1. Quick Review: CAPM
2. Better Asset Pricing Model: APT
3. Market Efficiency
4. Variance Ratio Tests
5. Forecastability of Returns: Univariate and Bivariate Systems
6. Long-Horizon Forecasts of Returns
7. Multivariate Systems (VAR): Reduced forms
Quick Review: CAPM

- Recall from last class:

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

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:** \( Cov(R^M_t \varepsilon_{i,t}) = 0 \)
• 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).
3 MULTIFACTOR MODELS:

- Generalization of the one-factor model
- 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...
- In this example we have \( k \) factors.
- As before, all factors are uncorrelated with \( \varepsilon_{i,t} \)
- Ideally, the factors are mutually uncorrelated, i.e.
  \[ \text{cov}(f_{i,t}, f_{j,t}) = 0, \text{ all } i \neq j. \]
  Whether this is true or not depends on how we obtain the factors.
- The $1M$ question is: What are those other factors?
- This “generalization” of the CAPM is called the Arbitrage Pricing Theory (APT) model.
- If the factors are not necessarily mutually uncorrelated, then the model is called ICAPM (intertemporal CAPM).
- 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
  \[ R_{i,t} - R^f = a_i + \beta_i (R^M_t - R^f) + u_{i,t} \]
  and \( a_i, \) and \( \beta_i \) are time-invariant (no \( t \)-supscript).

- Is this implication of the CAMP realistic?

- Should the CAPM be “generalized?”

- How?

- Is FED policy risk priced? Can you answer this question within the framework of the APT?

- Market Efficiency (Chapter 2)

- Earlier Models of Asset Pricing:
  - Fair Game: Cardano’s (c.1565) *LiberdeLudoAleae* (CLM, 1997)

  The most fundamental principle of all in gambling is simply equal conditions, e.g. of opponents, of bystanders, of money, of situation, of the dice box, and of the die itself. To the extent to which you depart from that equality, if it is in your opponent’s favor, you are a fool, and if in your own, you are unjust.

- We have already seen the market efficiency condition in class 1, but now we will go deeper and provide a test based on the price process.
\[ E(P_{t+1} | I_t) = P_t \]

which implies that

\[ P_{t+1} = P_t + \varepsilon_{t+1} \]

with \( E(\varepsilon_{t+1} | I_t) = 0 \)

- A martingale is a special case of the above. If \( I_t = \{P_t, P_{t-1}, \ldots\} \), so that \( E(P_{t+1} | P_t, P_{t-1}, \ldots) = P_t \)
• Note: This is an AR(1) model with $\phi = 1$ (non-stationary).

• The martingale condition (also known as a “Random Walk”) places restrictions only on the first moments on the process. However, it is silent on second, third, fourth, etc. moments.
  – Specifically, we know that there has got to be a trade-off between risk and return.

• LeRoy (1973), Lucas (1978): The Martingale condition is neither a necessary nor a sufficient condition for rational expectations models of asset prices.

• However, the martingale is an important starting point.
• There are three versions of the Random Walk model.

• The variations place various restrictions on the innovations (surprise, news) process, $\varepsilon_t$ in

$$P_t = \mu + P_{t-1} + \varepsilon_t$$

– $\varepsilon_t$ are i.i.d $E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2$.

– If we place some restrictions, we can drop the “identically distributed” condition, but keep the “independent” condition. I.e. $\varepsilon_t$ are independent.

– $\varepsilon_t$ are uncorrelated (but they might be dependent).

• Note that the above process has a “stochastic trend”. To see why, notice that, if we solve recursively the above process, we have

$$P_t = \mu t + \sum_{i=1}^{t} \varepsilon_i + P_0$$

– $E(P_t) = \mu t$

– $Var(P_t) = \sigma^2 t$

– Note that all three forms have the same two moments.

• Review Q: Do you see why $P_t$ is not (covariance) stationary?
• Is the process $P_t = \mu + P_{t-1} + \varepsilon_t$ adequate to model stock prices? Here is a hint:
We will use $p_t = \log P_t$ (or $P_t = e^{p_t}$) and

$$p_t = \mu + p_{t-1} + \varepsilon_t$$

so that $p_t = \mu t + \sum_{i=1}^{t} \varepsilon_t + p_0$, or to get back to prices, $P_t = e^{P_0} e^{\mu t} e^{\sum_{i=1}^{t} \varepsilon_t}$.

