2. Mean-variance Equivalence and the CAPM

In this section we prove that when asset returns can be assumed to follow an elliptical distribution, then even if the investor's utility function is very complicated, the problem of maximizing expected utility can be reduced to maximizing a multivariate quadratic function. For such distributions, the optimal expected-utility portfolio is the mean-variance portfolio. First, we need to recall some notation and concepts from the previous lecture.

2.1. **Mean-variance equivalence.** Let $\boldsymbol{h} \in \mathbb{R}^n$ denote the portfolio holdings, measured in dollars or an appropriate numeraire currency, at some time t in the future. Let $\boldsymbol{h}_0$ denote the current portfolio. let $\boldsymbol{r} \in \mathbb{R}^n$ denote the return over the interval $[t, t+1]$. Hence $\boldsymbol{r} \in \mathbb{R}^n$ is an n-dimensional vector whose i-th component is

$$r_i = p_i(t+1)/p_i(t) - 1$$

where $p_i(t)$ is the i-th asset's price at time t (adjusted for splits or capital actions if necessary).

Hence the (one-period) wealth random variable is

$$\tilde{w} = \boldsymbol{h}'\boldsymbol{r}$$

and the expected-utility maximizer chooses $\boldsymbol{h}$ to satisfy

$$\boldsymbol{h}^* = \operatorname{argmax} \mathbb{E}[u(\tilde{w})] \tag{2.1}$$

Definition 2.1. A utility function $u : \mathbb{R} \to \mathbb{R}$ is called *standard* if it is increasing, concave, and continuously differentiable.

These properties make economic sense. Even great philanthropists have increasing utility of wealth in their investment portfolio – they would prefer to be able to do more to end hunger, disease etc. Concavity corresponds to risk-aversion as we showed previously. Finally, if the utility function is not continuously differentiable, it implies that there is a certain particular level of wealth for which one penny below that is very different than one penny above.

Definition 2.2. The underlying asset return distribution, $p(\boldsymbol{r})$, is said to be *mean-variance equivalent* if first and second moments of the distribution exist, and for any standard utility function u, there exists some constant $\kappa > 0$ (where κ depends on u) such that

$$\boldsymbol{h}^* = \text{argmax}\,\{\mathbb{E}[\tilde{w}] - (\kappa/2)\mathbb{V}[\tilde{w}]\}$$

where $\boldsymbol{h}^*$ is defined by (2.1).

The multivariate Cauchy distribution is elliptical, but its moments of orders 1 and higher are all infinite/undefined. Therefore, it is not mean-variance equivalent because the requisite means and variances would be undefined.

Which distributions, then, are mean-variance equivalent? We showed previously that the normal distribution is; this is easy. Many distributions, including heavy-tailed distributions such as the multivariate Student-t, are also mean-variance equivalent, but not all two-parameter families.

Definition 2.3. For a given asset return distribution $p(\boldsymbol{r})$ having finite first and second moments, we say *expected utility is a function of mean and variance* if for any increasing utility function u, one has $\mathbb{E}[u(w_1)] = \mathbb{E}[u(w_2)]$ whenever $\mathbb{E}[w_1] = \mathbb{E}[w_2]$ and $\mathbb{V}[w_1] = \mathbb{V}[w_2]$, where w_1 and w_2 are lotteries arising from holding portfolios of risky assets.

Definition 2.3, like Definition 2.2, is a property of the asset return distribution $p(\boldsymbol{r})$; some distributions have this property, and some do not. Chamberlain (1983) has shown that the distributions satisfying Definition 2.3 are precisely the *elliptical* distributions, where the term 'elliptical' is given below in Definition 2.5.

Definition 2.4. When Definition 2.3 holds, an investor's preferences among portfolios depend only on the expected return and variance of those portfolios. Defining $\mu = \boldsymbol{h}'\mathbb{E}[\boldsymbol{r}]$ and $\sigma^2 = \boldsymbol{h}'\boldsymbol{\Omega}\boldsymbol{h}$ then the expected utility of holding portfolio h can be represented as some function $Q(\mu, \sigma)$. An *indifference curve* is the curve in the two-dimensional (μ, σ)-plane which is a level curve of Q.

