r_f)^2\right]}, \tag{12}$$

where the denominator of (12) is positive because of the concavity of the utility function, which ensures that $U''(Y)$ is negative. The overall sign of the expression therefore depends solely on the numerator which is determined by the risk premium $(\widetilde{r}-r_f)$. Suppose that realized returns can take on two possible values: a high return $r_H > r_f$ or low return $r_L < r_f$, and that the individual displays decreasing absolute risk aversion. In the high return case, for $\phi \geq 0$, we have that the uncertain end-of-period wealth $Y_H \geq Y_0(1+r_f)$ so that

$$R_A(Y_H) \leq R_A(Y_0(1+r_f)). \tag{13}$$

Multiplying both sides of (13) by $-U'(Y_H)(r_H - r_f)$, which is a negative quantity due to the first derivative being positive and $r_H > r_f$, yields

$$U''(Y_H)(r_H - r_f) \geq -U'(Y_H)(r_H - r_f)\,R_A(Y_0(1+r_f)). \tag{14}$$

In the low return case, for $\phi \geq 0$, we have that the uncertain end-of-period wealth $Y_L \leq Y_0(1+r_f)$ so that

$$R_A(Y_L) \geq R_A(Y_0(1+r_f)). \tag{15}$$

Multiplying both sides of (15) by $-U'(Y_L)(r_L - r_f)$, which is a positive quantity because $r_L < r_f$, yields

$$U''(Y_L)(r_L - r_f) \geq -U'(Y_L)(r_L - r_f)\,R_A(Y_0(1+r_f)). \tag{16}$$

We note that the inequalities in (14) and (16) are of the same form. Taking expectations therefore yields the final result

$$\mathbb{E}\left[U''\left(\widetilde{Y}_1\right)(\widetilde{r}-r_f)\right] \geq -\mathbb{E}\left[U'\left(\widetilde{Y}_1\right)(\widetilde{r}-r_f)\right] R_A\left(Y_0\left(1+r_f\right)\right) \geq 0, \tag{17}$$

where the last inequality comes about from substituting in the FOC in (4). Thus, decreasing absolute risk aversion implies $\phi'(Y_0) > 0$ in (12) so that the individual invests an increasing amount of wealth in the risky asset for larger amounts of initial wealth. As discussed above, this implies that two individuals with the same utility function, but different levels of initial wealth, will invest different amounts in the risky asset. With decreasing absolute risk aversion, the wealthier individual will invest a larger money amount in the risky asset. The opposite is true for individuals with increasing risk aversion, which can be shown using similar arguments.

The above analysis deals with the absolute money amount ϕ invested in the risky assets. However, we may similarly be interested in the *proportion* ϕ/Y_0 of initial wealth invested in the risky asset. This can be studied using the concept of relative risk aversion. Define the elasticity measuring the proportional increase in the risky asset for an increase in initial wealth as

$$\eta = \frac{d\phi/\phi}{dY_0/Y_0} = \frac{d\phi}{dY_0}\frac{Y_0}{\phi}. \tag{18}$$

Let $R_R(Y) = -Y\frac{U''(Y)}{U'(Y)}$ be the Pratt-Arrow coefficient of relative risk aversion. This coefficient can be increasing, decreasing, or constant as a function of wealth. If $R_R(Y)$ decreases with wealth, i.e., $\frac{dR_R(Y)}{dY_0} = R_R'(Y) < 0$, then the individual is said to display decreasing relative risk aversion and will increase the fraction of wealth invested in the risky asset when wealth increases. If the individual display increasing relative risk aversion, then risk aversion increases with wealth so that the fraction of wealth decreases when wealth increases. If the individual displays constant relative risk aversion, then the fraction of wealth invested in the risky asset will remain unaffected by changes in wealth. These results, due to Arrow (1971), are summarized below.

Risk aversion	Coefficient	Investment behavior
Decreasing relative	$R_R'(Y) < 0$	$\eta > 1$
Constant relative	$R_R'(Y) = 0$	$\eta = 1$
Increasing relative	$R_R'(Y) > 0$	$\eta < 1$

We can prove these relations using arguments identical to those for the case of absolute risk aversion. To see this, add $1 - \frac{\phi}{\phi}$ to the right-hand side of (18) so that the equation becomes

$$\eta = 1 + \frac{(d\phi/dY_0)\,Y_0 - \phi}{\phi}. \tag{19}$$

Next, substitute in the expression for $d\phi/dY_0$ from (12) and re-arrange terms to obtain

$$\eta = 1 + \frac{Y_0\left(1+r_f\right)\mathbb{E}\left[U''\left(\widetilde{Y}_1\right)\left(\widetilde{r}-r_f\right)\right] + \phi\mathbb{E}\left[U''\left(\widetilde{Y}_1\right)\left(\widetilde{r}-r_f\right)^2\right]}{-\phi\mathbb{E}\left[U''\left(\widetilde{Y}_1\right)\left(\widetilde{r}-r_f\right)^2\right]}. \tag{20}$$

Collecting terms in the numerator then yields

$$\eta = 1 + \frac{\mathbb{E}\left[U''\left(\widetilde{Y}_1\right)\left(\widetilde{r}-r_f\right)\left\{Y_0\left(1+r_f\right)+\phi\left(\widetilde{r}-\right)\right\}\right]}{-\phi\mathbb{E}\left[U''\left(\widetilde{Y}_1\right)\left(\widetilde{r}-r_f\right)^2\right]} \tag{21}$$

$$= 1 + \frac{\mathbb{E}\left[U''\left(\widetilde{Y}_1\right)\left(\widetilde{r}-r_f\right)\widetilde{Y}_1\right]}{-\phi\mathbb{E}\left[U''\left(\widetilde{Y}_1\right)\left(\widetilde{r}-r_f\right)^2\right]}, \tag{22}$$

where the denominator is always positive because $U''(Y) < 0$. Again, the numerator determines the demand elasticity for the risky asset. Using a setup completely analogous to above, one can demonstrate that an individual with decreasing relative risk aversion has $\eta > 1$ and invests a higher fraction of her wealth in the risky asset as wealth increases. If an individual has a utility function displaying increasing relative risk aversion ($\eta < 1$), the fraction of his initial wealth that he invests in the risky asset would decline as his initial wealth increased. Last, an investor with constant relative risk aversion ($\eta = 1$) invests a constant fraction of his wealth.

3. The mean-variance model

This section provides an introduction to the portfolio selection model of Markowitz (1952, 1959). We start by introducing assumptions on preferences and asset returns. We then illustrate the intuition within the mean-variance model using the well known two asset examples before ultimately turning to the mathematics of the portfolio frontier for many assets.

3.1. Mean-variance preferences

The mean-variance model of Markowitz (1952) represents a monumental progress in our understanding of portfolio selection. The model assumes that investors have a preference for expected returns and an aversion to variance, which is implied by the concavity of the utility function of a risk averse individual. To see this, let $U\left(\widetilde{R}_p\right)$ denote an individual's utility function defined over end-of-period gross portfolio returns (or, equivalently, end-of-period wealth) and consider a Taylor series expansion around the expected gross portfolio return

$$\begin{aligned} U\left(\widetilde{R}_p\right) &= U\left(\mathbb{E}\left[\widetilde{R}_p\right]\right) + U'\left(\mathbb{E}\left[\widetilde{R}_p\right]\right)\left(\widetilde{R}_p - \mathbb{E}\left[\widetilde{R}_p\right]\right) \\ &\quad + \frac{1}{2}U''\left(\mathbb{E}\left[\widetilde{R}_p\right]\right)\left(\widetilde{R}_p - \mathbb{E}\left[\widetilde{R}_p\right]\right)^2 + H_3, \end{aligned} \tag{23}$$

where

$$H_3 = \sum_{i=3}^{\infty} \frac{1}{i!} U'^i \left(\mathbb{E}\left[\widetilde{R}_p\right] \right) \left(\widetilde{R}_p - \mathbb{E}\left[\widetilde{R}_p\right] \right)^i \tag{24}$$

is a remainder term and U'^i denotes the ith derivative of the utility function. Taking expectations to (23) yields and expression for expected utility (in the von Neumann and Morgenstern (1944) sense)

$$\mathbb{E}\left[U\left(\widetilde{R}_p\right)\right] = U\left(\mathbb{E}\left[\widetilde{R}_p\right]\right) + \frac{1}{2} U''\left(\mathbb{E}\left[\widetilde{R}_p\right]\right) \sigma^2_{\widetilde{R}_p} + \mathbb{E}\left[H_3\right], \tag{25}$$

which indicates a preference for expected returns and an aversion to variance ($U''(\cdot) < 0$) for an individual with a strictly concave utility function. Importantly, however, the expression in (25) also illustrates that expected utility cannot be defined exclusively over mean and variance for arbitrary return distributions and preferences.[1] Albeit higher-order moments may additionally matter to investors in the real-world, we can motivate the mean-variance model (i) by assuming that the remainder term is small, i.e., $\mathbb{E}[H_3] \approx 0$, (ii) for arbitrary return distributions if investors have quadratic utility, or (iii) for arbitrary preferences if returns are multivariate normally distributed since the normal distribution is fully characterized by its mean and variance so that $\mathbb{E}[H_3]$ becomes a function of mean and variance only.[2] Equipping individuals with quadratic utility is appealing in that it requires no assumptions on return distributions, but unrealistic as quadratic utility implies that individuals experience satiation at a certain point after which utility decreases and that individuals display increasing absolute risk aversion. While assuming normally distributed returns is appealing as the rate of return on a portfolio made up of assets having returns that are multivariate normally distributed is also normally distributed, the normal distribution is unbounded from below, which is inconsistent with limited liability and with economic theory that attributes little meaning to negative consumption.

