1. A Note on the Tangency Portfolio

Gordon Ritter

Dec 5, 2017

1.1. **Fully-Invested Portfolios with no riskless asset.** Consider first a universe of n risky assets, with no risk-free asset at all.

In the notation of Merton (1972), let E_i denote the expected return on the i-th security, and let Ω denote the covariance matrix of returns of a universe of n securities. The bold face $\mathbf{E} = (E_1, \ldots, E_n)$ denotes the vector of expected returns for all securities.

A portfolio is said to be a *frontier portfolio* if it has the smallest variance for a prescribed level E of expected return. Merton (1972) writes this as a constrained optimization problem:

$$\min \frac{1}{2}\sigma^2 \quad \text{subject to}$$
$$\sigma^2 = \langle \mathbf{x}, \Omega \mathbf{x} \rangle, \quad E = \langle \mathbf{x}, \mathbf{E} \rangle, \quad 1 = \mathbf{1} \cdot \mathbf{x}$$

One may then see by applying the KKT conditions that all solutions are of the form

$$\mathbf{x} = \gamma_1 \Omega^{-1} \mathbf{E} + \gamma_2 \Omega^{-1} \mathbf{1}$$

Indeed, defining

$$A = \langle \mathbf{1}, \Omega^{-1} \mathbf{E} \rangle, \quad B = \langle \mathbf{E}, \Omega^{-1} \mathbf{E} \rangle, \quad C = \langle \mathbf{1}, \Omega^{-1} \mathbf{1} \rangle$$

Merton shows that

$$\gamma_1 = \frac{CE - A}{D}, \quad \gamma_2 = \frac{B - AE}{D}$$

where $D := BC - A^2 > 0$, and it follows using non-singularity of Ω that $B > 0$ and $C > 0$.

One of the Lagrangian first-order conditions is

$$\Omega \mathbf{x} = \gamma_1 \mathbf{E} + \gamma_2 \mathbf{1}.$$

If we take the inner product of both sides of this equation on the left by $\mathbf{x}$, we obtain

$$\sigma^2 = \langle \mathbf{x}, \Omega \mathbf{x} \rangle = \gamma_1 E + \gamma_2$$

where we use the fully-invested constraint that $\langle \mathbf{x}, \mathbf{1} \rangle = 1$. Substituting the expressions above for γ_1 and γ_2, we find

$$\sigma^2 = \frac{C}{D} E^2 - \frac{2A}{D} E + \frac{B}{D}.$$

Thus the frontier in mean-variance space is a parabola. The unique minimum point (the minimum-variance portfolio) is located at $\overline{E} = A/C$ with variance $\overline{\sigma}^2 = 1/C$.

The weights of the minimum-variance portfolio are

$$\overline{\mathbf{x}} = \frac{1}{C}\Omega^{-1}\mathbf{1}.$$

Suppose one makes a change of coordinates, representing E as a function of standard deviation σ, instead of variance σ^2. Also, we switch the axes, representing E on the ordinate and σ on the abscissa. In this parameterization, the frontier is a hyperbola whose leftmost point $(\overline{\sigma}, \overline{E})$ corresponds to the minimum variance portfolio. If, bizarrely, you insisted upon even lower return than the minimum variance portfolio (while remaining fully invested and without being able to access a risk-free asset), then you would have to accept more risk than $\overline{\sigma}$. This contradicts rationality, so no-one would ever do it. Hence the lower half of the hyperbola is not relevant or important; only the upper half is important. The whole hyperbola is called the *frontier*; the upper half is called the *efficient frontier*.

This is nicely visualized by measuring the historical frontiers for stocks and bonds over a century, for various holding periods. In each case, we see that 100% bonds is in the bottom half of the hyperboloid – the silly region that one should never consider. It is also worth emphasizing that these are based on *historical* returns and variances, which should never be confused with predictive or *ex ante* versions of same.

