Financial Analytics
Imperial College London

Mean-Variance Optimization and the CAPM

These lecture notes provide an introduction to mean-variance analysis and the capital asset pricing model (CAPM). We begin with the mean-variance analysis of Markowitz (1952) when there is no risk-free asset and then move on to the case where there is a risk-free asset available. We also discuss the difficulties of implementing mean-variance analysis in practice and outline some approaches for resolving these difficulties. Because optimal asset allocations are typically very sensitive to estimates of expected returns and covariances, these approaches typically involve superior estimation methods or (robust) optimization models that explicitly account for parameter uncertainty.

Mean-variance analysis leads directly to the *capital asset pricing model* or CAPM. The CAPM is a one-period equilibrium model that provides many important insights to the problem of asset pricing. The language and terminology associated with the CAPM has become ubiquitous in finance.

1 Markowitz's Mean-Variance Analysis

Consider a one-period financial market with n securities which have identical expected returns and variances, i.e. $\mathrm{E}[R_i] = \mu$ and $\mathrm{Var}(R_i) = \sigma^2$ for $i = 1, \ldots, n$ where R_i is the return on security i. We also suppose $\mathrm{Cov}(R_i, R_j) = 0$ for all $i \neq j$. Let w_i denote the fraction of wealth invested in the i^{th} security at time $t = 0$. Note that we must have $\sum_{i=1}^n w_i = 1$ for any portfolio. Consider now two portfolios:

Portfolio A: 100% invested in security # 1 so that $w_1 = 1$ and $w_i = 0$ for $i = 2, \ldots, n$.

Portfolio B: An equi-weighted portfolio so that $w_i = 1/n$ for $i = 1, \ldots, n$.

Let R_A and R_B denote the random returns of portfolios A and B, respectively. Then

$$\begin{aligned} R_A &= R_1 \\ R_B &= \frac{1}{n}\sum_{i=1}^n R_i \end{aligned}$$

and from this it follows that

$$\begin{aligned} \mathrm{E}[R_A] &= \mathrm{E}[R_B] = \mu \\ \mathrm{Var}(R_A) &= \sigma^2 \\ \mathrm{Var}(R_B) &= \sigma^2/n. \end{aligned}$$

The two portfolios therefore have the same expected return but very different return variances. A risk-averse investor should clearly prefer portfolio B because this portfolio benefits from diversification without sacrificing any expected return. This was the cental insight of Markowitz in 1958 when he recognized that investors should look to minimize variance for a given level of expected return or, equivalently, look to maximize expected return for a given level of variance.

Before formulating and solving the mean variance problem consider Figure 1 below. There were $n = 6$ securities with given mean returns, variances and covariances. We generated $m = 200$ random portfolios from these n securities and computed the expected return and volatility, i.e. standard deviation, for each of them. They are plotted in the figure and are labelled "inefficient". This is because every one of these random portfolios can be improved. In particular, for the same expected return it is possible to find a portfolio with a smaller volatility. Alternatively, for the same volatility it is possible to find a portfolio with a higher expected return.

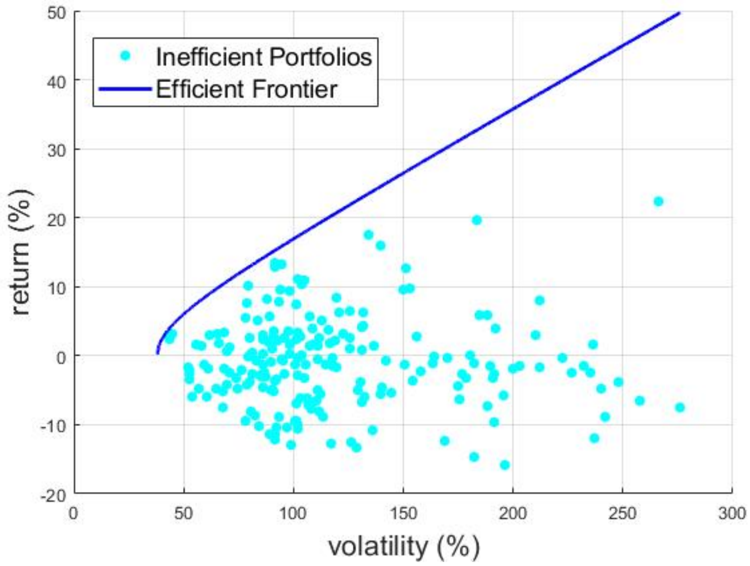

Figure 1: Sample portfolios and the efficient frontier without a risk-free security.