The mean-variance model is therefore not a general model of asset choice, but its contribution to the development of finance theory and its simple, yet remarkable, insights on diversification still stand as strongly today. The central role of the mean-variance model in standard finance theory can be attributed to its analytical tractability and its clear and rich empirical predictions. Markowitz's brilliant insight was to recognize that, in allocating wealth among various risky assets, a risk averse investor should focus on the expectation and the risk of her combined

[1] Expected utility is, of course, more general and can depend on things like skewness (the third moment of returns). For example, the observation that individuals purchase lottery tickets with negative expected returns, but a large upside, suggests a role for skewness in utility. Kraus and Litzenberger (1976) develop a single-period model that augments the mean-variance model to account for investors with a preference for skewness. Harvey and Siddique (2002) examines the asset pricing effects of skewness empirically.

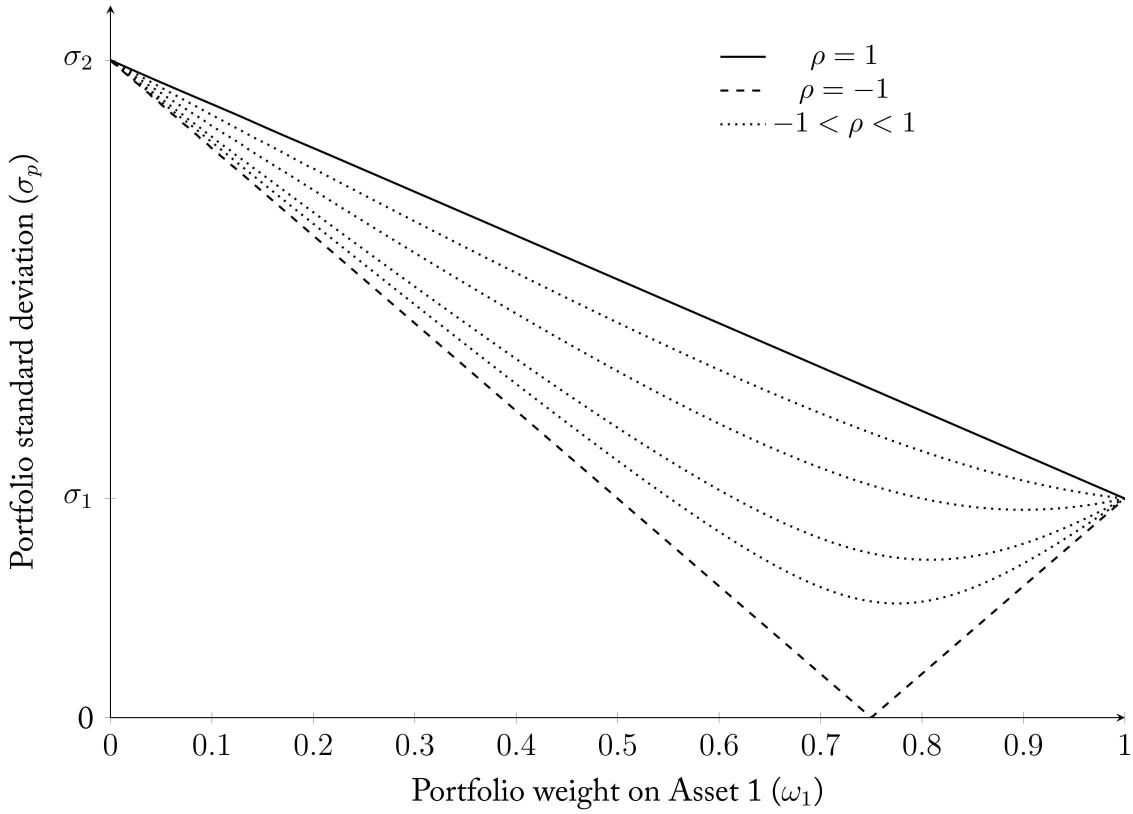
[2] A general misconception is that Markowitz assumed that returns are normally distributed and/or that the investor's utility function is quadratic. Rather, he simply noted that both provides sufficient approximations over a surprisingly large range of return. See Markowitz (2010) for an excellent discussion of his own views on portfolio theory today.

portfolio's return — a return that is affected by the individual assets' diversification possibilities. Due to diversification benefit, the attractiveness of a particular asset when held in a portfolio can differ significantly from its appeal when viewed in isolation. The mean-variance model applies the theory of rational decision making under uncertainty described by expected utility theory (von Neumann and Morgenstern, 1944) to the problem of choosing a portfolio. A rational investor would want to choose a portfolio of assets that efficiently trades off higher expected return for lower variance of return.

3.2. Building intuition with two assets

We begin the analysis by considering the problem of optimally combining two assets using the classic mean-variance analysis of Markowitz (1952) that judges assets and portfolios by their first two moments of returns. Although higher-order moments may also matter to investors in the real world, mean-variance analysis delivers a rich and useful intuition about portfolio formation and the importance of diversification.

Figure 1: Diversification for different levels of return correlation



3.2.1. Two risky assets

To illustrate the influence of diversification on portfolio selection, we consider a simple, but highly illustrative, example. Suppose that there are two risky assets, asset 1 and asset 2, traded in a frictionless economy with expected returns μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, and return correlation $-1 \leq \rho_{1,2} \leq 1$. Throughout, we will assume for simplicity – and without loss of generality – that $\mu_1 < \mu_2$ and that $\sigma_1^2 < \sigma_2^2$. Suppose that we form a portfolio with a fraction ω_1 invested in asset 1 and a fraction $\omega_2 = (1 - \omega_1)$ invested in asset 2. The expected portfolio return on this two asset portfolio is

$$\mu_p = \omega_1 \mu_1 + (1 - \omega_1) \mu_2, \tag{26}$$

since portfolio weights must sum to one. We note that realized returns are linear in portfolio weights as well, i.e.,

$$r_p = \omega_1 r_1 + (1 - \omega_1) r_2. \tag{27}$$

Importantly, the relation in (26) tells us that the expected portfolio return is a weighted average of the expected returns of the individual assets in the portfolio. The portfolio variance is

$$\begin{aligned}
\sigma_p^2 &= \text{Var}\left[\omega_1 r_1 + (1 - \omega_1) r_2\right] && (28)\\
&= \omega_1^2 \text{Var}\left[r_1\right] + (1 - \omega_1)^2 \text{Var}\left[r_2\right] - 2\omega_1 (1 - \omega_1) \text{Cov}\left[r_1, r_2\right] && (29)\\
&= \omega_1^2 \sigma_1^2 + (1 - \omega_1)^2 \sigma_2^2 - 2\omega_1 (1 - \omega_1) \sigma_{1,2}, && (30)
\end{aligned}$$

where the quadratic relation between ω_1 and σ_p^2 implies that a plot of means against variances is a parabola, and that a plot of means against standard deviations is a hyperbola. The portfolio standard deviation is

$$\sigma_p = \left[\omega_1^2 \sigma_1^2 + (1 - \omega_1)^2 \sigma_2^2 + 2\omega_1 (1 - \omega_1) \sigma_1 \sigma_2 \rho_{1,2}\right]^{\frac{1}{2}}, \tag{31}$$

where we quickly notice that the value of $\rho_{1,2}$ controls the benefits from diversification. For example, we have that $\sigma_{1,2} = \sigma_1 \sigma_2 \rho_{1,2} \leq \sigma_1 \sigma_2$, which implies that

$$\sigma_p \leq \omega_1 \sigma_1 + (1 - \omega_1) \sigma_2. \tag{32}$$

The inequality in (32) tells us that for $\rho_{1,2} < 1$, we have diversification benefits and the portfolio standard deviation is less than a weighted average of the individual assets' standard deviations. This is the benefit of diversification and its ability to reduce portfolio risk when correlation is not perfectly positive. In the situation where the two risky assets are perfectly positively correlated ($\rho_{1,2} = 1$), we have no diversification and the portfolio variance is simply a weighted average of the variances of the individual assets.

