1. Today’s Agenda

1. Announcement:
   – HW 2 posted–Due in two weeks.

2. Conditional Expectation and Linear Regression Analysis (OLS–Ordinary Least Squares)
   (a) Consistency–importance
   (b) Asymptotic Normality–importance
   (c) Identification–importance
   (d) Two applications–AR(1) and predictive regressions

3. Asset Pricing: CAPM

4. Better Asset Pricing: APT

5. Explanation of Reading Material and HW 2
• Recall from last time, that we can write any variable $y$ as $y = E(y|x) + \{y - E(y|x)\}$

• We denote $\varepsilon = \{y - E(y|x)\}$.

• Note that $E(\varepsilon|x) = 0$ by construction.

• Recall that we were looking at $E(y|x)$ as a function of $x$.

• Now, we assume that $E(y|x)$ is an affine function of $x$, or $E(y|x) = \alpha + \beta x$.

• The problem is that we don’t know the parameters $\alpha$ and $\beta$.

• But...we can estimate them...and this is the goal of a linear regression analysis.
• Linear Regression analysis, also known as Ordinary Least Squares (or OLS) is by far the most popular model of a conditional mean

• This popularity is due to three factors:
  – The analysis and estimation of a linear function is far, far easier than the estimation of nonlinear functions...but you will find that out for yourselves.
  – Any function \( g(x) \) can be approximated (Taylor expansion) with a line around the points of interest.
  – Most finance (and economic) models are linear.
  – Aside: Benefits of non-linearities and complexity
    * Cost-Benefit Analysis: What are the benefits from modelling higher order terms? If the benefits outweigh the cost of dealing with complex systems, then it is worth doing.
• BIGGER QUESTION: Why do we estimate the relationship between $y_t$ and $x_t$?
• We want to know what is the effect of $x_t$ on $y_t$.
• IMPORTANT: If you have two variables $y_t$ and $x_t$, can we always estimate the linear relationship between them?
• I.e. can we always estimate $\beta$ accurately?
• ANSWER: No. We need the condition $E(\varepsilon_t|x_t) = 0$, or that $\varepsilon_t$ is truly the residual in the regression and that it is not correlated with the explanatory variables $x_t$.
• The variable $x_t$ is called exogenous. It is not related to the shocks $\varepsilon_t$ that affect $y_t$. If it is related to the shocks, we will not be able to distinguish $x_t$ from the shocks, and we will not be able to estimate $\beta$ consistently.
  – If $E(\varepsilon_t|x_t) = 0$ is not satisfied, our estimator of $\beta$ will not be consistent. In other words, it will not be close to $\beta$.
• This is the cardinal sin in empirical work!!!
2 Time Series Linear Regression Models

\[ y_t = x_t' \beta + \varepsilon_t \]

• Given the observations \((y_t, x_t)\), the ordinary least squares estimate of \(\beta\), denoted by \(\hat{\beta}^{ols}\), or simply \(\hat{\beta}\), is the value of \(\beta\) that minimizes the residual sum of squares, or

\[ \hat{\beta}^{ols} = \arg \min_{\beta} \sum_{t=1}^{T} (y_t - x'_t \beta)^2 \]

• The solution is

\[ \hat{\beta} = \left( \sum_{t=1}^{T} x_t x'_t \right)^{-1} \left( \sum_{t=1}^{T} x_t y_t \right) \]

• Note that

\[ \hat{\beta} = \left( \sum_{t=1}^{T} x'_t x_t \right)^{-1} \left( \sum_{t=1}^{T} x'_t y_t \right) \]
\[ = \left( \sum_{t=1}^{T} x'_t x_t \right)^{-1} \left( \sum_{t=1}^{T} x_t \left( x_t \beta + \varepsilon_t \right) \right) \]
\[ = \left( \sum_{t=1}^{T} x'_t x_t \right)^{-1} \left( \sum_{t=1}^{T} x'_t x_t \right) \beta + \left( \sum_{t=1}^{T} x'_t x_t \right)^{-1} \sum_{t=1}^{T} x'_t \varepsilon_t \]
\[ = \beta + \left( \sum_{t=1}^{T} x_t x'_t \right)^{-1} \sum_{t=1}^{T} x_t \varepsilon_t \]
• Note: The solution $\hat{\beta}$ to the minimization problem is very simple.

