Tobias Tip: CAPM

I strongly recommend you read Brian's review from last quarter on the CAPM (available on the class website).

The CAPM

The central idea behind the CAPM is that it is a statement about *expectations*; specifically, we aren't making a claim about what an assets return will be in any given period, but rather what the "long run" return should be on average.

Going through Brian's math gives us a lot of insight:

$$r_t = \beta_m \cdot r_{m,t} + \epsilon_t$$
$$\mathbb{E}[r_t] = \beta_m \cdot \mathbb{E}[r_{m,t}] + \mathbb{E}[\epsilon_t]$$
$$\mathbb{E}[r_t] = \beta_m \cdot \mathbb{E}[r_{m,t}] + 0$$
$$\therefore \mathbb{E}[r_t] = \beta_m \cdot \mathbb{E}[r_{m,t}]$$

The implication here is that in expectation, the return of an asset can be fully explained by two things: its beta to the market, and the expected return of the market itself. The natural way to test this is to run a linear regression of asset returns against market returns, but *with* an intercept term. The reason for this is that we want to check whether the intercept is statistically different from zero, which would indicate that there are other factors at play beyond just market beta.

The way that we can check this is using the standard t-test for the intercept term (in the case of a single asset regression). Recall from the Data Analysis and Regression reviews that the t-statistic for a coefficient is given by:

$$t = \frac{\hat{\alpha} - \alpha_0}{SE(\hat{\alpha})}$$

$$\text{SE}(\hat{\alpha}) = \hat{\sigma} \cdot \sqrt{\frac{1}{n} + (n-1) \cdot \sigma_i^2}$$

$$\hat{\sigma} = \sqrt{\frac{\sum \hat{\epsilon}_i^2}{(n-2)}}$$

So really what we care about is the standard error of the intercept term. Moreover, we often say that we don't care about R^2 in CAPM regressions, which is true, because:

$$R^2 = 1 - \frac{(n-1) \cdot \text{Var}(\epsilon_t)}{(n-1) \cdot \text{Var}(r_t)} = 1 - \frac{\text{Var}(\epsilon_t)}{\text{Var}(r_t)}$$

So really what R^2 is telling us is the proportion of *idiosyncratic risk* (i.e., the variance of the error term) to total risk. However, from the CAPM perspective, we are throwing expectations on the formula, and therefore idiosyncratic risk "vanishes" in expectation. It is worth noting that there is some sleight of hand going on, because while we don't care about idiosyncratic risk in expectation, the variance of the residuals (driver of the R^2) term does impact the standard error of the intercept term, and therefore our ability to statistically test the CAPM!

Diving a bit deeper into the variance decomposition, we can see that:

$$\text{Var}(r_t) = \beta_m^2 \cdot \text{Var}(r_{m,t}) + \text{Var}(\epsilon_t) = (\text{Systematic Risk}) + (\text{Idiosyncratic Risk})$$

This decomposition is important because it tells us that the total risk (variance) of an asset can be broken down into two components: the systematic risk (driven by market movements) and the idiosyncratic risk (specific to the asset). Thus, if our only driver of expected returns is market beta, and our total risk is made up of both systematic and idiosyncratic components, then it follows that we are not being compensated for taking on idiosyncratic risk.

One natural conclusion from this is that any portfolio that contains idiosyncratic risk is *not* an MV-optimal portfolio (under the CAPM), because if we can diversify away any idiosyncratic risk, that would strictly reduce our portfolio variance without impacting expected returns. How can we construct a portfolio with no idiosyncratic risk? By holding the market portfolio!

Another way to get at this idea is through the lens of the tangency portfolio. The CAPM implies that the market portfolio is the tangency portfolio. Remember from Class 1 that the tangency portfolio is the highest Sharpe Ratio portfolio. Therefore, if we do some algebra, using the fact that univariate regression β can be expressed as:

$$\beta_m = \frac{\text{Cov}(r_t, r_{m,t})}{\text{Var}(r_{m,t})}$$

We can use the expectation formula above to get things in terms of Sharpe Ratios and correlations:

$$\begin{aligned}
\mathbb{E}[r_t] &= \frac{\text{Cov}(r_t, r_{m,t})}{\text{Var}(r_{m,t})} \cdot \mathbb{E}[r_{m,t}] \\
&= \frac{\text{Corr}(r_t, r_{m,t}) \cdot \sigma_{r_t} \cdot \sigma_{r_{m,t}}}{\sigma^2_{r_{m,t}}} \cdot \mathbb{E}[r_{m,t}] \\
\frac{\mathbb{E}[r_t]}{\sigma_{r_t}} &= \text{Corr}(r_t, r_{m,t}) \cdot \frac{\mathbb{E}[r_{m,t}]}{\sigma_{r_{m,t}}} \\
\therefore \text{Sharpe}(r_t) &= \text{Corr}(r_t, r_{m,t}) \cdot \text{Sharpe}(r_{m,t})
\end{aligned}$$