Figure 1 provides a simple illustration of this idea by plotting the portfolio variance against ω_1 for a range of possible values of $\rho_{1,2}$, where the main take-away is that portfolio risk reduces as $\rho_{1,2}$ goes towards negative one and that a risk-free portfolio can be made from two risky assets when there is perfect negative correlation ($\rho_{1,2} = -1$).

Figure 2: Efficient frontier with two risky assets

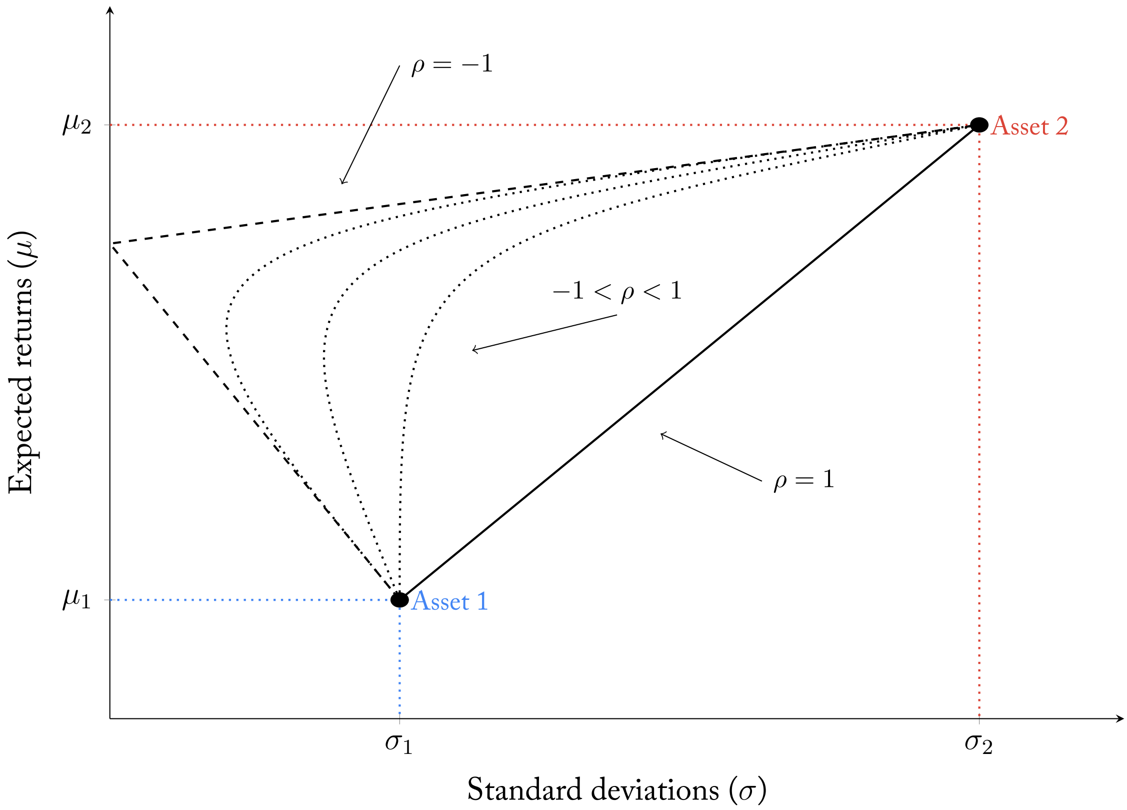


Figure 2 summarizes the possible portfolio frontiers that can be generated from our two hypothetical risky assets in mean-standard deviation (μ–σ) space for a broad range of possible values of $\rho_{1,2}$. We will limit this discussion to long positions only to build intuition here, but we will be more general later in this note. It is customary to distinguish between three special cases: i) perfect positive correlation, ii) perfect negative correlation, and iii) imperfect correlation. We begin with the special case of perfect positive correlation in which $\rho_{1,2} = 1$. Since the two assets always move in lockstep, there is no diversification gain. To see this, note that the standard deviation in (31) becomes a perfect square such that

$$\begin{aligned}\sigma_p &= \left[\omega_1^2\sigma_1^2 + (1-\omega_1)^2\sigma_2^2 + 2\omega_1(1-\omega_1)\sigma_1\sigma_2\right]^{\frac{1}{2}} \\ &= \omega_1\sigma_1 + (1-\omega_1)\sigma_2,\end{aligned} \tag{33}$$

which is a simple weighted average of the individual assets' standard deviations. Solving (33) for ω_1 yields the solution $\omega_1 = \frac{\sigma_p - \sigma_2}{\sigma_1 - \sigma_2}$ for a given portfolio volatility and the weigh places on asset 2 is $(1-\omega_1) = \frac{\sigma_1 - \sigma_p}{\sigma_1 - \sigma_2}$. Substituting the weights into (26) yields a closed form expression for the efficient frontier

$$\mu_p = \mu_1 + \frac{\mu_2 - \mu_1}{\sigma_2 - \sigma_1}(\sigma_p - \sigma_1), \tag{34}$$

which is a straight line between the two assets in μ–σ space as illustrated in Figure 2 as the solid black line. Importantly, there are no diversification benefits in this case.

Next, the second special case occurs when $\rho_{1,2} = -1$ so that the two risky assets are perfectly negatively correlated. In this case, the portfolio standard deviations once again becomes a perfect square

$$\begin{aligned}\sigma_p &= \left[\omega_1^2\sigma_1^2 + (1-\omega_1)^2\sigma_2^2 - 2\omega_1(1-\omega_1)\sigma_1\sigma_2\right]^{\frac{1}{2}} \\ &= \omega_1\sigma_1 - (1-\omega_1)\sigma_2,\end{aligned} \tag{35}$$

and we can show that the minimum variance portfolio is risk-free and that the efficient frontier is linear. Solving (35) for ω_1 gives is the weight placed on asset 1, i.e., $\omega_1 = \frac{\sigma_p+\sigma_2}{\sigma_1+\sigma_2}$ and the expression for the portfolio frontier becomes

$$\mu_p = \frac{\sigma_1\mu_2 + \sigma_2\mu_1}{\sigma_1+\sigma_2} \pm \frac{\mu_1-\mu_2}{\sigma_1+\sigma_2}\sigma_p. \tag{36}$$

The expected return on the risk-free minimum variance portfolio can be found by setting σ_p equal to zero in (36), i.e., the intercept represents the expected return on the minimum variance portfolio. The weight placed on asset 1 in this portfolio is equal to $\omega_1 = \frac{\sigma_2}{\sigma_1+\sigma_2}$, which implies that the risk-free portfolio can be formed with positive investments in the two risky assets. Figure 2 illustrates the portfolio frontier for this case in the dashed black lines, where only the upward sloping part is mean-variance efficient. That is, all portfolios on the downward sloping line are mean-variance dominated by the portfolios on the upward sloping part. No rational investor would place herself on the inefficient part of the frontier. The exact placement on the efficient frontier is determined by the investor's risk aversion. x The final special case occurs when the assets are imperfectly correlated ($-1 < \rho_{1,2} < 1$). In this case, the relationship between portfolio expected return and standard deviation is no longer linear, but a hyperbolic (bullet shaped) as illustrated in Figure 2 by the dotted black lines. It is no longer possible to create a risk-free portfolio, so the minimum variance portfolio will have $\sigma_p > 0$. However, as $\rho_{1,2}$ approaches -1, the less the variance of the minimum variance portfolio becomes and the portfolio frontier moves further to the left.

3.2.2. One risky and one risk-free asset

Suppose that one of the assets is risk-free, e.g., $\sigma_1 = 0$ and $\mu_1 = r_f$, then the efficient frontier becomes a straight line that originates from the vertical axis at the risk-free return. The portfolio standard deviation is

$$\sigma_p = (1-\omega_1)\sigma_2 \tag{37}$$

and the expected portfolio return is

$$\mu_p = \mu_1 + \frac{\mu_2-\mu_1}{\sigma_2}\sigma_p = r_f + \frac{\mu_2 - r_f}{\sigma_2}\sigma_p, \tag{38}$$

which follows directly as a simplification of the above discussion for two risky assets. Specifically, we can show that the weight on the risk-free asset is $\omega_1 = \frac{\sigma_p}{\sigma_2}$. The efficient frontier in (38) is often referred to as the *Capital Allocation Line (CAL)* and its slope is the Sharpe ratio of the risky asset, i.e., $SR = \frac{\mu_2 - r_f}{\sigma_2}$. The Sharpe ratio measures the excess return earned per unit of risk, where risk is measured by the standard deviation. Any portfolio that combines a single risky asset with the risk-free asset has the same Sharpe ratio as the risky asset itself. Therefore, all portfolios on the CAL has the same Sharpe ratio.