Tobin (1958) assumed that expected utility is a function of mean and variance, and showed mean-variance equivalence as a consequence. Unfortunately, Tobin's proof was flawed – it contained a derivation which is only valid for elliptical distributions. The flaw in Tobin's proof, and a counterexample, was pointed out by Feldstein (1969). After presenting a correct proof, we will discuss the flaw.

Recall that for a scalar-valued random variable X, the *characteristic function* is defined by

$$\phi_X(t) = \mathbb{E}\left[e^{itX}\right],$$

If the variable has a density, then the characteristic function is the Fourier transform of the density, and it is a special case of the Fourier transform of a measure. The characteristic function of a real-valued random variable always exists, since it is the integral of a bounded continuous function over a finite measure space.

Generally speaking, characteristic functions are especially useful when analyzing moments of random variables, and linear combinations of random variables. Characteristic functions have been used to provide especially elegant proofs of some of the key results in probability theory, such as the central limit theorem.

If a random variable X has moments up to order k, then the characteristic function ϕ_X is k times continuously differentiable on $\mathbb{R}$. In this case

$$\mathbb{E}[X^k] = (-i)^k \phi_X^{(k)}(0).$$

If ϕ_X has a k-th derivative at zero, then X has all moments up to k if k is even, but only up to $k-1$ if k is odd, and

$$\phi_X^{(k)}(0) = i^k \mathbb{E}[X^k]$$

If $X_1, \ldots, X_n$ are independent random variables, then

$$\phi_{X_1+\cdots+X_n}(t) = \phi_{X_1}(t) \cdots \varphi_{X_n}(t).$$

Definition 2.5. A multivariate density on $\mathbb{R}^n$ is said to be *elliptical* if its characteristic function, defined by $\phi(\mathbf{t}) := \mathbb{E}[\exp(i\,\mathbf{t}'\mathbf{x})]$ takes the form

$$\phi(\mathbf{t}) = \exp(i\,\mathbf{t}'\boldsymbol{\mu})\psi(\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}) \tag{2.2}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ is the vector of medians, and $\boldsymbol{\Omega}$ is a matrix, assumed to be positive definite, known as the *dispersion matrix*. The function ψ does not depend on n.

The name “elliptical” arose because the isoprobability contours are ellipsoidal. Ingersoll (1987) points out that if variances exist, then the covariance matrix is proportional to $\mathbf{\Omega}$, and if means exist, $\boldsymbol{\mu}$ is also the vector of means.

Eq. (2.2) does not imply that the random vector $\mathbf{x}$ has a density, but if it does, then the density must be of the form

$$f_n(\mathbf{x}) = |\mathbf{\Omega}|^{-1/2}\, g_n\left[(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Omega}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right] \tag{2.3}$$

Eq. (2.3) is sometimes used as the definition of elliptical distributions, when existence of a density is assumed.

The multivariate normal is the most well-known elliptical family; for the normal, one has $g_n(s) = c_n \exp(-s/2)$ (where c_n is a normalization constant) and $\psi(T) = \exp(-T/2)$. Note that g_n depends on n while ψ does not. The elliptical class also includes many non-normal distributions, including examples which display heavy tails, and are therefore better suited to modeling asset returns. For example, the multivariate Student-t distribution with ν degrees of freedom has density of the form (2.3) with

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

and for $\nu = 1$ one recovers the multivariate Cauchy.

One could choose $g_n(s)$ to be identically zero for sufficiently large s, which would make the distribution of asset returns bounded above and below. Thus, one criticism of the CAPM – that it requires assets to have unlimited liability – is not a valid criticism.