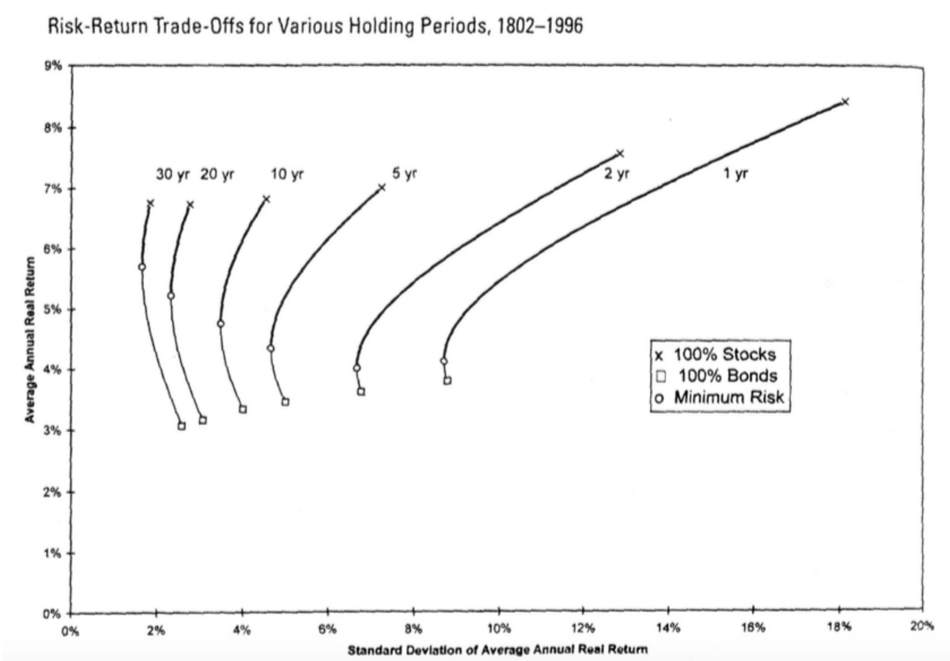

FIGURE 1.1.

In some sense, the slightly weird-looking shape of the hyperbolae in Fig. 1.1 is closely-related to the constraint $\sum_i x_i = 1$. This means, intuitively, that all of your money must be invested in one of the menu of risky assets; you cannot opt out, even with part of your account. This leads to various (perhaps counterintuitive) properties such as the fact that there is an absolute minimum variance, and you cannot achieve lower variance no matter what you do in this model.

The constraint $\sum_i x_i = 1$ is also responsible for the fact that if you insisted upon lower return than the minimum variance portfolio's return, you would have to accept more risk. After all, the capital has to go somewhere, and if it isn't going to the minimum variance weights, where's it going to go? You guessed it: something else, which must then display more than the minimum variance.

Clearly, different efficient portfolios have different Sharpe ratios. If this seems odd, it's also essentially due to the constraint $\sum_i x_i = 1$. That constraint basically means that in order to get risk that is very different from the minimum risk $\overline{\sigma}$ while still remaining fully invested, you might have to construct rather strange portfolios which have suboptimal risk-return tradeoffs.

The Sharpe ratio is

$$S = \frac{E - r_f}{\sigma}$$

where r_f is the risk-free rate. In a world with no riskless asset at all, I suppose $r_f = 0$, in which case the Sharpe ratio is simply E/σ. In the coordinates, (σ, E) the Sharpe ratio at a point on the efficient frontier is the slope of a line from the origin to that point. One can show with a little calculus (exercise) that the maximum such slope is realized when the line from the origin is tangent to the hyperbola.

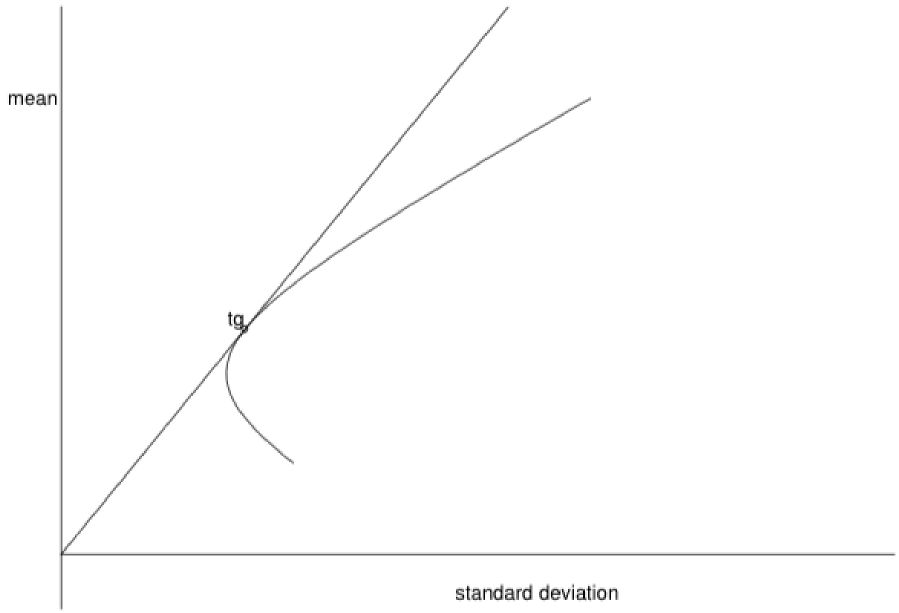

FIGURE 1.2. The maximum Sharpe ratio is when the line from the origin is tangent to the hyperbola.