• But, we can also solve the problem by brute force, by numerically minimizing $\sum_{i=1}^{T} (y_i - x_i'\beta)^2$ w.r.t. $\beta$.

• We should obtain the same solution as $\hat{\beta}$, if our numerical minimization routine is good.

• This is the goal of simpleregression2.m
• Note:
\[ \hat{\beta} - \beta = \left( \sum_{t=1}^{T} x_t x_t' \right)^{-1} \sum_{t=1}^{T} x_t \varepsilon_t \]

• If \( \hat{\beta} \) is a good estimate of \( \beta \), it better be the case that \( \left( \sum_{t=1}^{T} x_t x_t' \right)^{-1} \sum_{t=1}^{T} x_t' \varepsilon_t \) is small.

• In fact, we can argue that this piece is “small” and “decreases” as \( T \to \infty \). For that, we need that \( E(x_t \varepsilon_t) = 0 \). Then, we can argue that \( \text{plim}_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} x_t \varepsilon_t = E(x_t \varepsilon_t) = 0 \) and that \( \frac{1}{T} \sum_{t=1}^{T} x_t x_t' \) converges to a finite number.
• Note that $\hat{\beta}$ is a random variable. For different samples $(y_t, x_t)$, we will get different estimates $\hat{\beta}$ of $\beta$.
• Therefore, as any random variable, $\hat{\beta}$ has a density.
• What is the density of $\hat{\beta}$?
• Q: Why do we want to know that?
• A: Because we want to know whether $\beta$ is “close” to some hypothesized number $\beta_0$ i.e. we will want to test the null hypothesis: $\beta = \beta_0$.
• We can test the hypothesis by seeing how far away $\hat{\beta}$ is from $\beta_0$.
• Q: What do we mean by “far away”?
• A: To answer this question, we need to know the density of $\hat{\beta}$. 
Illustration: Suppose in our simulated regression (simpleregression.m), we want to test whether $\beta = 3.1$ and we have $T = 100$.

Suppose $T = 1000$ and we want to test the same hypothesis. How about for $T = 5000$?

What do we conclude?

IMPORTANT: We were able to do this testing because we were able to simulate the distribution of $\hat{\beta}$.

But when we work with data, we don’t know what is the true data generating process (DGP).

We need to derive some theoretical results that tell us what is the distribution of $\hat{\beta}$ without relying on simulations.
Before we go any further, we need to (re?) introduce the CLT (Central Limit Theorem).

- CLT: Suppose that \( \{x_t\}_{t=1}^{T} \) is a random sample from a distribution with mean \( \mu \) and variance \( \sigma^2 < \infty \). The sample mean, denoted by \( \bar{X} = (1/T) \sum_{t=1}^{T} x_t \), scaled by \( \sqrt{T} \) has an asymptotically normal distribution, or

\[
\sqrt{T} (\bar{X} - \mu) \rightarrow^d N \left( 0, \sigma^2 \right)
\]

- Interpretation 1: As we have more and more observations (information), the rescaled mean \( \sqrt{T} \bar{X} \) converges to a normal distribution with mean \( \mu \) and variance \( \sigma^2 \).

- Interpretation 2: As we have more and more observations (information), the mean \( \bar{X} \) converges to a normal distribution with mean \( \mu \) and variance \( \frac{\sigma^2}{T} \). This is often written as:

\[
\bar{X} \sim^a N \left( \mu, \frac{\sigma^2}{T} \right)
\]
• Note: The variance of $X$ is $\sigma^2$, but the variance of $\bar{X}$ is $\sigma^2/T$.

• Note: As $T \to \infty$, the variance of $\bar{X}$ decreases, i.e. we gain certainty about the exact location of the population mean $\mu$. 
• Now, let’s go back to the regression analysis, where
\[ \hat{\beta} - \beta = \left( \sum_{t=1}^{T} x'_t x_t \right)^{-1} \sum_{t=1}^{T} x_t \varepsilon_t \]

• First, assuming that \( x_t \) is ergodic for the second moment, then \( \frac{1}{T} \sum_{t=1}^{T} x'_t x_t \to \Sigma_x \), where \( \Sigma_x \) is ??????????