- Note that $p_t - p_{t-1} = \mu + \varepsilon_t = r_t$ (in our previous notation), is the continuously compounded return.

- If we assume that $\varepsilon_t$ (and hence $r_t$) is normally distributed, then we are back to the case where $r_t$ is normally distributed and $1 + R_t = e^{r_t}$ is lognormally distributed.

- The process $p_t = \mu + p_{t-1} + \varepsilon_t$ is a discrete (arithmetic) Brownian motion

- The process $P_t = e^{p_t}$ is a discrete geometric Brownian motion.
Note: An immediate observation of the above specification for prices is that returns are not forecastable by past returns or by any other conditioning variable, since

\[ r_t = \mu + \varepsilon_t \]

and

- \( \varepsilon_t \) are either uncorrelated or independent.
- \( \mu \) is the mean and it does not vary with time.
4 Variance Ratio Tests

- The variance ratio test is a neat and “natural” way to test market efficiency.
- We cannot work with prices $p_t$ or $P_t$, because those series are non-stationary. So far, we know how to work only with stationary time-series.
- Recall: DO NOT work with non-stationary time series, unless you know what you are doing.
- The non-stationary processes of interest to us ($p_t$) can be stationarized....How?
- Therefore, we will have a preference to work with ....
• As we saw above, an implication of the Random Walk model of prices is that:
  – Continuously compounded (or log) returns are unforecastable
  – The increments (log returns) are uncorrelated (or independent, or iid)

• Here is a useful observation about the variance of a two period (log) return \( r_{t,t+2} \)
  \[
  \text{Var} \left( r_{t,t+2} \right) = \text{Var} \left( r_{t+1} + r_{t+2} \right) \\
  = \text{Var} \left( r_{t+1} \right) + \text{Var} \left( r_{t+2} \right) \\
  = 2\sigma_r^2
  \]

• In general, for the q-th period (long-horizon) return \( r_{t,t+q} \):
  \[
  \text{Var} \left( r_{t,t+q} \right) = \text{Var} \left( r_{t+1} + ... + r_{t+q} \right) \\
  = q\sigma_r^2
  \]

• Those implications were derived under the hypothesis (the null) that returns are unforecastable.
• So, we have a “natural” test for the Random Walk model (or equivalently, for unforecastable returns):

\[ \text{Var}(r_{t,t+q}) \overset{\text{Under the Null of a Random Walk}}{=} q \text{Var}(r_{t+1}) \]

or

\[ VR(q) = \frac{\text{Var}(r_{t,t+q})}{q \text{Var}(r_{t+1})} = 1 \]

• This is a simple test.

• Compute the variance of the \( q \)-period returns.

• Compute the variance of the 1-period returns.

• It must be the case that the ratio of \( \frac{\text{Var}(r_{t,t+q})}{q \text{Var}(r_{t+1})} \) must be close to 1.

• Q: How close? We should not forget that the test \( VR(q) \) is a random variable with a corresponding density, etc.

• Q: For what values of \( VR(q) \) is the model a RW and for what values it is not?

• ASIDE: We can relate the \( VR(q) \) test to autocorrelations (see CLM, 1997).

• ASIDE: What \( q \) to choose?
So, under the null hypothesis that $r_t$ are i.i.d., we can use the CLT to show that
\[ \sqrt{Tq \left( V R(q) - 1 \right)} \sim^a N \left( 0, 2(q - 1) \right) \]
(recall last lecture and the results about $\bar{X}$ and $\hat{\beta}$)
or
\[ \left( V R(q) - 1 \right) \sim^a N \left( 0, \frac{2(q - 1)}{Tq} \right) \]
Therefore, we can conduct testing in the usual way:
Form a statistic
\[ Z = \frac{V R(q) - 1}{\sqrt{\frac{2(q - 1)}{Tq}}} \sim^a N \left( 0, 1 \right) \]
So, if $Z$ is greater than 1.96, we reject the null of IID returns at the 5% level.
• Note: We must be careful when computing the long-horizon returns \( r_{t,t+q} \) from overlapping observations.

• Q: Why?