Let $\boldsymbol{v} = \mathbf{T}\mathbf{x}$ denote a fixed (non-stochastic) linear transformation of the random vector $\mathbf{x}$. It is of interest to relate the characteristic function of $\boldsymbol{v}$ to that of $\mathbf{x}$.

$$\begin{aligned}
\phi_{\boldsymbol{v}}(\mathbf{t}) &= \mathbb{E}[e^{i\,\mathbf{t}'\boldsymbol{v}}] = \mathbb{E}[e^{i\,\mathbf{t}'\mathbf{T}\mathbf{x}}] && (2.4)\\
&= \phi_{\mathbf{x}}(\mathbf{T}'t) && (2.5)\\
&= e^{i\,\mathbf{t}'\mathbf{T}\boldsymbol{\mu}}\psi\left(\mathbf{t}'\mathbf{T}\mathbf{\Omega}\mathbf{T}'\mathbf{t}\right) && (2.6)\\
&= e^{i\,\mathbf{t}'\mathbf{m}}\psi\left(\mathbf{t}'\Delta\mathbf{t}\right) && (2.7)
\end{aligned}$$

where for convenience we define $\mathbf{m} := \mathbf{T}\boldsymbol{\mu}$ and $\mathbf{\Delta} := \mathbf{T}\mathbf{\Omega}\mathbf{T}'$.

Even with the same function ψ, the functions f_n, g_n appearing in the density (2.3) can have rather different shapes for different n (= the dimension of $\mathbb{R}^n$), and we give some examples below. However the function ψ does not depend on n. For this reason, one sometimes speaks of an elliptical “family” which is identified with a single function ψ. The single ψ corresponds to a family of functions g_n determining the densities – one such function for each dimension of Euclidean space. Above we have shown that marginalization of an elliptical family results in a new elliptical of the same family (ie. the same ψ-function).

Theorem 2.1. If the distribution of $\boldsymbol{r}$ is elliptical, and if u is differentiable and concave, then there exists a function Q such that $\mathbb{E}[u(\tilde{w})] = Q(\mu, \omega)$ where $\mu := \boldsymbol{h}'\boldsymbol{\mu}$ and $\omega^2 := \boldsymbol{h}'\boldsymbol{\Omega}\boldsymbol{h}$. Moreover,

$$Q_2(\mu, \omega) := \frac{\partial}{\partial \omega} Q(\mu, \omega) \leq 0.$$

Proof. For the duration of this proof, fix a portfolio with holdings vector $\boldsymbol{h}$ and let $Z = \boldsymbol{h}'\boldsymbol{r}$ denote the wealth increment (also denoted $\tilde{w}$ above). Let $\mu = \boldsymbol{h}'\boldsymbol{\mu}$ and $\omega^2 = \boldsymbol{h}'\boldsymbol{\Omega}\boldsymbol{h}$ denote moments of Z. Applying the marginalization property (2.7) with the $1 \times n$ matrix $\mathbf{T} = \boldsymbol{h}'$ yields

$$\phi_Z(t) = e^{i\,t\mu} \psi(t^2 \omega^2).$$

The k-th central moment of Z will be

$$i^{-k} \frac{d^k}{dt^k} \psi(t^2 \omega^2) \Big|_{t=0}$$

From this it is clear that all odd moments will be zero, and all even moments will be proportional to ω^k. Therefore the full distribution of Z is completely determined by μ, ω.

In particular, $\mathbb{E}[u(Z)]$ is completely determined by μ, ω. Hence there exists some function Q such that $\mathbb{E}[u(\tilde{w})] = Q(\mu, \omega)$. To show that $\frac{\partial}{\partial \omega} Q(\mu, \omega) \leq 0$, ie. that variance is disliked, write

$$Q(\mu, \omega) = \mathbb{E}[u(Z)] = \int_{-\infty}^{\infty} u(Z) f(Z) dZ\,. \tag{2.8}$$

Note that the integral is over a one-dimensional variable. Using the special case of eq. (2.3) with $n = 1$, we have

$$f_1(Z) = \omega^{-1} g_1[(Z - \mu)^2 / \omega^2]. \tag{2.9}$$

Using (2.9) to update (2.8), we have

$$Q(\mu, \omega) = \mathbb{E}[u(Z)] = \int_{-\infty}^{\infty} u(Z) \omega^{-1} g_1[(Z - \mu)^2 / \omega^2]\, dZ\,.$$

Now make the change of variables $\xi := (Z - \mu)/\omega$ and $dZ = \omega\, d\xi$, which yields

$$Q(\mu, \omega) = \int_{-\infty}^{\infty} u(\mu + \omega \xi) g_1(\xi^2) d\xi\,.$$