• Second, assume that the product \( x_t \varepsilon_t \) is a random variable, so that (from the CLT)
\[
\left( \frac{1}{T} \sum_{t=1}^{T} x_t \varepsilon_t \right) \sim^a N \left( E(x_t \varepsilon_t), \frac{E(x_t \varepsilon_t \varepsilon_t x'_t)}{T} \right)
\]
Here comes a crucial step in regression analysis, a step that we will use over and over again in different “flavors”

\[ E(x_t \varepsilon_t) = 0 \]

The above condition comes from the FOC of the minimization problem.

This condition implies that the explanatory variables \( x_t \) and the errors \( \varepsilon_t \) are uncorrelated.

Recall that our model was: \( y_t = \alpha + x_t' \beta + \varepsilon_t \). We want \( E(y|x) = \alpha + x_t' \beta \), so that \( E(\varepsilon_t|x_t) = 0 \).

But we also want \( E(\varepsilon_t) = 0 \). What does this imply for the relationship between \( \varepsilon_t \) and \( x_t \)?

You will have to show for homework that if \( E(\varepsilon_t|x_t) = 0 \), then \( E(\varepsilon_t x_t) = 0 \).

Therefore, the \( E(\varepsilon_t x_t) = 0 \) condition is really an implication of our model.

As mentioned above \( x_t \) is called exogenous.
Given that

\[ E(x_t \varepsilon_t) = 0 \]

we have

\[(\frac{1}{T} \sum_{t=1}^{T} x_t \varepsilon_t) \sim^a N \left( 0, \frac{E(x_t \varepsilon_t \varepsilon_t x'_t)}{T} \right)\]
Now comes another crucial assumption.

- We assume that $E(x_t \varepsilon_t \varepsilon_t' x_t') = E(x_t x_t') E(\varepsilon_t \varepsilon_t) = \sum_x \sigma^2_{\varepsilon}$, where $\sum_x$ is the covariance matrix of $x_t$ and $\sigma^2_{\varepsilon}$ is the variance of the residuals.

- This assumption will be explained later....but it has a lot of sub-assumptions. For example
  - The $\varepsilon_t$ are serially uncorrelated
  - The $\varepsilon_t$ are stationary
  - The variance of $\varepsilon_t$ does not depend on $x_t$.

- The assumption implies that:
  $$\left(\frac{1}{T} \sum_{t=1}^{T} x_t \varepsilon_t \right) \sim^a N \left(0, \frac{\sum_x \sigma^2_{\varepsilon}}{T}\right)$$
Now we put all the pieces together:

$$\hat{\beta} - \beta = \left( \sum_{t=1}^{T} x_t' x_t \right)^{-1} \sum_{t=1}^{T} x_t \varepsilon_t$$

$$= \frac{1}{T} \sum_{t=1}^{T} x_t \varepsilon_t \rightarrow^{p} 0$$

• Or, $\text{plim} \hat{\beta} = \beta$. Intuitively, as $T$ increases, the estimate $\hat{\beta}$ gets closer and closer to $\beta$.

• To emphasize once again, if $E(\varepsilon_t x_t) \neq 0$, then

$$\frac{1}{T} \sum_{t=1}^{T} x_t \varepsilon_t \rightarrow^{p} E(\varepsilon_t x_t) \neq 0,$$

and

$$\hat{\beta} - \beta = \frac{1}{T} \sum_{t=1}^{T} x_t \varepsilon_t \rightarrow^{p} \frac{E(\varepsilon_t x_t)}{\sum x} \neq 0$$

• No matter how much data we have, $\hat{\beta} = \beta + R.V.$

• We don’t have a good estimate of $\beta$. 
More interestingly,

\[
\begin{align*}
(\hat{\beta} - \beta) &= \frac{1}{T} \sum_{t=1}^{T} x_t \varepsilon_t \\
&\quad \div \frac{1}{T} \sum_{t=1}^{T} x_t' x_t \\
&\sim a \frac{N \left( 0, \frac{\Sigma_x \sigma_{\varepsilon}^2}{T} \right)}{\sqrt{\Sigma_x}} \\
&\sim a \frac{N \left( 0, \frac{\sigma_{\varepsilon}^2 \Sigma_x^{-1}}{T} \right)}{\sqrt{\Sigma_x}}
\end{align*}
\]

Or, we can also write

\[
\sqrt{T} \left( \hat{\beta} - \beta \right) \rightarrow^d N \left( 0, \sigma_{\varepsilon}^2 \Sigma_x^{-1} \right)
\]

- In words, the estimate $\hat{\beta}$ has an approximately normal distribution, centered at $\beta$, and with variance $\frac{\sigma_{\varepsilon}^2 \Sigma_x^{-1}}{T}$.
- We can estimate $\sigma_{\varepsilon}^2$ and $\Sigma_x^{-1}$ using their sample moments (and invoking ergoticity).
- Hence, we don’t have to rely on simulations to know the distribution of $\hat{\beta}$. 
We have made the strong assumption that 
\[ E(x_t \varepsilon_t \varepsilon_t x'_t) = E(x_t x'_t) E(\varepsilon_t \varepsilon_t). \] Suppose that this is not the case.