3.3. The mathematics of the efficient frontier

This section details the analytical solution to the problem of determining the portfolio frontier for many assets (with and without a risk-free asset). We consider the problem of finding portfolios that have minimum variance for a given required return. These are called minimum-variance portfolios, and are said to lie on the minimum-variance portfolio frontier. The analytical solution is provided by Merton (1972) who asked the following question: given the expected returns and the covariance matrix for N individual assets, what is the set of portfolio weights that minimizes the variance of the portfolio for each feasible expected portfolio return. The following sections outlines his analytical solutions to the problem.

3.3.1. Notation

We start by introducing notation. Define the $N \times 1$ vector of expected asset returns ($\boldsymbol{\mu}$) for the N risky asset as

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{bmatrix} \tag{39}$$

and let the $N \times 1$ vector of portfolio weights ($\boldsymbol{\omega}$) be given as

$$\boldsymbol{\omega} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_N \end{bmatrix} \tag{40}$$

and define the $N \times N$ variance-covariance matrix of risky asset returns as

$$\mathbf{V} \equiv \mathrm{Cov}\left(\widetilde{\mathbf{r}}\right) = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1,N} \\ \vdots & \ddots & \vdots \\ \sigma_{N,1} & \cdots & \sigma_N^2 \end{bmatrix}, \tag{41}$$

which we assume to be of full rank, i.e., that there are no redundant assets among the N assets. Finally, let $\mathbf{1}$ denote a $N \times 1$ vector of ones, let $\mu_p = \boldsymbol{\omega}^\top \boldsymbol{\mu}$ be a scalar denoting the required rate of return on the portfolio, and let $\sigma_p^2 = \boldsymbol{\omega}^\top \mathbf{V} \boldsymbol{\omega}$ be a scalar denoting the variance of the portfolio return. Using this matrix notation, the following sections will illustrate and solve the portfolio choice problem of investors judging assets by their first two moments.

3.3.2. Portfolio selection with risky assets

Suppose that there are $N \geq 2$ risky asset traded in a frictionless economy where unlimited short selling is allowed, and where assets are linearly independent (i.e., no redundancies in the selection of assets). In the mean-variance model, we say that a portfolio is a *frontier portfolio* if it has the minimum variance among portfolios with the same expected return (or the maximum return for the same variance). A portfolio p is a frontier portfolio if and only if its $N \times 1$ vector of portfolio weights $\boldsymbol{\omega}_p$ satisfies the following quadratic optimization problem

$$\begin{aligned} \min_{\boldsymbol{\omega}} \ & \frac{1}{2} \boldsymbol{\omega}^\top \mathbf{V} \boldsymbol{\omega} \\ \text{s.t. } & \boldsymbol{\omega}^\top \boldsymbol{\mu} = \mu_p \\ & \boldsymbol{\omega}^\top \mathbf{1} = 1, \end{aligned} \tag{42}$$

where $\boldsymbol{\mu}$ denotes an $N \times 1$ vector of expected returns, $\mathbf{1}$ is a vector of ones, and μ_p is the required return on the portfolio p. The problem in (42) minimizes the portfolio variance subject to the constraint that the expected portfolio return equals μ_p and that the portfolio weights sum to one. We note that short sales (negative weights) are allowed in the problem. The coefficient of $\frac{1}{2}$ in the objective function appears merely for notational convenience and does not change the form of the solution. We can solve the constrained optimization problem by writing up the Lagrangian

$$\min_{\{\boldsymbol{\omega}, \lambda, \gamma\}} \mathcal{L} = \frac{1}{2} \boldsymbol{\omega}^\top \mathbf{V} \boldsymbol{\omega} + \lambda \left(\mu_p - \boldsymbol{\omega}^\top \boldsymbol{\mu} \right) + \gamma \left(1 - \boldsymbol{\omega}^\top \mathbf{1} \right), \tag{43}$$

where λ and γ are Lagrange multipliers (positive constants) on the constraints. The necessary and sufficient first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\omega}} = \mathbf{V} \boldsymbol{\omega} - \lambda \boldsymbol{\mu} - \gamma \mathbf{1} = \mathbf{0} \tag{44}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \mu_p - \boldsymbol{\omega}^\top \boldsymbol{\mu} = 0 \tag{45}$$

$$\frac{\partial \mathcal{L}}{\partial \gamma} = 1 - \boldsymbol{\omega}^\top \mathbf{1} = 0, \tag{46}$$

where $\mathbf{0}$ is an $N \times 1$ vector of zeros. Solving for $\boldsymbol{\omega}$ in (44) yields the solution

$$\boldsymbol{\omega}_p^* = \lambda \mathbf{V}^{-1} \boldsymbol{\mu} + \gamma \mathbf{V}^{-1} \mathbf{1} \tag{47}$$

for an optimal portfolio in the mean-variance sense. Notice that the portfolio weights are increasing in $\boldsymbol{\mu}$ and decreasing in $\mathbf{V}$. However, λ and γ are still unknown at this point. We can determine the expressions for them as follows. First, pre-multiply both sides of (47) with $\boldsymbol{\mu}^\top$ and use the constraint in (45) to obtain (since $\boldsymbol{\omega}_p^\top \boldsymbol{\mu} = \boldsymbol{\mu}^\top \boldsymbol{\omega}_p = \mu_p$)

$$\boldsymbol{\mu}^\top \boldsymbol{\omega}_p^* = \lambda \boldsymbol{\mu}^\top \mathbf{V}^{-1} \boldsymbol{\mu} + \gamma \boldsymbol{\mu}^\top \mathbf{V}^{-1} \mathbf{1} = \mu_p. \tag{48}$$

Next, pre-multiply both sides of (47) with $\mathbf{1}^\top$ and use the constraint in (46) to obtain

$$\mathbf{1}^\top \boldsymbol{\omega}_p^* = \lambda \mathbf{1}^\top \mathbf{V}^{-1} \boldsymbol{\mu} + \gamma \mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1} = 1. \tag{49}$$

For notional simplicity, let us define the following constants

$$A = \mathbf{1}^\top \mathbf{V}^{-1} \boldsymbol{\mu} = \boldsymbol{\mu}^\top \mathbf{V}^{-1} \mathbf{1} \tag{50}$$

$$B = \boldsymbol{\mu}^\top \mathbf{V}^{-1} \boldsymbol{\mu} > 0 \tag{51}$$

$$C = \mathbf{1}^\top \mathbf{V}^{-1} \mathbf{1} > 0 \tag{52}$$

$$D = BC - A^2 > 0 \tag{53}$$

so that we can write (48) and (49) more compactly as

$$\mu_p = \lambda B + \gamma A \tag{54}$$

$$1 = \lambda A + \gamma C. \tag{55}$$

We can then solve for λ and γ using (54) and (55), respectively, to obtain the expressions

$$\lambda = \frac{\mu_p - \gamma A}{B} \tag{56}$$

$$\gamma = \frac{1 - \lambda A}{C}. \tag{57}$$

Determining λ Insert (57) into (56) and re-arrange to obtain the expression

$$\lambda = \frac{\mu_p - \frac{1-\lambda A}{C} A}{B} \tag{58}$$

$$= \frac{\mu_p}{B} - \frac{A - \lambda A^2}{BC} \tag{59}$$

$$= \frac{C\mu_p}{BC} - \frac{A - \lambda A^2}{BC}, \tag{60}$$

and then solve for λ to arrive at the final expression for λ using the definition of D

$$\lambda = \frac{C\mu_p - A}{BC - A^2} = \frac{C\mu_p - A}{D}. \tag{61}$$

Determining γ Insert (56) into (57) and re-arrange to obtain

$$\gamma = \frac{1 - \frac{\mu_p - \gamma A}{B} A}{C} \tag{62}$$
$$= \frac{1}{C} - \frac{A\mu_p - \gamma A^2}{BC} \tag{63}$$
$$= \frac{B}{BC} - \frac{A\mu_p - \gamma A^2}{BC}, \tag{64}$$

and then finally solve for γ to find the final expression

$$\gamma = \frac{B - A\mu_p}{BC - A^2} = \frac{B - A\mu_p}{D}. \tag{65}$$

Solution Combining (47) with the expression for λ and γ reveals that the solution to the optimization problem in (42) is then given by

$$\boldsymbol{\omega}_p^* = \underbrace{\frac{C\mu_p - A}{D}}_{\lambda} \mathbf{V}^{-1}\boldsymbol{\mu} + \underbrace{\frac{B - A\mu_p}{D}}_{\gamma} \mathbf{V}^{-1}\mathbf{1}. \tag{66}$$

We can equivalently write (66) as an affine function as follows

$$\boldsymbol{\omega}_p^* = \mathbf{g} + \mathbf{h}\mu_p \tag{67}$$

where

$$\mathbf{g} = \frac{1}{D}\left[B\mathbf{V}^{-1}\mathbf{1} - A\mathbf{V}^{-1}\boldsymbol{\mu}\right] \tag{68}$$
$$\mathbf{h} = \frac{1}{D}\left[C\mathbf{V}^{-1}\boldsymbol{\mu} - A\mathbf{V}^{-1}\mathbf{1}\right]. \tag{69}$$

We note that any frontier portfolio can be generated by (67) and that any portfolio that can be represented by (67) is a frontier portfolio. The collection of all frontier portfolios is called the portfolio frontier.