1.1 The Efficient Frontier without a Risk-free Asset

We will consider first the mean-variance problem when a risk-free security is not available. Again we have n risky securities with the corresponding return vector $\mathbf{R}$ being multivariate normally distributed with mean return vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, i.e.

$$\mathbf{R} \sim \mathsf{MVN}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{1}$$

The mean-variance portfolio optimization problem is formulated as:

$$\min_{\mathbf{w}} \frac{1}{2}\mathbf{w}^\top \boldsymbol{\Sigma}\mathbf{w} \tag{2}$$

$$\begin{aligned} \text{subject to} \quad & \mathbf{w}^\top \boldsymbol{\mu} = p \\ \text{and} \quad & \mathbf{w}^\top \mathbf{1} = 1. \end{aligned}$$

Note that the specific value of p will depend on the risk aversion of the investor. This is a quadratic optimization problem and because $\boldsymbol{\Sigma} \succeq 0$, i.e. it's a positive-definite matrix, it is also a convex optimization problem and has a unique local optimum solution which is the global optimum. This solution can be solved via standard Lagrange multiplier methods.

When we plot the mean portfolio return p against the corresponding minimal portfolio volatility / standard deviation we obtain the so-called **portfolio frontier**. We can also identify the portfolio having minimal variance among all risky portfolios: this is called the **minimum variance portfolio**. The points on the portfolio frontier with expected returns greater than the minimum variance portfolio's expected return, $\bar{R}_{mv}$ say, are said to lie on the **efficient frontier**. The efficient frontier is plotted as the upper blue curve in Figure 1 or alternatively, the blue curve in Figure 2.

The following theorem is a famous result from mean-variance theory and is sometimes credited with helping give rise to the mutual fund industry. It is sometimes referred to as a **2-fund theorem**.

Theorem 1 *Let $\mathbf{w}_1$ and $\mathbf{w}_2$ be mean-variance efficient portfolios corresponding to expected returns r_1 and r_2, respectively, with $r_1 \neq r_2$. Then all* **all** *efficient portfolios can be obtained as linear combinations of $\mathbf{w}_1$ and $\mathbf{w}_2$.*

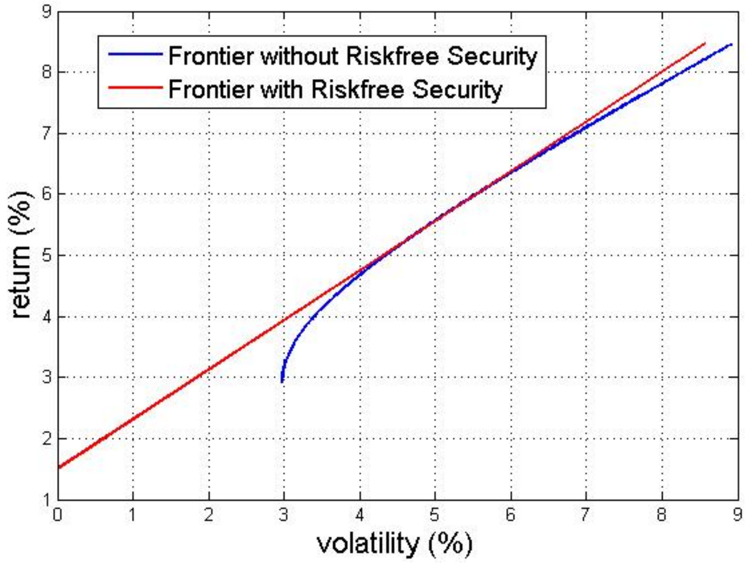

Figure 2: The efficient frontier with a risk-free security.