- Instead 
\[ E(x_t \varepsilon_t \varepsilon_t x'_t) = \sum_{x \varepsilon}. \]

- Then, (we still have that 
\[ E(x_t \varepsilon_t) = 0 \])
\[ \left( \frac{1}{T} \sum_{t=1}^{T} x_t \varepsilon_t \right) \sim^a N \left( 0, \frac{\Sigma_{x \varepsilon}}{T} \right) \]

- Therefore,
\[ \left( \hat{\beta} - \beta \right) = \frac{\frac{1}{T} \sum_{t=1}^{T} x_t \varepsilon_t}{\frac{1}{T} \sum_{t=1}^{T} x'_t x_t} \]
\[ \sim^a N \left( 0, \frac{\Sigma_{x \varepsilon}}{\Sigma_x} \right) \]
\[ \sim^a N \left( 0, \frac{\Sigma_x^{-1} \Sigma_{x \varepsilon} \Sigma_x^{-1}}{T} \right) \]

- Now, the question is: How do we estimate \( \Sigma_{x \varepsilon} \). This will wait.

- The point is that all the results still go through, with the exception that the standard error of \( \hat{\beta} \) is different.

- But the estimator is consistent. (It will not be if 
\[ E(\varepsilon x) \neq 0. \)
Since we have gotten too deep into econometrics and we really want to go to applications, here is a result that we will not prove.

- You probably know that if we have a random sample \( \{ x_t \}_{t=1}^T \), then \( \frac{\overline{X} - \mu}{s/\sqrt{T}} \sim^a N(0, 1) \) (From CLT).

- Similarly, in a regression, \( \frac{\hat{\beta} - \beta}{s_b/\sqrt{T}} \sim^a N(0, 1) \), where \( s_b = \text{diagonal element of the matrix} \sigma_x^2 \Sigma_x^{-1} \).

- This statistic is called a \( t \) statistic, because when the sample size is small (say 20 observations), the distribution of \( \frac{\hat{\beta} - \beta}{s_b/\sqrt{T}} \) is better approximated by a \( t \)-distribution. When the sample size increases, we use the normal distribution.

- Usually, we want to know whether \( \beta = 0 \). This is called a null hypothesis. Under this null hypothesis, the statistic becomes

\[
 t = \frac{\hat{\beta} - 0}{s_b/\sqrt{T}}
\]

- This test is fundamental in analyzing regressions.
We might be interested in knowing how well the data fits our model. But in order to answer this question, we first have to define a metric that measures the fit of the model.

- Recall that $\frac{1}{T} \sum_{t=1}^{T} y_t^2$ is the total variance of the data.

- But, we can write

$$\frac{1}{T} \sum_{t=1}^{T} y_t^2 = \frac{1}{T} \sum_{t=1}^{T} \left( x_t \hat{\beta} + \hat{\epsilon}_t \right)^2$$

$$= \frac{1}{T} \sum_{t=1}^{T} x_t^2 \hat{\beta}^2 + \hat{\epsilon}_t^2 + 2x_t \hat{\beta} \hat{\epsilon}_t$$

$$\approx \frac{1}{T} \sum_{t=1}^{T} x_t^2 \hat{\beta}^2 + \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t^2$$

- Therefore, dividing by $\frac{1}{T} \sum_{t=1}^{T} y_t^2$, we obtain

$$1 = \frac{\frac{1}{T} \sum_{t=1}^{T} x_t^2 \hat{\beta}^2}{\frac{1}{T} \sum_{t=1}^{T} y_t^2} + \frac{\frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t^2}{\frac{1}{T} \sum_{t=1}^{T} y_t^2}$$

- The first term is the variance of $y_t$ explained by the conditional mean. This term is denoted by $R^2$. Properties of $R^2$???