Portfolio variance We can show that the minimized portfolio variance σ_p^2 is a function of the Lagrange multipliers. In particular, write the portfolio variance as

$$\sigma_p^2 = \boldsymbol{\omega}_p^\top \mathbf{V}\boldsymbol{\omega}_p \tag{70}$$
$$= \boldsymbol{\omega}_p^\top \mathbf{V}\left(\lambda\mathbf{V}^{-1}\boldsymbol{\mu} + \gamma\mathbf{V}^{-1}\mathbf{1}\right) \tag{71}$$
$$= \lambda\boldsymbol{\omega}_p^\top\boldsymbol{\mu} + \gamma\boldsymbol{\omega}_p^\top\mathbf{1} \tag{72}$$
$$= \lambda\mu_p + \gamma, \tag{73}$$

where the second equality uses the definition of optimal portfolio weights and the last equality uses the constraints in (45) and (46). Next, insert the expressions for λ and γ to obtain

$$\sigma_p^2 = \frac{C\mu_p - A}{D}\mu_p + \frac{B - A\mu_p}{D} \tag{74}$$
$$= \frac{C\mu_p^2 - A\mu_p + B - A\mu_p}{D} \tag{75}$$
$$= \frac{1}{D}\left[C\mu_p^2 - 2A\mu_p + B\right], \tag{76}$$

which is a parabola in mean-variance space with vertex $\left(\frac{A}{C}, \frac{1}{C}\right)$. We can further manipulate the expression to arrive at a simple and informative expression. Factoring out the constant C and using that $B = \left(D + A^2\right)/C$ we get

$$\sigma_p^2 = \frac{C}{D}\left[\mu_p^2 - 2\frac{A}{C}\mu_p + \frac{B}{C}\right] \tag{77}$$
$$= \frac{C}{D}\left[\mu_p^2 - 2\frac{A}{C}\mu_p + \frac{D}{C^2} + \frac{A^2}{C^2}\right] \tag{78}$$
$$= \frac{C}{D}\left[\left(\mu_p - \frac{A}{C}\right)^2 + \frac{D}{C^2}\right] \tag{79}$$
$$= \frac{C}{D}\left(\mu_p - \frac{A}{C}\right)^2 + \frac{1}{C}. \tag{80}$$

We know that $C, D > 0$, and due to the quadratic nature of the expression, we immediately have that the minium variance portfolio has variance equal to $\frac{1}{C}$ and expected return equal to $\frac{A}{C}$. Substituting $\mu_p = \frac{A}{C}$ into (66) reveals that the weights on the minimum variance portfolio are given by the expression $\omega_{\text{MVP}} = \frac{1}{C}\mathbf{V}^{-1}\mathbf{1}$ (I encourage you to verify for yourself).

3.3.3. Portfolio separation: a two fund theorem

Tobin (1958) provides an important result within the mean-variance model of Markowitz (1952): the two-fund separation theorem. The result states that: "Every portfolio on the minimum-variance portfolio frontier can be replicated by a combination of any two frontier portfolios; and an individual will be indifferent between choosing among the N financial assets, or choosing a combination of just two frontier portfolios". This is a powerful and remarkable finding that has immediate and practical implications for portfolio selection. Specifically, if investors have homogenous expectations regarding the distribution of asset returns, then the portfolio frontier is characterized by (67) and any investor can form their preferred frontier portfolio by trading in as little as two frontier portfolios.

We are now going to prove this result using the affine function in (67). The first step is to understand the two vectors $\mathbf{g}$ and $\mathbf{g} + \mathbf{h}$. As we shall see, these vectors can be interpreted as portfolio weights corresponding to two easily recognizable frontier portfolios – something that is illustrated in Figure 3 in which two possible locations of $\mathbf{g}$ and $\mathbf{g}+\mathbf{h}$ are indicated together with the minimum variance portfolio. First, we claim that $\mathbf{g}$ is a vector of portfolio weights corresponding to a frontier portfolio with a zero expected returns. To see this, set $\mu_p = 0$ in (67) so that

$$\boldsymbol{\omega}_p^* = \mathbf{g} + \mathbf{h} \cdot 0 = \mathbf{g}. \tag{81}$$

Next, we claim that that $\mathbf{g}+\mathbf{h}$ is the vector of portfolio weights of a frontier portfolio with an expected return of 1, i.e.,

$$\boldsymbol{\omega}_p^* = \mathbf{g} + \mathbf{h} \cdot 1 = \mathbf{g} + \mathbf{h}. \tag{82}$$

Having verified that $\mathbf{g}$ and $\mathbf{g}+\mathbf{h}$ are two frontier portfolios (because they satisfy the solution), we are now going to claim that the entire set of frontier portfolios can be generated by $\mathbf{g}$ and $\mathbf{g}+\mathbf{h}$. Let q denote a frontier portfolio with expected return μ_q, and whose portfolio weights are given by

$$\boldsymbol{\omega}_q^* = \mathbf{g} + \mathbf{h}\mu_q, \tag{83}$$

which follows directly from (67). Since $\mathbf{g}$ and $\mathbf{g}+\mathbf{h}$ are frontier portfolios, we can invest a fraction of our wealth in each of them. Suppose that we consider the following portfolio weights on $\mathbf{g}$ and $\mathbf{g}+\mathbf{h}$: $\{(1-\mu_q), \mu_q\}$, then we can show that

$$(1-\mu_q)\,\mathbf{g} + \mu_q\,(\mathbf{g}+\mathbf{h}) = \mathbf{g} + \mathbf{h}\mu_q = \boldsymbol{\omega}_q^*, \tag{84}$$

so that the above chosen portfolio indeed generates the frontier portfolio q. Since the portfolio q is arbitrarily chosen, we have shown that the entire portfolio frontier can be generated by the two frontier portfolios $\mathbf{g}$ and $\mathbf{g}+\mathbf{h}$.

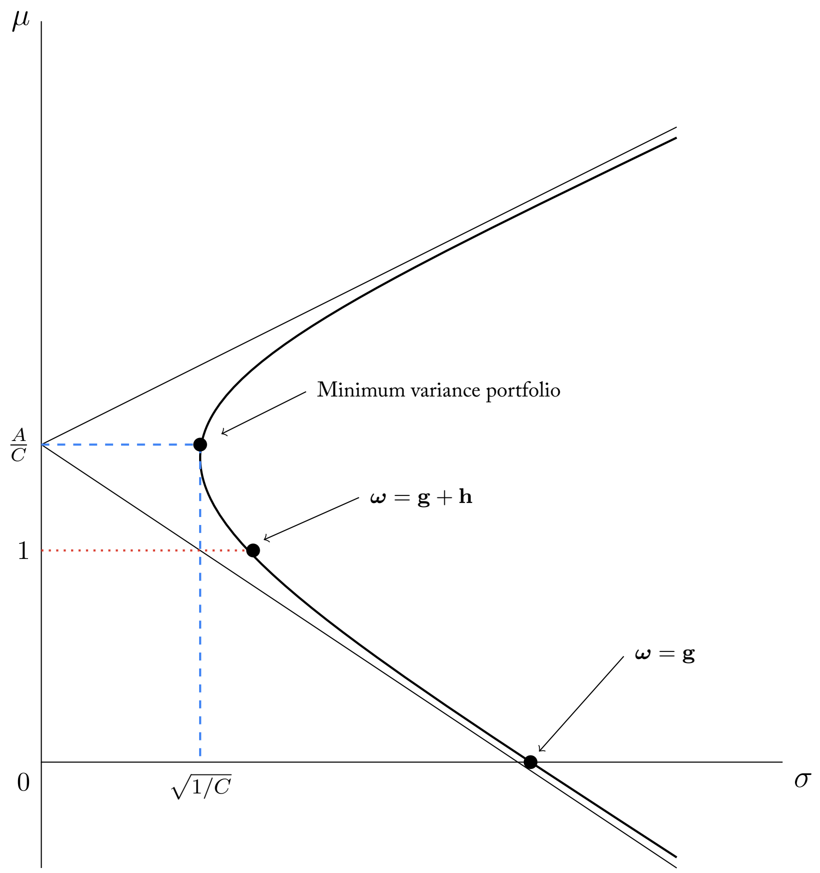
We can note that the above result relies only on the simple fact that $\mathbf{g}$ and $\mathbf{g}+\mathbf{h}$ have different expected returns. In turn, this implies a much stronger statement: the portfolio frontier can be generated by *any* two distinct frontier portfolios, not just the frontier portfolios $\mathbf{g}$ and $\mathbf{g}+\mathbf{h}$. To appreciate this result, let p_1 and p_2 be two distinct frontier portfolios $(\mu_1 \neq \mu_2)$ and let q be *any* frontier portfolio, then there exists a unique real number α such that