I'd like to re-emphasize that if you apply the above using historical returns the tangency portfolio is the one that *was* the maximum Sharpe ratio *in the past*, and really all statements about return and risk should be suitably transformed into the past tense. The "minimum variance portfolio" is the one that was the minimum variance in the past, etc. This should not be confused with a portfolio that is forecasted to be the minimum variance portfolio in the future.

1.2. **Removing the fully-invested constraint.** Continue from the previous section considering a universe of n risky assets, with no risk-free asset at all. It makes sense to ask whether any of the foregoing technology could possibly apply to dollar-neutral, self-financing portfolios. Recall that a portfolio is said to be *self-financing* if $\sum_i x_i = 0$, which is achieved in equity markets when You borrow stock worth, say, 100 million dollars, sell it in the marketplace for precisely that amount, and then use the proceeds to purchase 100 million dollars' worth of other stocks. Such a portfolio is said to be "self-financing," but in fact, it requires stock borrow, which is a form of financing!

Without the constraint $\sum_i x_i = 1$, the original goal of finding the minimum risk for a desired level of expected return still makes sense:

$$\min \sigma^2(\mathbf{x}) \text{ s.t. } \mathbf{x} \cdot \mathbf{E} = E$$

Applying the method of Lagrange multipliers, this is equivalent to

$$\min \left\{ \frac{1}{2} \kappa \sigma^2(\mathbf{x}) - \mathbf{x} \cdot \mathbf{E} \right\} \tag{1.1}$$

The solution to this is $\kappa^{-1}\Omega^{-1}\mathbf{E}$, which has already been derived using the Arrow-Pratt theory of decision-making under uncertainty.

Indeed, as long as the asset returns follow an elliptical distribution, then for any concave increasing utility function, and for any initial wealth, there exists some $\kappa > 0$ such that the portfolio that optimizes expected utility of wealth is the same as the one optimizing (1.1).

Hence, as soon as we remove the constraint $\sum_i x_i = 1$, there is only one efficient portfolio of risky assets, up to scaling. The risk-aversion constant κ controls the size of the portfolio, trivially, because it scales with $1/\kappa$. All (unconstrained) efficient portfolios are proportional to $\Omega^{-1}\mathbf{E}$. An overall scaling does not change the Sharpe ratio, so all efficient portfolios have the same Sharpe ratio, and no other portfolio that can be constructed from the same set of assets has a higher Sharpe ratio.

If trading costs are included, then we are once again in a scenario where the Sharpe ratio depends on κ. As we have shown from utility theory, with trading costs, (1.1) becomes

$$\max \left\{ \mathbf{x} \cdot \mathbf{E} - \frac{1}{2} \kappa \sigma^2(\mathbf{x}) - \text{cost}(\mathbf{x}_0, \mathbf{x}) \right\} \tag{1.2}$$

where $\mathbf{x}_0$ is the initial portfolio that we are trading out of.

Now it is still true that κ controls the size of the portfolio, and it is still true that smaller κ means larger portfolios, but the relationship is no longer linear. Larger portfolios trade more in absolute dollar terms, and hence more of their pre-cost Sharpe ratio is eroded in the form of trading costs, with spread cost and market impact being some of the largest costs for large institutional investors.

Hence in the presence of trading costs, smaller κ translates into lower Sharpe ratio, and the maximum Sharpe ratio that could be achieved by any portfolio is the Sharpe ratio of the unconstrained Markowitz portfolio, but in reality that upper bound is never achieved. It should be noted that by "Markowitz portfolio" we are not implying that the mean and covariance estimates must be sample means or sample covariance. Rather, they should be your best and most reasonable *ex ante* (forward-looking) forecasts. The sample estimates for mean return and return covariance are typically not very good forecasts.

References

Merton, Robert C (1972). "An analytic derivation of the efficient portfolio frontier". In: *Journal of financial and quantitative analysis* 7.4, pp. 1851–1872.