- The second term is the fraction of the variance of $y_t$ that is still unexplained.
Let’s slow down and think of what we have just done:

- We have specified the conditional mean as a linear function that has parameters to be estimated.
- We have estimated those parameters in such a way that the sum of the squared residuals is minimized.
- In this fashion, we have obtained an estimate of the parameters, called $\hat{\beta}$. This estimate has some nice properties:
  - As we increase $T$, the estimate gets closer and closer to the true value $\beta$.
  - We know that $\hat{\beta}$ has a normal distribution with mean $\beta$ and a variance that we can estimate.
  - Therefore, we can construct a statistical test whether the true $\beta$ is zero (or any other value).
2.1 Two important applications of time series regressions

- AR(1) [AR(p) really] model
  \[ Y_t = c + \phi Y_{t-1} + \varepsilon_t \]

- Predictive regressions
  \[ Y_t = \mu + \beta x_{t-1} + \varepsilon_t \]
  where it is often the case that
  \[ x_t = c + \phi x_{t-1} + u_t \]
• AR(1): Is it legitimate to run the regression
  \[ Y_t = c + \phi Y_{t-1} + \varepsilon_t \]

• When I say “legitimate,” I mean would we get consistent estimates of the parameters of interest.

• From the above discussion we need that \( E(Y_{t-1}\varepsilon_t) = 0 \).

• Note that we can write
  \[
  Y_{t-1} = c + \phi Y_{t-2} + \varepsilon_{t-1} \\
  = c + \phi [c + \phi Y_{t-3} + \varepsilon_{t-2}] + \varepsilon_{t-1} \\
  = c + \phi c + \phi^2 Y_{t-3} + \varepsilon_{t-1} + \phi \varepsilon_{t-2} \\
  = \frac{c}{1 - \phi} + \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-1-i}
  \]

• Hence, we need
  \[ E \left( \left[ \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-1-i} \right] \varepsilon_t \right) = 0 \]

• In other words, we need the \( \varepsilon_t \)'s to be serially uncorrelated.

• If we have serially correlated \( \varepsilon_t \)'s, we have to take that correlation into account.

• How?: Add lags until the residuals are uncorrelated.
Predictive Regressions: Let’s look at the bivariate forecasting relation:

\[ y_t = x_{t-1} \beta + \varepsilon_t \]

- Note, \( x_{t-1} \) forecasts (lags) \( y_t \).
- This is not a contemporaneous relation.
- We need that \( E(x_{t-1} \varepsilon_t) = 0 \).
- Suppose that

\[ x_t = \phi x_{t-1} + u_t \]

- In other words, \( x_t \) is an autoregressive process.
- We will need that \( E(\varepsilon_t u_{t-j}) = 0 \) all \( j \)'s.
- If this is not the case, we will have problems!
- The problems will be more severe for more persistent \( x_t \).
- Aside: How can we define persistence?
- All this will become intuitive.
Digression: Robust (Quantile) Regression

- So far, we have focused on the conditional mean \( E(y_t|x_t) \) and assuming that \( E(y_t|x_t) = x_t^\prime \beta \), we decided to minimize the sum of the squared residuals, to obtain

\[
\hat{\beta}^{ols} = \arg \min_{\beta} \sum_{t=1}^{T} (y_t - x_t^\prime \beta)^2
\]

- But suppose that the conditional distribution of \( y_t \) given \( x_t \) is ill behaved (outliers, skewness, kurtosis, etc.)

- We want to estimate the entire distribution of \( y_t \) given \( x_t \).

- We will see later how to do that...But here we can do something else.

- Let the unconditional distribution of \( y_t \) be \( F_{y_t}(y) = \Pr(y_t \leq y) \). Then, for any quantile \( \tau \), \( 0 < \tau < 1 \), we can define the inverse of \( F_{y_t}(\cdot) \) as \( Q_{y_t}(\tau) = \inf \{ r : F_{y_t}(y) \geq \tau \} \). The function \( Q_{y_t}(\cdot) \) is called the unconditional quantile function of \( y \). \( Q_y(0.5) \) is the 50th quantile, or the median, of \( y_t \).
The introduction of conditional quantiles is easily understood by making an analogy to the familiar least squares estimation.