So we can see that the Sharpe Ratio of any asset is equal to the correlation of an asset with the market, multiplied by the Sharpe Ratio of the market. Therefore, since correlation is bounded between -1 and 1, the maximum Sharpe Ratio that any asset can have is equal to the Sharpe Ratio of the market (when correlation = 1). This implies that the market portfolio is indeed the tangency portfolio.

To conclude, these are the key takeaways from the CAPM:

- The *expected* return of an asset is purely a function of its market beta.
- Idiosyncratic risk is not compensated in expectation.
- The market portfolio is the tangency portfolio (i.e., highest Sharpe Ratio portfolio).
- We can test the CAPM by regressing asset returns against market returns with an intercept term, and checking if the intercept is statistically different from zero.

This is all well and good for a single asset. However, there are thousands of assets in the market, so sure, the CAPM might hold or fail for a single asset, but what about across the entire market? This is where the idea of a cross-sectional test of the CAPM comes in.

We collect a large number of assets, and for each asset, we estimate its market beta by regressing its returns against market returns (with an intercept). Then, we take the estimated betas and run a cross-sectional regression of average asset returns against their estimated betas. The CAPM predicts that the slope of this regression should be equal to the expected market risk premium.

To get some intuition on this, we have the following vectors:

$$\begin{bmatrix} \mathbb{E}[r_1] \\ \mathbb{E}[r_2] \\ \vdots \\ \mathbb{E}[r_N] \end{bmatrix} \begin{bmatrix} \beta_{m,1} \\ \beta_{m,2} \\ \vdots \\ \beta_{m,N} \end{bmatrix}$$

Where each of them is an individual asset's expected return and market beta. The cross-sectional test therefore lets us extract two different things:

- What the expected return of the market is, i.e., we are best-solving for $\mathbb{E}[r_{m,t}]$ – since each data point gives us an equation of the form $\mathbb{E}[r_i] = \beta_{m,i} \cdot \mathbb{E}[r_{m,t}]$.
- Whether the relationship between expected returns and market betas holds across the entire market. That is, does our best-solved $\mathbb{E}[r_{m,t}]$ cause all of the data points to line up, implying that $\mathbb{E}[r_i]$ is indeed proportional to $\beta_{m,i}$ for all assets i?

So, let's run the following regression:

$$\begin{bmatrix} \mathbb{E}[r_1] \\ \mathbb{E}[r_2] \\ \vdots \\ \mathbb{E}[r_N] \end{bmatrix} = \kappa + \gamma \begin{bmatrix} \beta_{m,1} \\ \beta_{m,2} \\ \vdots \\ \beta_{m,N} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}$$

So when we do this, we get a γ estimate. This γ is exactly the number that the CAPM predicts should be equal to the expected market risk premium, i.e., $\gamma = \mathbb{E}[r_{m,t}]$.

In this case, we very much care about the R^2 of this regression, and in fact if the CAPM holds, we should see $R^2 = 1$, because all of the variation in expected returns should be explained by market betas alone. That is, we are *jointly* testing all of our mini-CAPM (time series) regressions at once. We slightly care about $\kappa = 0$, however, fitting κ allows us to capture any misestimation of expected market returns. The capturing of misestimation can be seen from the Data Analysis and Regression review, where we saw that the intercept term in linear regression effectively demeans the data for us, allowing us to focus on the slope term. So the inclusion of κ gives us the flexibility to not worry about the mean but purely focus on the relationship between expected returns and market betas.