1.2 The Efficient Frontier with a Risk-free Asset

We now assume there is a risk-free security available with risk-free rate r_f. Let $\mathbf{w} := [w_1 \cdots w_n]^\top$ be the vector of portfolio weights on the n risky assets so that $1 - \sum_{i=1}^n w_i$ is the weight on the risk-free security. An investor's portfolio optimization problem may then be formulated as

$$\min_{\mathbf{w}} \ \frac{1}{2}\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \tag{3}$$

$$\text{subject to} \quad \left(1 - \sum_{i=1}^n w_i\right) r_f + \mathbf{w}^\top \boldsymbol{\mu} \ = \ p.$$

The optimal solution to (3) can again be found via Lagrange multiplier methods and is given by

$$\mathbf{w} = \xi \ \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f \mathbf{1}) \tag{4}$$

where $\xi := \sigma^2_{min}/(p - r_f)$ and σ^2_{min} is the minimized variance, i.e., twice the value of the optimal objective function in (3). It satisfies

$$\sigma^2_{min} \ = \ \frac{(p - r_f)^2}{(\boldsymbol{\mu} - r_f\mathbf{1})^\top \ \boldsymbol{\Sigma}^{-1} \ (\boldsymbol{\mu} - r_f\mathbf{1})} \tag{5}$$

where $\mathbf{1}$ is an $n \times 1$ vector of ones. While ξ (or p) depends on the investor's level of risk aversion it is often inferred from the market portfolio. For example, if we take $p - r_f$ to denote the average excess market return and σ^2_{min} to denote the variance of the market return, then we can take $\sigma^2_{min}/(p - r_f)$ as the average or market value of ξ.

We now assume[1] that $r_f < \bar{R}_{mv}$. When we allow our portfolio to include the risk-free security the efficient frontier becomes a straight line with a y-intercept equal to the risk-free rate r_f. This is plotted as the red line in Figure 2 and should be clear from (5) where we see that σ_{min} is linear in p. In particular, the red line in Figure 2 is given by the equation

$$\sigma_{min} = \frac{(p - r_f)}{\sqrt{(\boldsymbol{\mu} - r_f\mathbf{1})^\top \ \boldsymbol{\Sigma}^{-1} \ (\boldsymbol{\mu} - r_f\mathbf{1})}}$$

[1]The two other possible cases are $r_f = \bar{R}_{mv}$ and $r_f > \bar{R}_{mv}$. These cases are slightly more complicated and we will not discuss since they typically would not occur in practice and indeed those cases are not possible in the equilibrium setting of the CAPM.

which is a line in (p, σ_{min}) space. A further property of the efficient frontier is that it is tangent to the *risky* efficient frontier, i.e. the blue curve in Figure 2.

Note that this result is a **1-fund theorem** since every investor will optimally choose to invest in a combination of the risk-free security and a single risky portfolio, i.e. the **tangency portfolio**. The tangency portfolio, $\mathbf{w}^*$, is given by the optimal $\mathbf{w}$ of (4) except that it must be scaled so that its component sum to 1. (This scaled portfolio will not depend on p.)

Exercise 1 *Without using (5) show that the efficient frontier is indeed a straight line as described above. Hint: consider forming a portfolio of the risk-free security with* any *risky security or risky portfolio. Show that the mean and standard deviation of the portfolio varies linearly with α where α is the weight on the risk-free-security. The conclusion should now be clear and in fact this argument also explains why the efficient frontier is tangent to the risky efficient frontier.*

Exercise 2 *Describe the efficient frontier if no borrowing is allowed.*

The **Sharpe ratio** of a portfolio (or security) is the ratio of the expected excess return of the portfolio to the portfolio's volatility. The **Sharpe optimal portfolio** is the portfolio with maximum Sharpe ratio. It is straightforward to see in our mean-variance framework (with a risk-free security) that the tangency portfolio, $\mathbf{w}^*$, is the Sharpe optimal portfolio.

1.3 Including Portfolio Constraints and Transaction Costs

We can easily include linear portfolio constraints in the problem formulation and still easily solve the resulting quadratic program. No-borrowing or no short-sales constraints are examples of linear constraints as are leverage and sector constraints. Likewise we can include (quadratic) transaction costs in the objective function that penalize trades away from the current portfolio. While analytic solutions are generally no longer available with these extensions, the resulting problems are still easy to solve numerically. In particular, we can still determine the efficient frontier numerically.