The conditional quantile function $Q_{y|x}(\tau|X = x) = x_t' \beta(\tau)$ can be estimated by solving:

$$\hat{\beta}(\tau) = \arg\min_{\beta} \sum_t \rho_\tau(y_t - x_t' \beta)$$

where $\rho_\tau(.)$ is a piecewise linear “check function,” defined as $\rho_\tau(u) = u(\tau - I(u < 0))$ and $I(.)$ is the indicator function.
• The function $\rho_\tau(.)$ selects the quantile $\tau$ to be estimated (see Koenker and Hallock (2000)).

• For the case $\tau = 0.5$, $\rho_{0.5}(u) = |u|$ and the solution of the above problem, $\hat{\eta}(0.5)$, is equivalent to minimizing the sum of absolute values of the residuals.

• From the definition, $\hat{Q}_{y|x}(0.5|x = x_t) = x_t'\hat{\beta}(0.5)$ represents the estimate of the conditional median of $y_t$.

• For different values of $\tau$, the estimate $\hat{\beta}(\tau)$ is the effect of $x_t$ on the $\tau$th quantile of $y_t$.

• An estimate of the entire function $\hat{Q}_{y|x}(\tau|y = x_t)$ can be computed from the above relation. For a more detailed introduction to quantile regressions, please refer to Koenker and Hallock (2000), and Koenker (2000).
3  The Capital Asset Pricing Model (CAPM) as a Regression:

- Another application of regression analysis
  - We can test the CAPM in cross-sectional regressions
  - We can also test it in time-series regressions

- You must have seen the CAPM in some of your previous classes.

- The CAPM assumes that there is only one source of risk—Market Risk.

- If you invest in the stock market, you will be exposed to a certain degree to market risk.

- For the exposure to market risk, you will be compensated.

- The compensation will be proportional to your risk exposure.

- IMPORTANT: Only market risk affects expected returns.
The model is:

\[ R_{i,t} - R^f = \alpha_i + \beta_i \left( R^M_t - R^f \right) + \varepsilon_{i,t} \]

where

- \( R_{i,t} = \) return on asset \( i \) at time \( t \).
- \( R^f = \) return of riskless asset at time \( t \) (is it necessary to have a subscript??)
- \( R^M_t = \) return on the market portfolio at time \( t \).
- \( \alpha_i \) and \( \beta_i \) are the coefficients to be estimated.
- **AND:** \( \text{Cov} \left( R^M_t, \varepsilon_{i,t} \right) = 0 \)
Note that if $\alpha_i = 0$, then
\[ E(R_{i,t}) = R^f + \beta_i E(R^M_t - R^f) \]

- The expected return on asset $i$ will be the risk free rate $R^f$ plus $\beta E(R^M_t - R^f)$

- The quantity $E(R^M_t - R^f)$ is called the market risk premium, the difference between the return on the market portfolio and the return on a riskless bond.

- The coefficient $\beta$ is the sensitivity of return $i$ to market risk.
  - If $\beta = 0$, asset $i$ is not exposed to market risk. Therefore, the investor is not compensated with higher return.
  - If $\beta > 0$, asset $i$ is exposed to market risk and $R_{i,t} \geq R^f$, provided that $E(R^M_t - R^f) > 0$.

- Is it reasonable to assume that $E(R^M_t - R^f) > 0$.

- Historically, what is $ER^M_t = ??$? How about $R^f = ??$?
The Equity Premium Puzzle

- Note: The difference $E \left( R_t^M - R^f \right)$ has historically been very high.

- By “very high”, I mean that no reasonable economic model can justify such a huge premium for holding stocks versus a riskless asset.

- This puzzling difference, called the “Equity Premium Puzzle,” has been a topic of active research for the past 20 years.

- There are two possible (but not mutually exclusive) ways of rationalizing this puzzle:
  - Economic models are not realistic.
  - The equity premium is a statistical fluke.

- The simple difference $E \left( R_t^M - R^f \right)$ has generated literally thousands of articles in the academic and popular press.
Back to the CAPM:

- The assumption that only market risk is compensated, or priced, is a vice and virtue.

- VIRTUE: The simplicity of the CAPM makes it a natural benchmark against which to measure all other asset pricing models.