$$\mu_q = \alpha\mu_1 + (1-\alpha)\,\mu_2. \tag{85}$$

Now, consider the portfolio weights on p_1 and p_2: $\{\alpha, (1-\alpha)\}$, where we make use of the result in (67) and the relation in (85)

$$\begin{aligned}
\alpha\boldsymbol{\omega}_{p_1}^* + (1-\alpha)\,\boldsymbol{\omega}_{p_2}^* &= \alpha\,(\mathbf{g}+\mathbf{h}\mu_1) + (1-\alpha)\,(\mathbf{g}+\mathbf{h}\mu_2) && (86)\\
&= \mathbf{g} + \mathbf{h}\,[\alpha\mu_1 + (1-\alpha)\,\mu_2] && (87)\\
&= \mathbf{g} + \mathbf{h}\mu_q && (88)\\
&= \boldsymbol{\omega}_q^*. && (89)
\end{aligned}$$

Figure 3: Frontier portfolios



This shows that the entire portfolio frontier can be generated by any two distinct frontier portfolios as stated in the portfolio separation result.

Efficient portfolios The frontier portfolios that have expected returns that are strictly higher than that of the minimum variance portfolio, $\frac{A}{C}$, are called *efficient portfolios*. Portfolios whose returns are lower than $\frac{A}{C}$ are referred to as inefficient portfolios. They are inefficient because we can form an efficient portfolio with a higher expected return for the same variance — they are mean-variance dominated. Using an approach similar to above, we can show more generally that any linear combination of frontier portfolios is on the frontier and that any combination of efficient portfolios will itself be an efficient portfolio.

Let $\boldsymbol{\omega}_i$, $i = 1,2,\ldots,m$ be m frontier portfolios and α_i, $i = 1,2,\ldots,m$ be real numbers so that $\sum_{i=1}^{m} \alpha_i = 1$, then

$$\sum_{i=1}^{m} \alpha_i \boldsymbol{\omega}_i = \sum_{i=1}^{m} \alpha_i \left(\mathbf{g} + \mathbf{h}\mu_i\right) = \mathbf{g} + \mathbf{h} \sum_{i=1}^{m} \alpha_i \mu_i, \tag{90}$$

meaning that any linear combination of frontier portfolios is also a frontier portfolio. If the portfolios $\boldsymbol{\omega}_i$, $i = 1,2,\ldots,m$ are efficient, and if α_i, $i = 1,2,\ldots,m$ are non-negative, then

$$\sum_{i=1}^{m} \alpha_i \mu_i \geq \sum_{i=1}^{m} \alpha_i \frac{A}{C} = \frac{A}{C}, \tag{91}$$

implying that any linear combination of efficient portfolios will also be an efficient portfolio, and that the set of efficient portfolios is a convex set.

3.3.4. Portfolio selection with a risk-free asset

The analysis of portfolio selection has insofar assumed that investors only invest in risky assets. Investors would hold efficient portfolios comprised of risky assets whose placement on the portfolio frontier depends on the investors' level of risk aversion. This section shows that the introduction of a risk-free asset can greatly simplify the investor's portfolio choice problem, and that we can obtain an even stronger portfolio separation result. This augmented portfolio selection problem was described and solved by Tobin (1958).

Consider an economy with $N \geq 2$ risky assets with expected return vector $\boldsymbol{\mu}$ and one risk-free asset offering a safe return of r_f. A portfolio p of all $N+1$ assets is a frontier portfolio if and only if the $N \times 1$ vector of risky asset weights $\boldsymbol{\omega}$ satisfies the optimization problem

$$\begin{aligned} &\min_{\boldsymbol{\omega}} \ \frac{1}{2}\boldsymbol{\omega}^\top \mathbf{V}\boldsymbol{\omega} \\ &\text{s.t. } \boldsymbol{\omega}^\top \boldsymbol{\mu} + \left(1 - \boldsymbol{\omega}^\top \mathbf{1}\right) r_f = \mu_p, \end{aligned} \tag{92}$$

where δ denotes the Lagrange multiplier. We note that the problem is still formulated in terms of choosing the optimal weights $\boldsymbol{\omega}_p$ in risky assets, but where the portfolio is now completed by borrowing or lending at the risk-free rate. We therefore no longer require that $\boldsymbol{\omega}_p^\top \mathbf{1} = 1$. The Lagrangian is

$$\min_{\{\boldsymbol{\omega},\delta\}} \mathcal{L} = \frac{1}{2}\boldsymbol{\omega}^\top \mathbf{V}\boldsymbol{\omega} + \delta\left(\mu_p - \boldsymbol{\omega}^\top \boldsymbol{\mu} - \left(1 - \boldsymbol{\omega}^\top \mathbf{1}\right) r_f\right) \tag{93}$$

with necessary and sufficient first-order conditions

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\omega}} = \mathbf{V}\boldsymbol{\omega} - \delta\left(\boldsymbol{\mu} - r_f \mathbf{1}\right) = 0 \tag{94}$$

$$\frac{\partial \mathcal{L}}{\partial \delta} = \mu_p - \boldsymbol{\omega}^\top \boldsymbol{\mu} - \left(1 - \boldsymbol{\omega}^\top \mathbf{1}\right) r_f = 0, \tag{95}$$

so that the solution for the optimal weights in the N risky assets becomes

$$\boldsymbol{\omega}_p^* = \delta \mathbf{V}^{-1}\left(\boldsymbol{\mu} - r_f \mathbf{1}\right). \tag{96}$$

Notice again that the portfolio weights are increased in expected returns $\boldsymbol{\mu}$, but declining in variance $\mathbf{V}$. The constraint in the portfolio choice problem implies that $\mu_p - r_f = \boldsymbol{\omega}^\top\left(\boldsymbol{\mu} - r_f \mathbf{1}\right) =$

$(\boldsymbol{\mu} - r_f \mathbf{1})^\top \boldsymbol{\omega}$, so pre-multiplying (96) with $(\boldsymbol{\mu} - r_f \mathbf{1})^\top$ yields

$$\mu_p - r_f = (\boldsymbol{\mu} - r_f \mathbf{1})^\top \boldsymbol{\omega} = \delta (\boldsymbol{\mu} - r_f \mathbf{1})^\top \mathbf{V}^{-1} (\boldsymbol{\mu} - r_f \mathbf{1}) = \delta H \tag{97}$$

where $H \equiv (\boldsymbol{\mu} - r_f \mathbf{1})^\top \mathbf{V}^{-1} (\boldsymbol{\mu} - r_f \mathbf{1})$ is another constant, which can be similarly defined as $H = B - 2Ar_f + Cr_f^2$ using the constants from earlier. H is often interpreted as the squared Sharpe ratio of all frontier portfolios, which will be illustrated clearly below in the definition of the portfolio variance. One can verify that $H > 0$.

Solution Solving for δ in (97) and inserting into (96) yields the final solution for optimal portfolios

$$\boldsymbol{\omega}_p^* = \frac{\mu_p - r_f}{H} \mathbf{V}^{-1} (\boldsymbol{\mu} - r_f \mathbf{1}). \tag{98}$$

As above, we note that any frontier portfolio can be generated by (98) and that any portfolio that can be characterized by (98) is a frontier portfolio.