Let Q_2 denote the partial derivative of Q in the second variable, so

$$\begin{aligned} Q_2(\mu, \omega) &= \int_{-\infty}^{\infty} u'(\mu + \omega\xi)\xi g_1(\xi^2) d\xi \\ &= \int_0^{\infty} \xi g_1(\xi^2)[u'(\mu + \omega\xi) - u'(\mu - \omega\xi)] d\xi \end{aligned}$$

Note that we have changed the domain of integration from $(-\infty, \infty)$ to $(0, \infty)$. A differentiable function is concave on an interval if and only if its derivative is monotonically decreasing on that interval, hence

$$u'(\mu + \omega\xi) - u'(\mu - \omega\xi) < 0$$

while $g_1(\xi^2) > 0$ since it is a probability density function. Hence on the domain of integration, the integrand is ≤ 0 and hence $Q_2(\mu, \omega) \leq 0$, completing the proof of Theorem 2.1.

Recall Definition 2.4 above of indifference curves. Imagine the indifference curves written in the σ, μ plane with σ on the horizontal axis. If there are two branches of the curve, take only the upper one. Mean-variance equivalence boils down to two statements about the indifference curves:

(1) $d\mu/d\sigma > 0$ or an investor is indifferent about two portfolios with different variances only if the portfolio with greater σ also has greater μ,
(2) $d^2\mu/d\sigma^2 > 0$, or the rate at which an individual must be compensated for accepting greater σ (this rate is $d\mu/d\sigma$) increases as σ increases

These two properties say that the indifference curves are convex.

What, exactly, fails if the distribution $p(\boldsymbol{r})$ is not elliptical? The crucial step of this proof assumes that a two-parameter distribution $f(x; \mu, \sigma)$ can be put into "standard form" $f(z; 0, 1)$ by a change of variables $z = (x - \mu)/\sigma$. This is not a property of all two-parameter probability distributions; for example, it fails for the lognormal.

One can see by direct calculation that for logarithmic utility, $u(x) = \log x$ and for a log-normal distribution of wealth,

$$f(x; m, s) = \frac{1}{sx\sqrt{2\pi}} \exp(-(\log x - m)^2/2s^2)$$

then the indifference curves are not convex. The moments of x are

$$\mu = e^{m+s^2/2}, \quad \text{and} \quad \sigma^2 = (e^{m+s^2/2})^2(e^{s^2} - 1)$$

and, with a little algebra one has

$$\mathbb{E}u = \log \mu - \frac{1}{2}\log(\sigma^2/\mu^2 + 1)$$

One may then calculate $d\mu/d\sigma$ and $d^2\mu/d\sigma^2$ along a parametric curve of the form

$$\mathbb{E}u = \text{constant}$$

and we would see that $d\mu/d\sigma > 0$ but $d^2\mu/d\sigma^2$ changes sign.

Theorem 2.1 implies that for a given level of median return, investors always dislike dispersion. We henceforth assume, unless otherwise stated, that the first two moments of the distribution exist. In this case, (for elliptical distributions) the median is the mean and the dispersion is the variance, and hence the underlying asset return distribution is mean-variance equivalent in the sense of Def. 2.2. We emphasize that this holds for any smooth, concave utility.

2.2. **The Capital Asset Pricing Model in its original form.** We begin by introducing the Capital Asset Pricing Model in its *original form* due to Sharpe (1964). We may, time permitting, return to discuss extensions due to Black and Lintner.

The Capital Asset Pricing Model is an extremely idealized model of an open market place. Everyone knows it isn't a true description of how our markets actually function, as all of the assumptions one must make to derive it are gross idealizations. Nonetheless, sometimes it's useful to work out what would be true in an ideal world. It's then instructive to understand the exact ways in which our world deviates from the predictions of the idealized model.