- VICE: It is unreasonable to suspect that only market risk is compensated.
  - If we ignore other important sources of risk, what would happen to us????????
Digression: Diversification

- The realized return is:
  \[ R_{i,t} - R_t^f = \alpha + \beta \left( R_t^M - R_t^f \right) + \varepsilon_{i,t} \]

- But the expected return is
  \[ E R_{i,t} = R_t^f + \beta E \left( R_t^M - R_t^f \right) \]

- Therefore, we ignore one source of fluctuation, \( \varepsilon_t \).

- But, didn’t we say that variance of returns is always bad and investors should be compensated for being exposed to variance???????
Here is an illuminating example of what is going on.

- As a starting point, note that the total variance of $R_{i,t}$ is
  
  $$Var(R_{i,t} - R^f_t) = \beta^2 Var\left(R^M_t - R^f_t\right) + \sigma^2$$

- Suppose we have $N$ assets (stocks). The return of each of those stocks can be written as

  $$R_{i,t} - R^f_t = \alpha + \beta \left(R^M_t - R^f_t\right) + \varepsilon_{i,t}, \quad i = 1, \ldots, N$$

- Suppose, without loss of generality that $Var(\varepsilon_{1,t}) = Var(\varepsilon_{2,t}) = \ldots = Var(\varepsilon_{N,t}) = \sigma^2$ and they are uncorrelated.

- Now, let’s form a portfolio by putting an equal weight on each return, so that the return of this portfolio would be:

  $$R^p_t = \frac{1}{N} R_{1,t} + \ldots + \frac{1}{N} R_{N,t}$$

  $$= \frac{1}{N} \sum_{i=1}^{N} R_{i,t}$$

  $$= \frac{1}{N} \sum_{i=1}^{N} \left\{ R^f_t + \alpha + \beta \left(R^M_t - R^f_t\right) + \varepsilon_{i,t} \right\}$$

  $$= R^f_t + \alpha + \beta \left(R^M_t - R^f_t\right) + \frac{1}{N} \sum_{i=1}^{N} \varepsilon_{i,t}$$
Now, if we check the variance of the portfolio return

\[
Var(R_p^t - R_f^t) = \beta^2 Var \left( R_M^t - R_f^t \right) + \frac{1}{N^2} \sum_{i=1}^{N} \sigma^2
\]

\[
= \beta^2 Var \left( R_M^t - R_f^t \right) + \frac{1}{N^2} N \sigma^2
\]

\[
= \beta^2 Var \left( R_M^t - R_f^t \right) + \frac{1}{N} \sigma^2
\]

- Note that the first part, the part due to the market risk, is still the same.

- However, the part that is uncorrelated to the market risk has decreased.

- In fact, as \( N \to \infty \), the variance due to the idiosyncratic part of the variance will converge to zero.

- In other words, some parts of the fluctuations can be diversified in such a way that they do not matter.

- Other fluctuations, such as market risk, cannot be diversified away (because all assets are exposed to the same market risk).

- Q: How many stocks \( N = ? \) do we need to hold in order to be diversified? (Campbell et al (2000)—variance \( \sigma^2 \) has increased)
So, in the CAPM, only market risk is compensated. This is clearly not a reasonable assumption. Suppose that

\[ R_{i,t} - R^f = \alpha_i + \beta_i (R^M_t - R^f) + \beta_2 f_{2,t} + \varepsilon_{i,t} \]

But we erroneously estimate the CAPM

\[ R_{i,t} - R^f = \tilde{\alpha}_i + \tilde{\beta}_i (R^M_t - R^f) + u_{i,t} \]

where \( u_{i,t} = \beta_2 f_{2,t} + \varepsilon_{i,t} \).

- Q: Is this a big deal?
- Would we get good estimates of \( \tilde{\alpha}_i \) and \( \tilde{\beta}_i \) that are close to \( \alpha_i \) and \( \beta_i \)?
• Another concern: Roll’s Critique.

• Suppose that the CAPM is the true model, but...

• We don’t observe the “true” market return $R^M_t$. We have a proxy (say SP500), but this is ultimately a proxy, or

\[ R^{SP}_t = R^M_t + v_t \]

where $v_t$ is some random term, $E(v_t) = 0$.

• But there is no reason why $E(v_t \varepsilon_{i,t}) = 0$. In fact, it is very likely that $E(v_t \varepsilon_{i,t}) \neq 0$.