• A: Because, under the null hypothesis, \( r_{t,t+q} \) are i.i.d. If we use overlapping observations, they will NOT be i.i.d. by construction.

• Some people advocate the use of overlapping observations and correcting for correlation from the overlap as:

\[
\sqrt{Tq \left( VR_{\text{overlap}}(q) - 1 \right)} \sim^a N \left( 0, \frac{2(2q-1)(q-1)}{3q} \right)
\]

• The correction of the variance is supposed to correct for the overlap. At the same time we have more observations, so the test might be “better”, or more powerful.

• People are deluding themselves!

• If we correct exactly for the overlap, whether we use overlap or not would make no difference. But the uncertainty introduced around how to deal with the overlap makes the second test less desirable.

• At the end of the day, we have the same amount of information (returns), not more.
• NOTE: As part of your homework, you will have to program the \( VR \) test using non-overlapping and overlapping returns and for different portfolios.

• Compare your results to those in CLM (chapter 2).

• Answer the question: Do you believe the RW hypothesis?
5 Forecastability of Returns: Univariate and Bivariate Systems (CLM, Chapters 2 and 7)

- Thus far, we defined returns $r_t = \log (1 + R_t)$ as

$$r_t = p_t - p_{t-1} = \log \frac{P_t}{P_{t-1}}$$

- Of course, if $r_t = \mu + \varepsilon_t$ is true, we can also run simple regression tests:
  - We can write $r_t = \mu + \phi r_{t-1} + \varepsilon_t$, where under the null hypothesis, $\phi = 0$.
  - Run: $r_t$ on $r_{t-1}$ (this is an AR(1) model).
  - Test that the parameter $\phi$ is equal to zero.
  - Those tests are perfectly valid and can be related to the $VR$ test.
• ASIDE: There is some evidence that, at short horizons, $r_t$ is positively correlated for some stocks. That implies that if returns of certain stocks are higher than average at $t - 1$, they would tend to be higher than average at $t$. (Jagadeesh and Titman)

• This is called MOMENTUM and it is not uncontroversial.

• Momentum disappears at longer horizon and even “reverses” itself.
  – Q: If agents are rational and markets are efficient, why don’t people take advantage of momentum?
  – Q: Can we find “rational” explanations of momentum?

• There is no agreed upon consensus on momentum.
- If

\[ r_t = p_t - p_{t-1} \]
\[ = \log \frac{P_t}{P_{t-1}} \]

- Then

\[ e^{r_t} = (1 + R_t) = \frac{P_t}{P_{t-1}} \]
\[ R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \]

- Note: This definition has no place for dividends. But dividends must carry important information about the viability of a company.

- Can it be the case that, if we do the calculations correctly and incorporate dividends, then that information might forecast returns?
• The exact definition of a return is
  \[ R_{t+1} = \frac{P_{t+1} + D_{t+1} - P_t}{P_t} \]
  or
  \[ 1 + R_{t+1} = \frac{P_{t+1}}{P_t} + \frac{D_{t+1}}{P_t} \]

• Therefore, the increase in returns can come either from an increase in the stock price or from a dividend payment.

• Here, we will use a trick that has proven useful.

• Take logs
  \[ \log (1 + R_{t+1}) = \log (P_{t+1} + D_{t+1}) - \log (P_t) \]

• Now, I will do a trick using the Taylor’s expansion (see CLM, p. 261)

•
  \[ r_{t+1} \approx k + \rho p_{t+1} - p_t + (1 - \rho) d_{t+1} \]
  where
  – \( d_{t+1} = \log (D_{t+1}) \) is the log dividend
  – \( k \) is a constant term
  – \( \rho \) is a constant of linearization that we will treat as a parameter.
• The above transformation is “cool” because now returns, prices, and dividends are related by a linear relation.

• But there is a problem: This is not exactly a regression, because prices and dividends are not stationary. . . . Remember.