Here are all of the assumptions we need to make in order to derive the CAPM:

(a) an open market place, where all the risky assets are available to all
(b) there exists a risk-free asset (idealized treasury bond) for borrowing and/or lending in unlimited quantities with interest rate r_f
(c) all information is available to all such as covariances, variances, mean rates of return, etc.
(d) everyone is a risk-averse rational investor who uses Markowitz mean-variance portfolio theory
(e) there are no trading costs
(f) the market has reached equilibrium

Assumptions (c) and (d) together imply a unique, up to scalar multiplication, portfolio of risky assets. It must be proportional to $\Sigma^{-1}\mathbb{E}[r]$ where $\Sigma \in S^n_{++}$ is the common covariance matrix and r is an n-dimensional random vector denoting the next-period asset returns.

Let's spend a few moments discussing why all of these assumptions are false. For the record, I am not saying anything revolutionary here – even those responsible for the development of the CAPM and related theories understand very well that all of the assumptions are gross idealizations. This is probably only slightly worse than modeling a neuron as a sigmoid activation function applied to the output of a linear function. However, the latter approximation is extremely useful as well; without it, there would be no deep learning.

Sharpe (1990) says:

> *The initial version of the CAPM, developed over 25 years ago, was extremely parsimonious. It dealt with the central aspects of equilibrium in capital markets and assumed away many important aspects of such markets as they existed at the time.*
>
> — Sharpe (1990)

Assumption (a) is false because there are many assets that aren't tradable by everyone (think A-shares in China, shares like Berkshire with very high prices so that retail investors can't even afford one share, etc.)

Assumption (b) is perhaps the closest to being semi-realistic among assumptions (a)-(f). Of course no asset is risk-free, as governments and corporations frequently default on even relatively short-term debt. Also, individual investors do not have the same credit rating as, say, the US government and hence cannot borrow at the "low-risk rate" of a sovereign bond. No investor or institution can borrow or lend in infinite quantities. However, for institutional investors, (b) may actually be a reasonable approximation.

Assumption (c) is probably the most egregiously false of all of these, with (d) coming in a close second place. No-one knows what the asset return covariance matrix is going to look like in the future, although with a lot of effort, some still-noisy-but-useful statistical forecasts can be made. There isn't universal agreement on the "best" forecasting method, although some forecasting methods can be ruled out as obviously worthless.

The situation for the *first* moment of asset returns $\mathbb{E}[r]$ is even worse – no-one has much of an idea here, and those who do have typically come by this information via extremely labor-intensive processes and aren't prepared to freely share the results of their labor. The more expertise anyone has in predicting returns, the less inclined they are to share the results of their analysis. To come up with a statistical forecasting method which realizes out-of-sample a cross-sectional R^2 of 0.01 when forecasting residual returns to a factor model is possible, but difficult.

There is *no consensus* about $\mathbb{E}[r]$ even among sophisticated, professional investors with the same information set. Imagine the following argument happening sometime in 2001:

Investor A: I think the ipod is going to revolutionize the music industry, hence I project Apple revenue growth at X% and hence Apple is undervalued.
Investor B: I disagree; other companies already make portable MP3 players and I've observed their growth. It's less than one might have hoped. Apple is late to the game, and hence is at a disadvantage. I project Apple's revenue growth at Y% where Y is much less than X.
Investor A: I've seen all those same numbers, and I'm sticking with my original projections. This time is different.

Assumption (d) is also quite egregiously false. Most investors don't know how to use modern portfolio theory effectively, and so they don't use it, falling back instead on simple heuristics ("cut your losers, let your winners run," etc.) This is probably why the HFRX type indexes, which are supposed to be broad indexes of the hedge fund industry, have correlations of 0.7 (or more; one recent report from one of my prime brokers said 0.9) to the S&P 500. A few more sophisticated firms do use modern portfolio theory in one form or another, and realize lower correlation to the market and other factors.

Assumption (e) is also false. The more actively-traded a strategy is, the more trading costs eat into any potential alpha. Depending on exactly how they're constructed, naive implementations of Markowitz portfolios can turn over quite a lot. Arrow-Pratt utility theory calls for optimization of utility of final wealth anyway, and final wealth includes any transaction costs that are paid, so utility theory doesn't actually say to trade the (cost-unaware) Markowitz portfolio – it says we should optimize with costs in the consideration.