Portfolio variance We can further show that the variance of a frontier portfolio depends on the Lagrange multiplier δ and the constant H

$$\sigma_p^2 = \boldsymbol{\omega}_p^\top \mathbf{V} \boldsymbol{\omega}_p = \delta^2 (\boldsymbol{\mu} - r_f \mathbf{1})^\top \mathbf{V}^{-1} \mathbf{V} \mathbf{V}^{-1} (\boldsymbol{\mu} - r_f \mathbf{1}) = \delta^2 H = \frac{(\mu_p - r_f)^2}{H}, \tag{99}$$

where the interpretation of H as the squared Sharpe ratio is immediately visible. In fact, rearranging reveals that $H = \left(\frac{\mu_p - r_f}{\sigma_p}\right)^2$. Last, we can show that the portfolio frontier is a straight line identified by the expression

$$\mu_p = r_f \pm \sigma_p H^{\frac{1}{2}}. \tag{100}$$

The portfolio frontier of all assets is therefore composed of two half-lines originating from the point $(0,r_f)$ in (σ,μ)–space with slope $\pm H^{\frac{1}{2}}$, where $H^{\frac{1}{2}}$ is interpreted as the Sharpe ratio of all portfolios located on the efficient frontier.

Tangency portfolio The tangency portfolio is the only frontier portfolio composed only of risky assets, i.e., $\mathbf{1}^\top \boldsymbol{\omega}_T^* = 1$, whose weights are determined as

$$\boldsymbol{\omega}_T^* = \frac{1}{A - Cr_f} \mathbf{V}^{-1} (\boldsymbol{\mu} - r_f \mathbf{1}). \tag{101}$$

To see this, note that $\mathbf{1}^\top \boldsymbol{\omega}_T^* = 1$ implies that

$$\mathbf{1}^\top \boldsymbol{\omega}_T^* = \frac{\mu_T - r_f}{H} \mathbf{1}^\top \mathbf{V}^{-1} (\boldsymbol{\mu} - r_f \mathbf{1}) = 1. \tag{102}$$

Solving for μ_T yields the following expression

$$\mu_T = r_f + \frac{H}{A - Cr_f} \tag{103}$$

which when substituted back into (98) yields the final solution.

3.3.5. Portfolio separation: three special cases

When the investor has access to a risk-free asset for portfolio formation, we can obtain a strong portfolio separation result with clear and practical implications for portfolio selection. First, recall that the expected return on the minimum variance portfolio is $\frac{A}{C}$, and that this return defined the threshold used to determine whether a given frontier portfolio is efficient or inefficient. As long as $r_f \neq \frac{A}{C}$, we have portfolio separation. The statement is stronger here than for the case of risky asset only. In particular, we can state that: "Any portfolio on the efficient frontier can be replicated by a portfolio that holds only the riskless asset and a portfolio on the risky-asset-only frontier. This result implies, if all investor agree on the distribution of asset returns (homogenous beliefs) and, therefore, the linear efficient frontier, that they will all choose to hold risky asset in the same relative proportion given by the tangency portfolio". This is a hugely important result and implies that *all* investors hold the tangency portfolio, but differ in their allocation of wealth to it relative to the risk-free asset depending on their degree of risk aversion.

Figure 4: Portfolio separation and indifference curves

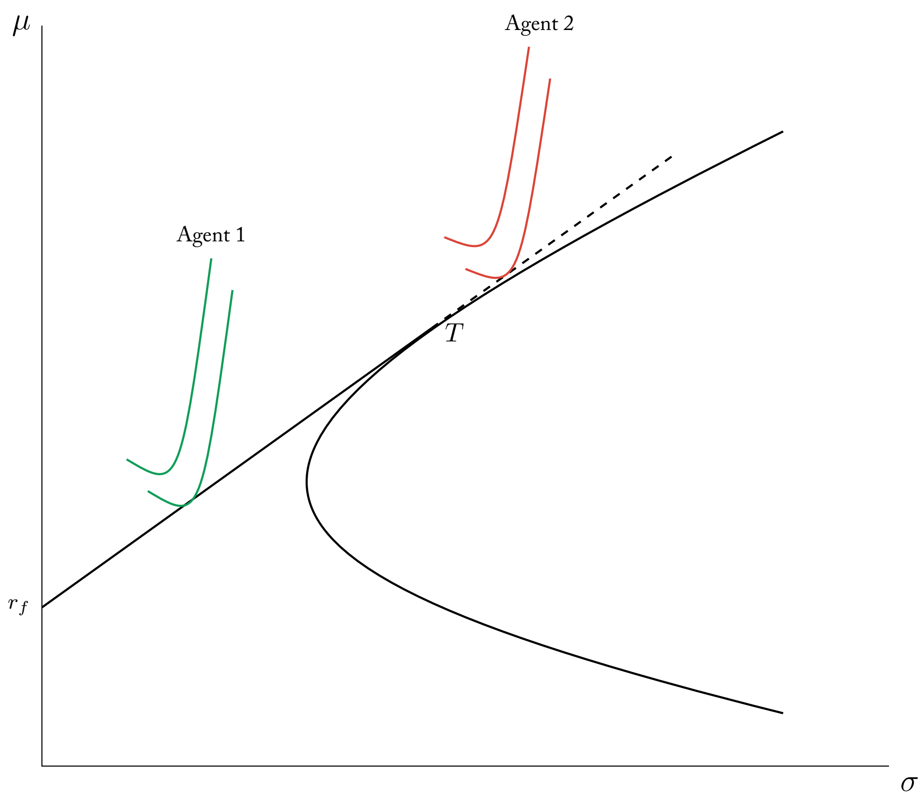
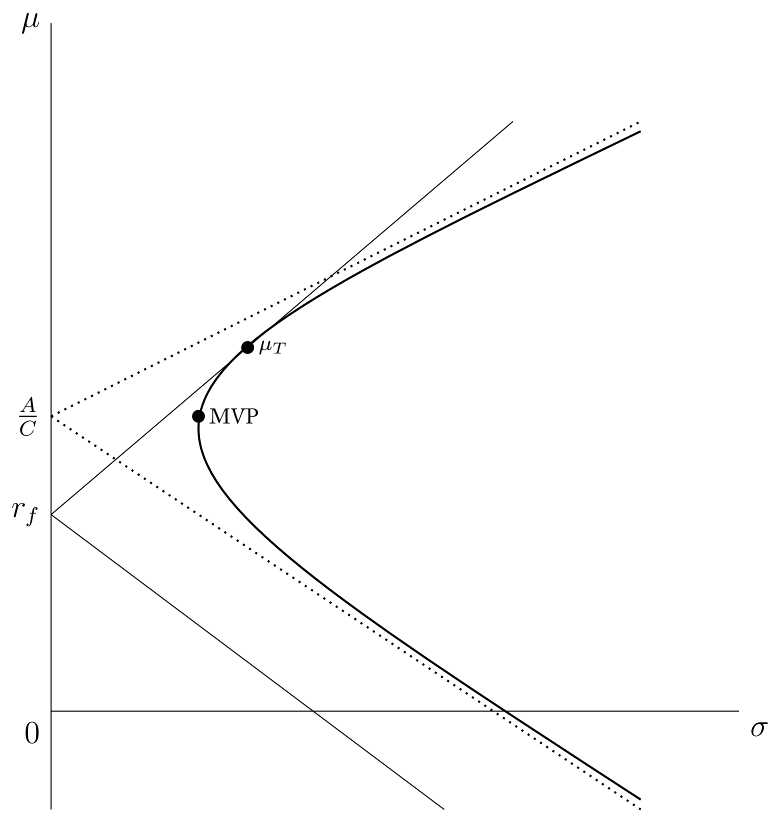


Figure 4 illustrates this idea using indifference curves for investors with varying degrees of risk aversion. For example, suppose that the green curves represents a more risk averse investor and the red a less risk averse investor. In this case, we see that the less risk averse investor invests more in the tangency portfolio of risky assets compared to the more risk averse investor.

We note that the statement that separation exists whenever $r_f \neq \frac{A}{C}$ implies the existence of three special cases, namely: i) $r_f < \frac{A}{C}$, ii) $r_f > \frac{A}{C}$, and iii) $r_f = \frac{A}{C}$. We detail these cases below, and demonstrate why we cannot have separation in the last case.