1.4 Problems with Traditional Mean-Variance Analysis

The traditional mean-variance analysis of Markowitz has some important weaknesses when applied naively in practice. They include the tendency to produce extreme portfolios combining extreme short positions with extreme long positions. As a result, portfolio managers (correctly) do not trust these extreme portfolios. This problem is invariably caused by estimation errors in the mean return vector and covariance matrix.
Consider Figure 3, for example, where we have plotted the same efficient frontier (of risky securities) as in Figure 2. In practice, investors can never compute this frontier since they do not know the true $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The best we can hope to do is to approximate it. But how might we do this? One approach would be to simply estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ using historical data. Each of the black dashed curves in Figure 3 is an **estimated frontier** that we computed by:

1. Simulating $m = 24$ sample returns from the true (in this case, multivariate normal) distribution
2. Using this simulated data to construct $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ which are estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively.
3. Using $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ to generate the estimated frontier.

Note that the blue curve in Figure 3 is the **true frontier** computed using the true mean vector and covariance matrix. The first observation is that the estimated frontiers are quite random and can differ greatly from the true frontier. They may lie below or above the true frontier or they may cross it. The **average estimated frontier** will generally[2] lie above the true frontier, however. An investor who naively uses an estimated frontier to make investment decisions may end up choosing a poor portfolio. How poor? The dashed red curves in Figure 3 are the **realized frontiers** that depict the true portfolio mean - volatility tradeoff that results from

[2]This can be made precise in a mathematical sense. It represents an example of what is sometimes called the *curse of optimization* where the act of optimizing causes estimation errors to be magnified.

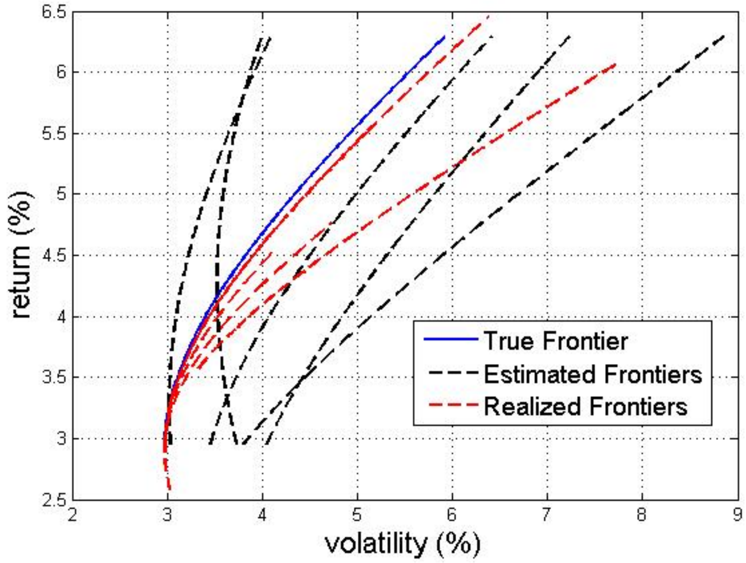

Figure 3: The true efficient frontier, estimated frontiers and realized frontiers.

making decisions based on the estimated frontiers. That is points on the realized frontier are given by $\left(\mathbf{w}^\top \boldsymbol{\mu}, \sqrt{\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}}\right)$ where $\mathbf{w}$ is an efficient portfolio for the *estimated parameters* $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$.

In contrast to the estimated frontiers, the realized frontiers must always (why?) lie below the true frontier. In Figure 3 some of the realized frontiers lie very close to the true frontier and so in these cases an investor might do very well. But in other cases the realized frontier is far from the (generally unobtainable) true efficient frontier.

These examples serve to highlight the importance of estimation errors in any asset allocation procedure. Note also that if we had assumed a heavy-tailed distribution for the true distribution of portfolio returns then we might expect to see an even greater variety of sample mean-standard deviation frontiers. In addition, it's worth emphasizing that in practice we may not have as many as 24 *relevant* observations available. For example, if our data observations are weekly returns, then using 24 of them to estimate the joint distribution of returns is hardly a good idea since we are generally more concerned with estimating *conditional* return distributions and so more weight should be given to more recent returns. A more sophisticated estimation approach should therefore be used in practice. More generally, it must be stated that estimating expected returns using historical data is very problematic and is not advisable!