Assumption (f) is unrealistic; stock prices routinely move by several percents over the time span of a few minutes even when there's no news or information that should affect the firm's intrinsic value in any way. To make an analogy with thermodynamic equilibrium, real markets much more closely resemble the surface of a bubbling, churning, boiling cauldron than a flat, placid, reflective mountain lake. The settling down into a thermal equilibrium never seems to occur.

Note that we have not assumed normality of returns anywhere. Sharpe confirms:

> *The CAPM makes no assumptions about the "return generating process." Hence, its results are completely consistent with any such process.*
>
> — Sharpe (1990)

With these caveats, let's continue with deriving the mathematical implications of the above assumptions. The efficient fund proportional to $\Sigma^{-1}\mathbb{E}[r]$ and used by all is called "the market portfolio" and will henceforth be denoted by M. Let V_i be the market capitalization of the i-th company. The total number of dollars in the entire equity market is then $V = \sum_{i=1}^{n} V_i$. If all investors hold the same portfolio (as the *risky* part of their allocation), then it doesn't matter whether there are a bunch of small investors or one big one. If there were one big one, then we know the total value of her investment in risk assets: it's V. The "weights" of the one big investor's portfolio are then $w_i = V_i/V$. Hence the vector $\mathbf{w}$ must be proportional to $\Sigma^{-1}\mathbb{E}[r]$.

Note that we found the (risky part of the) one big investor's portfolio without having to know Σ and $\mathbb{E}[r]$. This argument relied on all of the (tenuous) assumptions above. Note also that in this model, there are no assets with $\mathbb{E}[r] < r_f$ and hence no need for anyone to hold any short positions, although it isn't explicitly ruled out. One can see this as follows: if $\mathbb{E}[r_i] < r_f$ at the current price of asset i, then no investors will buy the asset at that price (they'd prefer to buy the risk-free bond). Hence no-one will buy it, period, unless the price is set low enough so that *at the new* *price* $\mathbb{E}[r_i] > r_f$. Only then will the herd of like-minded investors (or one big investor) buy it, and it acquires some value $V_i > 0$.

> *When a riskless asset is available, the only negative holdings in equilibrium will involve borrowing by investors with above-average risk tolerance who wish to finance added investment in a portfolio representing the overall capital market.*
>
> — Sharpe (1990)

Since there's only one risky portfolio in this model, all investors must hold some combination of this portfolio and the risk-free asset. If we insist on weights adding to 1, then all portfolios are then of the form $1 - h$ units of the risk-free asset, and h units of M. The expected return and risk of this portfolio are:

$$\overline{r} = (1-h)r_f + h\overline{r}_M, \quad \sigma = h\sigma_M$$

where, for notational convenience, we will denote $\mathbb{E}[r]$ as $\overline{r}$, and continue this notation throughout the lecture.

We can then eliminate $h = \sigma/\sigma_M$ to find

$$\overline{r} = (1 - \frac{\sigma}{\sigma_M})r_f + \frac{\sigma}{\sigma_M}\overline{r}_M = r_f + \frac{\sigma}{\sigma_M}(\overline{r}_M - r_f) \tag{2.10}$$

Eq. (2.10) has the impressive-sounding name *the capital market line.* Before we get too impressed, though, note that all of the mathematics we have used so far (solving linear equations in one variable) was already known to the ancient Babylonians. It tells us the expected return of any efficient portfolio, in terms of its standard deviation, in terms of the so-called *price of risk*

$$\frac{\overline{r}_M - r_f}{\sigma_M} \tag{2.11}$$

Theorem 2.2. Under the assumptions above, for any asset i we have

$$\overline{r}_i - r_f = \beta_i(\overline{r}_M - r_f) \tag{2.12}$$

where $\beta_i = \mathrm{cov}(r_i, r_M)/\sigma_M^2$.