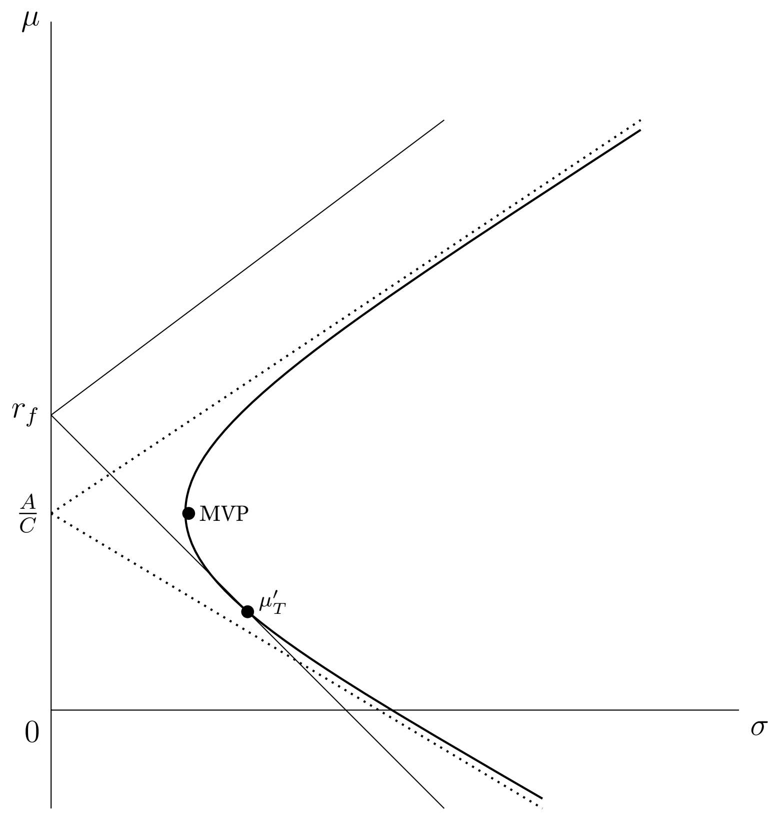
Case 1: Risk-free rate is less than A/C The case in which $r_f < \frac{A}{C}$ is the one most frequently encountered in textbook treatments of the mean-variance model (and in applications in general). Let r_f be the return on the risk-free asset and let μ_T (the expected return on the tangency portfolio) denote the tangent point of the half line $r_f + \sigma_p H^{\frac{1}{2}}$ and the portfolio frontier of all risky assets. Figure 5 illustrates the portfolio separation result. Any portfolio on the line segment from r_f to μ_T is a linear combination of the tangency portfolio μ_T and the risk-free asset r_f. Any portfolio on the half line $r_f + \sigma_p H^{\frac{1}{2}}$ beyond the tangent point μ_T involves short-selling the riskless asset and investing the proceeds in μ_T. That is, applying leverage. Finally, any portfolio on the half line $r_f - \sigma_p H^{\frac{1}{2}}$ involves short-selling the tangency portfolio μ_T and investing the proceeds in the riskless asset. Note that no rational investor would be on this downward sloping half line as they are all mean-variance dominated by the portfolios on the upward sloping half line.

Figure 5: Case 1: Risk-free rate is less than A/C



Case 2: Risk-free rate is larger than A/C The story for the second special case in which $r_f > A/C$ is slightly different, but altogether familiar. Figure 6 illustrates the intuition. Any portfolio on the half line $r_f + \sigma_p H^{\frac{1}{2}}$ involves short-selling μ'_T and investing the proceeds in the riskless asset. In this case, μ'_T is inefficient because $\mu'_T < \frac{A}{C}$ and r_f is efficient, making it the clear investment choice. Any portfolio on the half line $r_f - \sigma_p H^{\frac{1}{2}}$, on the other hand, involves a long position in the portfolio μ'_T. Note, however, this this half line remain inefficient.

Figure 6: Case 2: Risk-free rate is larger than A/C



Case 3: Risk-free rate is equal to A/C The final special case where $r_f = A/C$ is somewhat peculiar. In particular, this case implies that

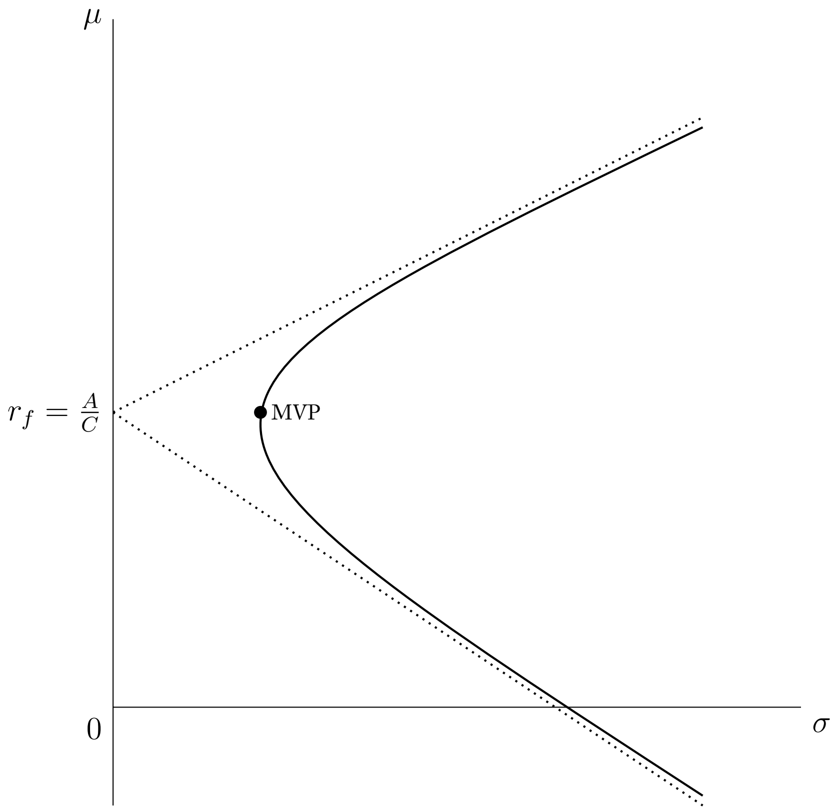
$$
\begin{aligned}
H &= B - 2Ar_f + Cr_f^2 && (104)\\
&= B - 2A\left(\frac{A}{C}\right) + C\left(\frac{A}{C}\right)^2 && (105)\\
&= \frac{BC - A^2}{C} = \frac{D}{C} > 0, && (106)
\end{aligned}
$$

which in turn implies that the risky asset portfolio frontier is generated as

$$\mu_p = r_f \pm \sigma_p \left(\frac{D}{C}\right)^{\frac{1}{2}}. \tag{107}$$

However, this is nothing but the equation for the asymptotes of the portfolio frontier for risky assets (drawn by dotted lines in the figures), which implies that the tangency portfolio can no longer be identified. As a result, there is no portfolio separation and the efficient frontier cannot

Figure 7: Case 3: Risk-free rate is equal to A/C



be generated by taking positions in the risk-free asset and the tangency portfolio (because the latter does not exist). The question then becomes: "How is the efficient frontier of all assets generated?" We can answer this question by substituting $r_f = A/C$ into the expression for the optimal portfolio weights in (98) and pre-multiplying wit $\mathbf{1}^\top$ to obtain

$$\mathbf{1}^\top \boldsymbol{\omega}_p^* = \frac{\mu_p - r_f}{H} \mathbf{1}^\top \mathbf{V}^{-1} \left(\boldsymbol{\mu} - \frac{A}{C} \mathbf{1} \right) \tag{108}$$

$$= \frac{\mu_p - r_f}{H} \left(A - \frac{A}{C} C \right) \tag{109}$$

$$= 0. \tag{110}$$

Any portfolio on the efficient frontier of all assets therefore involves investing everything in the riskless asset and holding an *arbitrage portfolio* of risky assets, that is, a portfolio whose weights sum to zero. This is sometimes also referred to as a zero-cost or zero net-investment portfolio.

References

Arrow, K.J., 1971. The theory of risk aversion, in: Essays in the Theory of Risk Bearing. Markham Publishing Co., Chicago, IL, pp. 90–120.

Campbell, J.Y., 2017. Financial decisions and markets: A course in asset pricing. Princeton University Press.

Gollier, C., 2001. The economics of risk and time. MIT Press, Cambridge MA.

Harvey, C.R., Siddique, A., 2002. Conditional skewness in asset pricing tests. Journal of Finance 55, 1263–1295.

Huang, C., Litzenberger, R.H., 1988. Foundations for financial economics. Prentice Hall.

Kraus, A., Litzenberger, R.H., 1976. Skewness preference and the valuation of risk assets. Journal of Finance 31, 1085–1100.

Markowitz, H.M., 1952. Portfolio selection. Journal of Finance 7, 77–91.

Markowitz, H.M., 1959. Portfolio selection: Efficient diversification of investments. John Wiley, New York.

Markowitz, H.M., 2010. Portfolio theory: As I still see it. Annual Review of Financial Economics 2, 1–23.

Merton, R.C., 1972. An analytical derivation of the efficient portfolio frontier. Journal of Financial and Quantitative Analysis 7, 1851–1872.

Pennacchi, G., 2008. Theory of asset pricing. Pearson/Addison-Wesley.

Pratt, J.W., 1964. Risk aversion in the small and in the large. Econometrica 32, 122–136.

Tobin, J., 1958. Liquidity preference as behavior towards risk. Review of Economic Studies 25, 65–86.

von Neumann, J., Morgenstern, O., 1944. Theory of Games and Economic Behavior. Princeton University Press.