1.5 Addressing these Estimation Problems

As a result of these weaknesses, portfolio managers traditionally have had little confidence in mean-variance analysis and therefore applied it very rarely in practice. There have been many efforts to overcome these problems over the years. These efforts include:

1. The use of better estimation techniques such as **shrinkage estimators**. These estimators shrink the standard estimators towards some constant vector (in the case of $\widehat{\boldsymbol{\mu}}$) or some constant matrix (in the case of $\widehat{\boldsymbol{\Sigma}}$). While shrinkage introduces some bias into the resulting estimators the latter are less noisy and can lead to better outcomes, i.e. portfolios in this case.

2. The use of Bayesian techniques such as the **Black-Litterman** framework introduced in the early 1990's. In addition to mitigating the problem of extreme portfolios (which typically occur because some components of $\widehat{\boldsymbol{\mu}}$ are extreme), the Black-Litterman framework allows users to specify their own subjective views on the market in a consistent and tractable manner.

3. The development of **robust optimization** techniques to explicitly handle parameter uncertainty in the portfolio optimization problem. In Figure 4 we show an estimated frontier that was computed using a

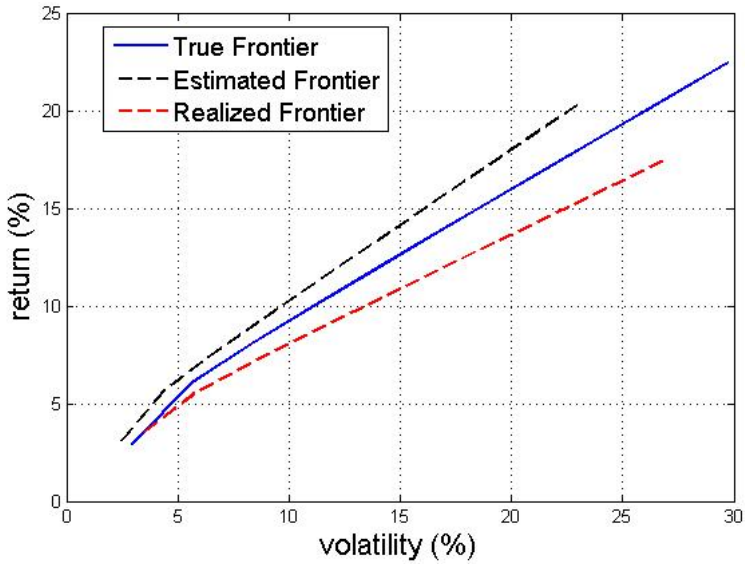

Figure 4: Robust Estimation of the Efficient Frontier.

robust optimization formulation of the mean-variance problem. We see that it lies much closer to the true frontier which is also the case with it's corresponding realized frontier.

Many of these techniques are now used routinely in general asset allocation settings. It is worth mentioning that the problem of extreme portfolios can also be mitigated in part by placing no short-sales and / or no-borrowing constraints on the portfolio.

1.6 Portfolio Management Relative to a Benchmark

In practice it's quite common for portfolio managers to assess performance relative to a fixed **benchmark portfolio** $\mathbf{w}_B$. This benchmark portfolio typically represents a particular asset class. A **passive manager** working in this asset class, e.g. U.S. equities, global equities, emerging market equities etc., might aim to replicate this benchmark whereas an **active manager** would aim to outperform the benchmark. Our mean-variance framework can be easily adapted to this latter case. In particular, the expected return is now replaced by the expected **excess return** where the excess return is $\mathbf{R}^\top(\mathbf{w}-\mathbf{w}_B)$, the portfolio return minus the benchmark return. Similarly the return variance is replaced by the variance of the **tracking error**, i.e. $\mathrm{Var}(\mathbf{R}^\top(\mathbf{w}-\mathbf{w}_B))$. We still end[3] up with a convex quadratic optimization problem which is easy to solve. For example, a passive asset manager might solve

$$\min_{\mathbf{w}} \ \frac{1}{2}(\mathbf{w}-\mathbf{w}_B)^\top \boldsymbol{\Sigma}(\mathbf{w}-\mathbf{w}_B) \tag{6}$$