Time Diversification

By far the most important formula that you need to remember (from last note on coherent risk measures) is the following:

$$\mathrm{Var}(A + B) = \mathrm{Var}(A) + \mathrm{Var}(B) + 2 \cdot \mathrm{Cov}(A, B)$$

Or, in standard deviation terms:

$$\sigma_{A+B} = \sqrt{\sigma_A^2 + \sigma_B^2 + 2 \cdot \rho_{A,B} \cdot \sigma_A \cdot \sigma_B}$$

With the critical insight that volatility is subadditive:

$$\sigma_{A+B} \leq \sigma_A + \sigma_B$$

This is the crux of time diversification. Specifically, instead of considering the variance of two assets, we are considering the variance of two time periods, T_1 and T_2. So now we consider the effect of holding an asset over two time periods instead of one. Using the formula above, we have:

$$\sigma_{T_1+T_2} = \sqrt{\sigma_{T_1}^2 + \sigma_{T_2}^2 + 2 \cdot \rho_{T_1,T_2} \cdot \sigma_{T_1} \cdot \sigma_{T_2}}$$

So, let's consider a few special cases and assume that the variance is the same across both time periods, i.e., $\sigma_{T_1} = \sigma_{T_2}$:

- If returns are independent across time, then $\rho_{T_1,T_2} = 0$, and we have:

  $$\sigma_{T_1+T_2} = \sqrt{\sigma_{T_1}^2 + \sigma_{T_2}^2} = \sqrt{2} \cdot \sigma_{T_1}$$

  Note that this is exactly the volatility scaling rule we saw in Class 1!
- If returns are perfectly positively correlated across time, then $\rho_{T_1,T_2} = 1$, and we have:

  $$\sigma_{T_1+T_2} = \sqrt{\sigma_{T_1}^2 + \sigma_{T_2}^2 + 2 \cdot \sigma_{T_1} \cdot \sigma_{T_2}} = 4 \cdot \sigma_{T_1}$$

  In this case, volatility scales linearly with time.
- If returns are perfectly negatively correlated across time, then $\rho_{T_1,T_2} = -1$, and we have:

  $$\sigma_{T_1+T_2} = \sqrt{\sigma_{T_1}^2 + \sigma_{T_2}^2 - 2 \cdot \sigma_{T_1} \cdot \sigma_{T_2}} = 0$$

  And thus our returns are riskless.

A critical insight therefore, is that when considering time diversification, the *only* thing that matters is that our returns are not perfectly autocorrelated across time. If returns are independent, or even positively correlated, then $\sigma_{T_1+T_2} < \sigma_{T_1} + \sigma_{T_2}$, and thus volatility grows sublinearly with time. Of course, if returns are negatively correlated across time, then volatility can actually decrease with time.

> *The only thing that needs to hold for time diversification to work is that $\rho_{T_1,T_2} < 1$. If this holds, then volatility grows sublinearly with time.*

The final piece of the puzzle is that our expected return scales linearly with time in all cases, because of the linearity of expectation:

$$\mathbb{E}[T_1 + T_2] = \mathbb{E}[T_1] + \mathbb{E}[T_2]$$

Therefore, as we hold an asset for longer periods of time, our expected return grows linearly, while our volatility grows sublinearly (assuming $\rho_{T_1,T_2} < 1$).

Now that we know roughly how to scale volatility with time, we can consider the probability of a negative cumulative return over multiple time periods. Assuming log-returns are iid normal, it then follows that our total return volatility over T periods is given by:

$$\sigma_T = \sigma_{1 \text{ period}} \cdot \sqrt{T}$$

And our expected return over T periods is given by:

$$\mu_T = \mu_{1 \text{ period}} \cdot T$$

So, the probability of a negative return over T periods is given by:

$$\begin{aligned} P(r_T < 0) &= P\left(Z < \frac{0 - \mu_T}{\sigma_T}\right) \\ &= P\left(Z < \frac{-\mu_{1 \text{ period}} \cdot T}{\sigma_{1 \text{ period}} \cdot \sqrt{T}}\right) \\ &= P\left(Z < -\frac{\mu_{1 \text{ period}}}{\sigma_{1 \text{ period}}} \cdot \sqrt{T}\right) \\ &= \Phi\left(\frac{-\mu_{1 \text{ period}}}{\sigma_{1 \text{ period}}} \cdot \sqrt{T}\right) \end{aligned}$$

Where $\Phi(\cdot)$ is the CDF of the standard normal distribution. Therefore, as T increases, the argument to $\Phi(\cdot)$ becomes more negative (assuming $\mu_{1\text{ period}} > 0$), and thus the probability of a negative return decreases with time. That is, the linearity of μ_T combined with the sublinearity of σ_T causes the probability of a negative return to decrease with time.