Proof. Form a portfolio with holding λ in asset i and $1-\lambda$ in the market M. Trivially, this portfolio has expected return and variance

$$\overline{r}(\lambda) = \lambda(\overline{r}_i - \overline{r}_M) + \overline{r}_M \tag{2.13}$$

$$\begin{aligned}\sigma^2(\lambda) &= \lambda^2(\sigma_i^2 + \sigma_M^2 - 2\,\mathrm{cov}(r_i, r_M)) \\ &\qquad + 2\lambda(\mathrm{cov}(r_i, r_M) - \sigma_M^2) + \sigma_M^2\end{aligned}$$

Think of (2.13) as parametrizing a curve in (σ, r) space, just as (2.10) is a line in (σ, r) space.

Then its trivial to see that the line (2.10) is tangent to the curve (2.13) at the point $\lambda = 0$. This follows by evaluating (2.13) at $\lambda = 0$ to show that it is the market point $(\overline{r}_M, \sigma_M)$, and other points on the curve aren't efficient portfolios, so they lie below the line (2.10). The slope of this tangent line is, of course, (2.11), so we can equate this to the slope calculated by taking the derivative:

$$\frac{\overline{r}_M - r_f}{\sigma_M} = \left.\frac{d\overline{r}(\lambda)}{d\sigma(\lambda)}\right|_{\lambda=0} = \frac{\overline{r}_i - \overline{r}_M}{(\mathrm{cov}(r_i, r_M) - \sigma_M^2)/\sigma_M}$$

Solving this equation for the unknown variable $\overline{r}_i$ leads directly to (2.12), completing the proof. □

Theorem 2.3. Consider a portfolio of risky assets with holdings vector $h \in \mathbb{R}^n$ where $h_i \geq 0$ and $\sum_i h_i = 1$. Then one has

$$\overline{r}_h - r_f = \beta_h(\overline{r}_M - r_f) \quad \text{where} \quad \beta_h := \sum_i h_i\beta_i = \boldsymbol{h}'\boldsymbol{\beta} \tag{2.14}$$

Proof. Note that

$$\overline{r}_h - r_f = -r_f + \sum_i h_i\overline{r}_i = \sum_i h_i(\overline{r}_i - r_f) \tag{2.15}$$

$$= \sum_i h_i\beta_i(\overline{r}_M - r_f) = \beta_h(\overline{r}_M - r_f) \tag{2.16}$$

To go from the first line to the second line, we used (2.12). This concludes the proof of Theorem 2.3.

2.3. **Estimation.** Eq. (2.12) suggests that we could attempt to estimate β_i from a linear regression of $r_i - r_f$ on $r_m - r_f$. Indeed, Sharpe seems to allow for this:

> *The value of β may be given an interpretation similar to that found in regression analysis utilizing historic data, although in the context of the CAPM it is to be interpreted strictly as an ex ante value based on probabilistic beliefs about future outcomes.*
>
> — Sharpe (1990)

Of course, it may be in many cases that a historically estimated regression coefficient is a good *ex ante* forecast. This is likely to be true only in cases where the capital structure of the underlying company remains stationary. If the company has just undergone a major spinoff or restructuring, or a major debt issue, then the historical beta estimated via regression is likely less accurate as a forecast of ex ante beta. We now look at the empirical data for Apple over the three-year period from 2012-2014.