• But, we can write

\[ p_t \approx k + \rho p_{t+1} - r_{t+1} + (1 - \rho) d_{t+1} \]

• Solve FOREWARD (not backward as before)

\[ p_{t+1} \approx k + \rho p_{t+2} - r_{t+2} + (1 - \rho) d_{t+2} \]

or

\[ p_t \approx k + \rho \left[ k + \rho p_{t+2} - r_{t+2} + (1 - \rho) d_{t+2} \right] - r_{t+1} + (1 - \rho) d_{t+1} \]

\[ \approx K + \rho^2 p_{t+2} - \rho r_{t+2} - r_{t+1} + \rho (1 - \rho) d_{t+2} + (1 - \rho) d_{t+1} \]

\[ \approx \ldots \]

\[ \approx K + \rho^j p_{t+j} - \sum_{i=0}^{j} \rho^i r_{t+1+i} + \sum_{i=0}^{j} \rho^i (1 - \rho) d_{t+1+i} \]

• Now, we have to assume a no-bubbles condition (interpretation):

\[ \lim_{j \to \infty} \rho^j p_{t+j} = 0 \]
Therefore:

\[ p_t \approx K - \sum_{i=0}^{\infty} \rho^i r_{t+1+i} + \sum_{i=0}^{\infty} \rho^i (1 - \rho) d_{t+1+i} \]

- This is quite intuitive: If price is “high”, it means that either
  - dividends in the future would be high
  - returns in the future would be low
  - This presumes a no-bubbles condition. Note a difference between fundamentals and the price can exist, but it cannot last indefinitely.
The last step is to subtract the above expression from \( d_t \) to obtain:
\[
d_t - p_t = K - \sum_{i=0}^{\infty} \rho^i (1 - \rho) \Delta d_{t+1+i} + \sum_{i=0}^{\infty} \rho^i r_{t+1+i}
\]

This is a very, very useful relation, because
- The log dividend price ratio \( d_t - p \) is stationary
- \( \Delta d_{t+1+i} \) is stationary
- \( r_{t+1+i} \) is stationary.
- The log dividend price ratio must forecast either future dividend growth or future returns
- This is the relation we had obtained a few slides ago, but now it is in a “regression” form.
- We can use regressions to test this relation.

The last observation is that we are dealing with future values. But this has never stopped us.
\[
d_t - p_t = K - E_t \left\{ \sum_{i=0}^{\infty} \rho^i (1 - \rho) \Delta d_{t+1+i} + \sum_{i=0}^{\infty} \rho^i r_{t+1+i} \right\}
\]
Before going further, let’s take a step back and think of what we have done so far:

- In the absence of dividends, we argued that returns are unforecastable.
- Now, we argue that dividends can help us forecast returns.
- Isn’t that inconsistent?
Testing the above relation:

- Regress: $r_{t+1}$ on $d_t - p_t$
- Regress: $\Delta d_{t+1}$ on $d_t - p_t$
- Which relationship is stronger?
- Note: This is a forecasting relationship: the dividend price ratio today helps us forecast tomorrow’s (expected) returns
- It is different from the CAPM, APT, where we were trying to explain ex-post variations in returns.
- Here, we are doing an ex-ante regression.
6 Long-Horizon Regressions:

People have noticed that the relation
\[ d_t - p_t = K - \sum_{i=0}^{\infty} \rho^i (1 - \rho) \Delta d_{t+1+i} + \sum_{i=0}^{\infty} \rho^i r_{t+1+i} \]
might not hold in the short run, but it has got to hold in the long run.

- The above observation has prompted people to run regressions:
  \[ -r_{t+1} + \ldots + r_{t+k} = \alpha_1 + \beta_1 (d_t - p_t) + u_{t+k} \]
  \[ -\Delta d_{t+1} + \ldots + \Delta d_{t+k} = \alpha_2 + \beta_2 (d_t - p_t) + e_{t+k} \]

- But then, it must be the case that
  \[ \beta_1 + \beta_2 = 1 \]

  by definition. We have not used anything, but the definition of returns, no-bubbles condition, and a log-linearization.

- The above restriction is never used in practice. How can it be implemented?
Remark: As in the variance ratio tests, it is important to run the long horizon regressions using non-overlapping returns.

If we use overlapping returns, we get spurious results.