It's also worth noting that this is really saying something about sample size, not just time. Specifically, as we increase the number of independent return observations (i.e., increase our sample size), our estimate of the mean return becomes more precise, and thus the probability of observing a negative cumulative return decreases. This is a driving factor in HFT strategies having Sharpe Ratios in excess of 30, because they have so many return observations made across hundreds of thousands of trades per day.

As an extension, let's consider the probability that our T-period return is negative with some autocorrelation ρ between periods. In this case, our T-period volatility is given by:

$$\sigma_T = \sigma_{1\text{ period}} \cdot \sqrt{T + 2 \cdot \sum_{i=1}^{T-1} (T - i)\, \rho^i}$$

It's pretty easy to see here that when ρ is negative, the sum term will be negative, and thus volatility will grow more slowly with time. Conversely, when ρ is positive, the sum term will be positive, and thus volatility will grow more quickly with time. We can now just plug in this new σ_T into our probability of negative return formula from above to get the desired result. It's also worth noting that the scaling looks kind of like a geometric series! So for extremely large T, ρ^i will go to zero, and thus the scaling will approach the independent returns case.

Readings

A somewhat interesting article from The Economist on applying a CAPM-like framework to bonds:

Factor-based investing spreads from stocks to bonds

A data boom enables an algorithmic approach to fixed-interest investing

COMPARED with equity investing, bond investing can seem stuck in the dark ages. As hedge funds and asset managers use whizzy algorithms to trade shares automatically, bond-fund managers still often call traders by phone. So when new investing strategies do arise, they make an even bigger splash. "Factor" investing is the latest example.

This is the idea, credited to economists Eugene Fama and Kenneth French, that predictable, persistent factors explain long-term asset returns. Their 1992 model for equities used the size of firms and what became known as "value" (the tendency for cheap assets to outperform pricey ones). Later models added factors such as "momentum" (the tendency of prices to keep moving in the same direction). Factor-based analysis has squeezed active managers (since it explains much of their returns) and helped drive the rise of passive investing. Investors can access factors in equities, often called "smart beta", through cheap index-tracking funds or exchange-traded funds (ETFs) from the likes of BlackRock and State Street Global Advisors.

Messrs Fama and French considered factors in bond returns as early as 1993, though not the same ones as for equities (they reckoned, for instance, that for bonds value had "no obvious meaning"). Federal requirements since 2002 to disclose transaction prices and volumes have enabled closer analyses. A recent paper by researchers at AQR Capital Management, a $226bn hedge fund founded by a student of Mr Fama that specialises in factor investing for equities, looks at four factors for global sovereign bonds and American corporate ones: carry (high-yielding bonds beat low-yielding ones), quality (safer assets have better risk-adjusted returns), value and momentum.

These not only would have provided consistently good returns over the past two decades, but were also largely uncorrelated with factors in equity markets, credit risk for bonds and macroeconomic variables such as inflation. Since active bond-fund managers tend to make excess returns mainly by buying riskier bonds, and a traditional bond index-tracking fund means exposure to the firms and countries that issue the most debt, factors provide a third, distinctive investment option.

AQR's first dedicated fixed-income offering, a fund of American high-yield (that is, junk-rated) bonds, was launched in mid-2016. It outperformed the benchmark index by 2.1 percentage points in its first year, and 2.6 points in its second. Tony Gould of AQR credits not only the factor modelling for its success. He says that the higher cost of trading bonds compared with equities needs to be built into the bond-picking process. The firm has since started two more bond funds. Other such firms that used to focus on equities are looking into bonds, too. Man Numeric, for instance, a quant fund in Boston, wants to apply its expertise in company-level analysis to high-yield bonds.

Among the mass-market offerings are BlackRock's first smart-beta bond fund, launched in 2015. It switched from active management to index-tracking in 2018, and the firm now has several index-tracking bond ETFs that use factors (mostly quality and value). Fidelity Investments launched two bond factor ETFs in March, and Invesco launched eight on July 25th.

Factor investing for bonds is still so new that many investors have not even heard of it. But opportunities to use it are growing because of recent European regulations mandating price and volume disclosure for bonds. Just five years ago a fund manager would have struggled to find enough data for non-American bonds, says Collin Crownover of State Street Global Advisors. Now the firm is applying quality- and value-factor analysis to corporate bonds in euros and sterling. The way index-trackers and smart-beta approaches laid waste to stock-pickers suggests that managers of active bond funds should be quaking. ■