$$\text{subject to} \quad \mathbf{w}^\top \mathbf{1} = 1$$

whereas an active manager might solve

$$\max_{\mathbf{w}} \ (\mathbf{w}-\mathbf{w}_B)^\top \boldsymbol{\mu} \tag{7}$$

$$\text{subject to} \quad \frac{1}{2}(\mathbf{w}-\mathbf{w}_B)^\top \boldsymbol{\Sigma}(\mathbf{w}-\mathbf{w}_B) \leq \sigma^2$$

$$\text{and} \quad \mathbf{w}^\top \mathbf{1} = 1.$$

[3]See Chapter 6 of *Optimization Methods in Finance* by Cornuéjols, Peña and Tütüncü for further details.

Of course it's straightforward to also account for transactions costs and other linear constraints (as described earlier in Section 1.3) in these formulations. In fact, the optimal solution to (6) is trivially $\mathbf{w} = \mathbf{w}_B$ unless we include transaction costs or some other portfolio constraints. Note that (6) does not involve $\boldsymbol{\mu}$ and so in practice is a much easier problem to solve than (7) which would require us to estimate $\boldsymbol{\mu}$ and as we know, that's a difficult exercise!

---

2 The Capital Asset Pricing Model (CAPM)

If every investor is a mean-variance optimizer then we can see from Figure 2 and our earlier discussion that each of them will hold the same tangency portfolio of risky securities in conjunction with a position in the risk-free asset. Because the tangency portfolio is held by all investors and because markets must clear, we can identify this portfolio as the **market portfolio**. The efficient frontier is then termed the **capital market line**. We have the following theorem.

Theorem 2 (CAPM)
Let R_m and $\bar{R}_m$ denote the return and expected return, respectively, of the market, i.e. tangency, portfolio. Then there is a linear relationship between the expected return, $\bar{R} = \mathrm{E}[R]$ say, of any security (or portfolio) with random return R and the expected return of the market portfolio. Specifically, we have

$$\bar{R} = r_f + \beta\,(\bar{R}_m - r_f) \tag{8}$$

where

$$\beta := \frac{\mathrm{Cov}(R, R_m)}{\mathrm{Var}(R_m)}.$$

Proof: In order to prove (8), consider a portfolio with weights α and weight $1-\alpha$ on the risky security and market portfolio, respectively. Let R_α denote the (random) return of this portfolio as a function of α. We then have

$$\begin{aligned} \mathrm{E}\left[R_\alpha\right] &= \alpha\bar{R} + (1-\alpha)\bar{R}_m && (9)\\ \sigma^2_{R_\alpha} &= \alpha^2\sigma^2_R + (1-\alpha)^2\sigma^2_m + 2\alpha(1-\alpha)\sigma_{R,R_m} && (10) \end{aligned}$$

where $\sigma^2_{R_\alpha}$, σ^2_R and σ^2_m are the return variances of the portfolio, security and market portfolio, respectively. We use σ_{R,R_m} to denote $\mathrm{Cov}(R, R_m)$. Now note that as α varies, the mean and standard deviation, $(\mathrm{E}\left[R_\alpha\right], \sigma_{R_\alpha})$, trace out a curve that cannot (why?) cross the efficient frontier. This curve is depicted as the dashed curve in Figure 5 below.
Therefore at $\alpha = 0$ this curve must be tangent to the capital market line. Therefore the slope of the curve at $\alpha = 0$ must equal the slope of the capital market line. Using (9) and (10) we see the former slope is given by

$$\begin{aligned} \left.\frac{d\,\mathrm{E}[R_\alpha]}{d\,\sigma_{R_\alpha}}\right|_{\alpha=0} &= \left.\frac{d\,\mathrm{E}[R_\alpha]}{d\,\alpha}\Big/\frac{d\,\sigma_{R_\alpha}}{d\,\alpha}\right|_{\alpha=0}\\ &= \left.\frac{\sigma_{R_\alpha}\left(\bar{R}-\bar{R}_m\right)}{\alpha\sigma^2_R - (1-\alpha)\sigma^2_m + (1-2\alpha)\sigma_{R,R_m}}\right|_{\alpha=0}\\ &= \frac{\sigma_m\left(\bar{R}-\bar{R}_m\right)}{-\sigma^2_m + \sigma_{R,R_m}}. && (11) \end{aligned}$$