• Then, we run

\[ R_{i,t} - R^f = a_i + \beta_i \left( R^{SP500}_t - R^f \right) + \varepsilon_{i,t} \]

but $E \left( R^{SP500}_t \varepsilon_{i,t} \right) \neq 0$

• Our estimates of $a$ and $\beta$ will be inconsistent.
• Suppose that $E(f_{2,t}, R^M_t) \neq 0$. (Usually factors are correlated)

• Therefore $E(R^M_t u_{i,t}) \neq 0$.

• What does this imply about the estimates of $\tilde{\alpha}$ and $\tilde{\beta}$?

• Why?

• Conclusion: If the CAPM is misspecified, in the sense that there are omitted factors, we get inconsistent estimates of $\alpha$ and $\beta_1$.

• Therefore, our tests of the CAPM will be invalid.
How do we test the CAPM?

Two ways (unrelated):
- Test whether the pricing errors $\alpha_i$ are small, or whether $\alpha_i = 0$.
  * We can think of doing this test for each asset.
  * There is a more efficient way, by jointly testing if $\alpha_1 = \alpha_2 = \ldots = \alpha_N = 0$.
- Test whether adding other factor changes the estimates of $\beta$ (suggesting misspecification).
MULTIFACTOR MODELS: We generalize the one-factor model as:

- Idea: Suppose we write returns as
  \[ R_{i,t} - R_f = \alpha_i + \beta_{1,i}f_{1,t} + \beta_{2,i}f_{2,t} + \ldots + \beta_{k,i}f_{k,t} + \varepsilon_{i,t} \]
  where
  \[ f_{1,t} = R_t^M - R_f \]
  \[ f_{2,t} \] is the second factor, etc...
  \[ \text{In this example we have } k \text{ factors.} \]
  \[ \text{As before, all factors are uncorrelated with } \varepsilon_{i,t} \]
  \[ \text{Ideally, the factors are mutually uncorrelated, i.e.} \]
  \[ cov(f_{i,t}, f_{j,t}) = 0, \text{ all } i \neq j. \] Whether this is true or not depends on how we obtain the factors.
  \[ \text{The $1M$ question is: What are those other factors?} \]
  \[ \text{This “generalization” of the CAPM is called the} \]
  Arbitrage Pricing Theory (APT) model.
  \[ \text{If the factors are not necessarily mutually uncorrelated, then the model is called ICAPM (intertemporal CAPM).} \]
  \[ \text{For this class the difference between ICAPM and APT will not be emphasized.} \]
There are a few ways of identifying the factors.
  – Economic arguments (Chen, Roll and Ross (1986))—business cycle variables, etc.
  – Statistical arguments (principal components, discriminant analysis, factor analysis, dynamic factor analysis, etc.)
  – Other arguments (Fama and French (1993) 3-factor model)
  – Dimensional Funds Advisors (DFA).
• The CAPM and the APT are at the core of modern asset pricing (although in slightly more sophisticated forms).

• Why do we call it asset pricing, when we are really looking at returns?

• Why don’t we look directly at prices?

• Why are the CAPM and APT so popular (in academia as well as in practice).

• Because they are simple, linear models and can be estimated and tested using ......
For the CAPM regression, we need:

- Risky assets return
- Riskless asset return
- Market return

For the APT, in addition to the above, we need

- The other factors besides the market return
Tests of APT:

- $\alpha_i = 0$ (pricing errors). Why?

$$R_{i,t} - R^f = \alpha_i + \beta_{1,i}f_{1,t} + \beta_{2,i}f_{2,t} + \ldots \beta_{k,i}f_{k,t} + \varepsilon_{i,t}$$

We always run OLS with intercept (Rule!).

- $\beta_i$ are also of interest (what factors are priced)

- We are still searching for the “right” model for equity returns.

- Q: Can we price fixed income securities or derivatives with the CAPM/APT approach?
In next week’s assignment, you will have to estimate a CAPM and an APT model.

Can we really price assets using the CAPM and the APT?

What use are they?
To think:

- We always assume that
  \[ R_{i,t} - R^f = a_i + \beta_i \left( R^M_t - R^f \right) + u_{i,t} \]
  and \( a_i \) and \( \beta_i \) are time-invariant (no \( t \)-superscript).
- Is this implication of the CAMP realistic?
- Should the CAPM be “generalized?”
- How?
What have we learned today?

- Linear Regression:
  - Condition for consistency
  - Condition for asymptotic (and finite sample) normality
  - Efficiency
  - Identification