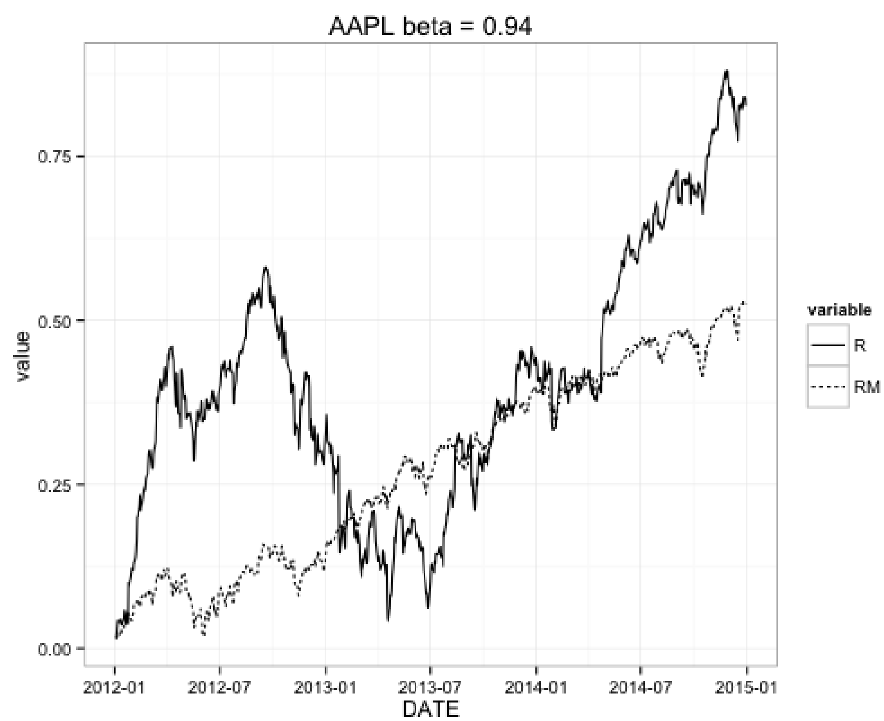

FIGURE 2.1. Cumulative returns to AAPL and the S&P 500 over a three-year period.

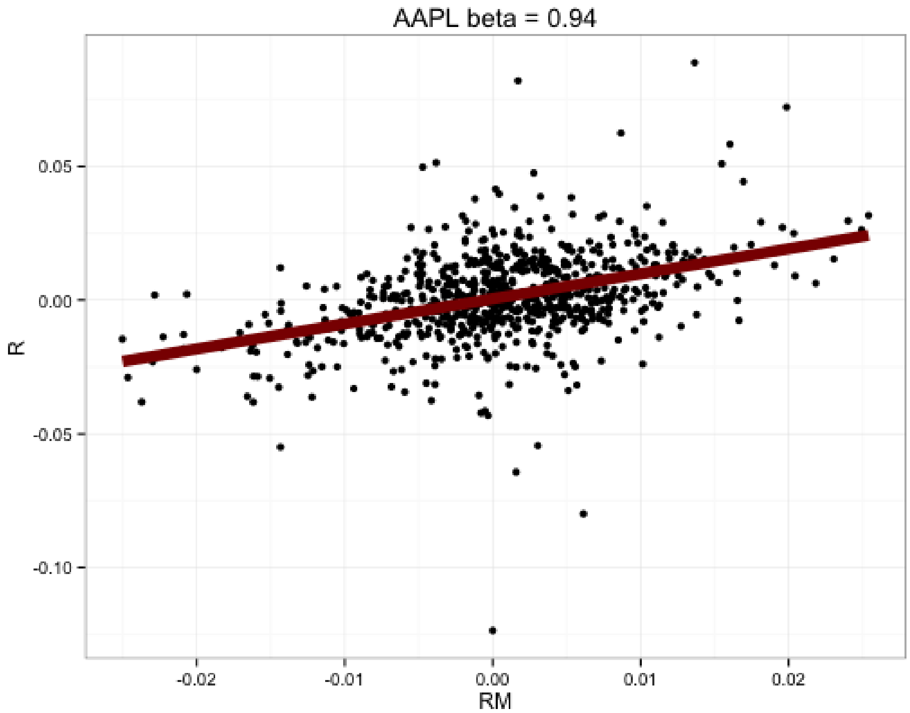

FIGURE 2.2. Scatterplot of returns to AAPL and the S&P 500 together with the line suggested by a CAPM regression.

REFERENCES

Chamberlain, Gary (1983). "A characterization of the distributions that imply meanVariance utility functions". In: *Journal of Economic Theory* 29.1, pp. 185–201.

Feldstein, Martin S (1969). "Mean-variance analysis in the theory of liquidity preference and portfolio selection". In: *The Review of Economic Studies* 36.1, pp. 5–12.

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

Sharpe, William F (1964). "Capital asset prices: A theory of market equilibrium under conditions of risk". In: *The Journal of Finance* 19.3, pp. 425–442.

– (1990). "Capital asset prices with and without negative holdings". In: *Nobel Lecture, December 7, 1990.*

Tobin, James (1958). “Liquidity preference as behavior towards risk”. In: *The review of economic studies* 25.2, pp. 65–86.