The slope of the capital market line is $\left(\bar{R}_m - r_f\right)/\sigma_{R_m}$ and equating the two therefore yields

$$\frac{\sigma_m\left(\bar{R}-\bar{R}_m\right)}{-\sigma^2_m + \sigma_{R,R_m}} = \frac{\bar{R}_m - r_f}{\sigma_m} \tag{12}$$

which upon simplification gives (8). ■

The central insight of the CAPM is that in equilibrium the riskiness of an asset is not measured by the standard deviation of its return but by its **beta**. The CAPM result is one of the most famous results in all of finance and,

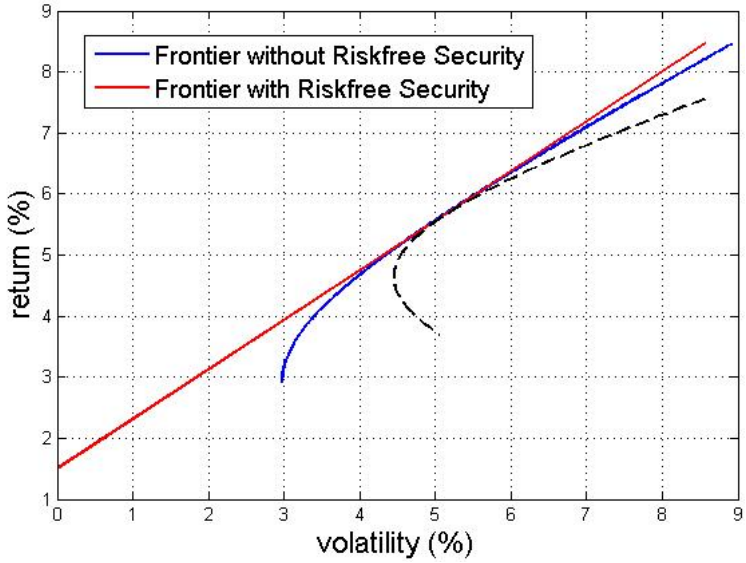

Figure 5: Proving the CAPM relationship.

even though it arises from a simple one-period model, it provides considerable insight to the problem of asset-pricing. For example, it is well-known that riskier securities should have higher expected returns in order to compensate investors for holding them. But how do we measure risk? Counter to the prevailing wisdom at the time at which the CAPM was developed, the riskiness of a security is not measured by its return volatility. Instead it is measured by its beta which is proportional to its covariance with the market portfolio. This is a very important insight. Nor, it should be noted, does this contradict the mean-variance formulation of Markowitz where investors do care about return variance. Indeed, we derived the CAPM from mean-variance analysis!

Exercise 3 *Why does the CAPM result not contradict the mean-variance problem formulation where investors do measure a portfolio's risk by its variance?*

The CAPM is an example of a so-called 1-factor model with the market return playing the role of the single factor. Other factor models can have more than one factor. For example, the Fama-French model has three factors, one of which is the market return. Many empirical research papers have been written to test the CAPM. Such papers usually perform regressions of the form

$$R_i - r_f \;=\; \alpha_i + \beta_i\,(R_m - r_f) + \epsilon_i \tag{13}$$

where α_i (not to be confused with the α we used in the proof of (8)) is the intercept and ϵ_i is the **idiosyncratic** or residual risk which is assumed to be independent of R_m and the idiosyncratic risk of other securities. If the CAPM holds then we should be able to reject the hypothesis that $\alpha_i \neq 0$. The evidence in favor of the CAPM is mixed. But the language inspired by the CAPM is now found throughout finance. For example, we use β's to denote factor loadings and α's to denote excess returns even in non-CAPM settings.

Definition 1 *Within the context of the CAPM, the efficient frontier (the red line in Figure 5) is often called the* **capital market line***. The slope of the capital market line is $K := (\bar{R}_m - r_f)/\sigma_m$ and is often called the* **market price-of-risk**.

The market price-of-risk tells you how much expected return you should receive in equilibrium per unit of volatility.

Definition 2 *The* **security market line** *is simply a plot of $\bar{R}$ as a function of β. This function is given by the CAPM relation (8).*