Correcting for the overlap is difficult and introduces another layer of complication.
Vector Autoregressions (VAR)

- First, we should not confuse VAR (vector autoregression) with VaR (Value at Risk).
- The VAR is a natural generalization of the autoregressive process, for multivariate series.
- Suppose we are interested in the joint, dynamic interaction between a few series, say returns and volatility.
  - We suppose that higher volatility must lead to higher returns
  - We also think that there might be some feedback effect from returns to volatility Campbell and Hentschel (1992).
  - Ultimately, we are not sure about the dynamic relationship between those two series
Then, we write

\[ Y_t = \begin{bmatrix} r_t \\ \sigma_t \end{bmatrix} \]

where \( r_t \) is the return and \( \sigma_t \) is its volatility (standard deviation) at time \( t \). We want to write the following dynamic relationship

\[
\begin{align*}
  r_t &= \alpha_1 + \beta_{11} r_{t-1} + \beta_{12} \sigma_{t-1} + \varepsilon_{1,t} \\
  \sigma_t &= \alpha_2 + \beta_{21} r_{t-1} + \beta_{22} \sigma_{t-1} + \varepsilon_{2,t}
\end{align*}
\]

- Both series depend on their own lagged values (as in the AR process)
- Both series depend on the lagged realization of the other process
- The residuals \([\varepsilon_{1t}, \varepsilon_{2t}]\) have a covariance matrix \( \Sigma \).
- This system can be written in a more elegant form as:

\[
Y_t = \alpha + \Phi Y_{t-1} + \varepsilon_t
\]

where \( \Phi = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \) and \( \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \).
The beauty of VARs

- Very easy to estimate: Run regressions of variables on their lagged values and on the lagged values of all other variables
- Run the regressions, equation by equation, and then stack them together.
- We don’t need a priori theory to know what variable causes what variable, etc.
- The VARs help us make the notion of statistical “causality” very precise.
Statistical (Granger) causality:
- Example: Martian observes that when people bring umbrellas, it starts raining. She concludes that people bringing umbrellas causes rain.
- But it is the other way around...The anticipation of rain makes people bring their umbrellas.
- In this example, the statistical causality runs from “umbrellas” to “rain”
- The structural causality runs from “rain” to “umbrellas”.
- Weather and orange juice futures example. – argument for market efficiency.
- Problem: In empirical work, if we don’t have a model (or more information about the system), we cannot distinguish between the two alternatives.
Causality in a bivariate VAR:

\[ r_t = \alpha_1 + \beta_{11} r_{t-1} + \beta_{12} \sigma_{t-1} + \varepsilon_{1,t} \]
\[ \sigma_t = \alpha_2 + \beta_{21} r_{t-1} + \beta_{22} \sigma_{t-1} + \varepsilon_{2,t} \]

- Suppose that \( \beta_{12} = 0 \). Then \( r_t = \alpha_1 + \beta_{11} r_{t-1} + \varepsilon_{1,t} \), or \( r_t \) is not affected by \( \sigma_{t-1} \). In such a case, we say that \( \sigma_{t-1} \) does not Granger-cause (or just cause) \( r_t \).

- However, since \( \beta_{21} \neq 0 \), \( r_{t-1} \) does Granger-cause \( \sigma_t \).

- It is important to understand that Granger-causality gives us the timing (umbrella, then rain) but not the economic story.

- This is a very common mistake in academia, in practice, and in everyday life.
Test for Granger-causality:

- We run the two regressions.
- For the hypothesis: “$\sigma_{t-1}$ Granger-causes $r_t$”, we test $\beta_{12} = 0$
- For the hypothesis: “$r_{t-1}$ Granger-causes $\sigma_t$”, we test $\beta_{21} = 0$
- It is always a good idea to start an empirical work with some background Granger-causality tests, be it only to get a feel for the data.
- But it is not a good idea to start weaving economic stories based on that evidence alone.
- We will come back to VARs.
Here is a real research problem (Goto and Valkanov (2001)):

- Fact: The effect of the FED actions on the stock market is negative. In other words, a contractionary monetary policy (aimed at curbing inflation) will result in lower returns for some time in the future.

- Question: Why is this so? There are two possibilities
  - Fed has a better forecast of the state of the economy, but its policy has no real effect on stock fundamentals (the umbrella). This is only Granger-causation without structural effect.
  - Fed has an impact on the economy and by contracting the economy, cash flows go down, returns decrease (the cloud). This is Granger-causation and a structural effect.

- Q: Is there something about the stock market that would help us distinguish between the two alternatives?