Today’s Agenda

● Announcements
  – Final: 2-sided, 8.5 by 11 inch cheat-sheet allowed.
  – Open book
  – Study:
    ∗ Lecture notes (50% of your time)
    ∗ Homework (35% of your time)
    ∗ Book (15% of your time)

1. Non-Linear Dependence
2. Market Microstructure
3. Bootstrap and Simulation of Standard Errors
4. Filtering: Time Domain
5. Bayesian Estimation
6. Next Time
   – Portfolio Selection
   – New Methods
Comment: The Fama-French (1993) way to form risk factors–DFA

Recall that the Fama-French regression is:

\[ R_{i,t} - r^f_t = \alpha_i + \beta_{1,t} (R^S_{t} - r^f_t) + \beta_{2,t} (R^S_{t} - R^L_{t}) + \]
\[ \beta_{3,t} (R^V_{t} - R^G_{t}) + \beta_{4,t} (R^M_{t} - R^R_{t}) + \varepsilon_{i,t} \]

where

- \((R^S_{t} - r^f_t)\) is the market risk premium
- \((R^S_{t} - R^L_{t})\) is the small stocks premium
- \((R^V_{t} - R^G_{t})\) is the value premium
- \((R^M_{t} - R^R_{t})\) is the momentum premium

Note that:

1. All these factors are zero-cost (i.e., long-short) portfolios
2. The factors have been constructed by sorting cross-sectionally monthly, quarterly, or annually.
3. Any other factor might be constructed in the same way
4. Recent examples:
   (a) Real Estate holdings (Tusel (2007))
   (b) Inventory premium
   (c) New stock issuance (Baker and Wurgler (2005))
1 Non-linear Dependence

- In finance, the characterization of the variation of returns and the dependence between returns is crucial.
- Thus far, the variation has been captured by the variance.
- The dependence has been captured by the covariance.
- The covariance is a good measure only with linear models.
- Non-linear dependence needs to be addressed.
- Suppose we have two assets $R$ and $F$ with returns $R_S$ and $R_F$ whose distribution is $F_{R_S}$ and $F_{R_F}$.
- The “complete” behavior is defined by their distributions.
• We might want to know how the entire distribution of the two variables. In other words, we want to know the joint distribution, $F_{R_S, R_F}$.

• Q: Why?

• A: Because there might be nonlinearity in the relation between the two variables.

• Find a general way of relating the two distributions, $F_{R_S}$ and $F_{R_F}$.

• For instance, how would the tails of the distributions change during extreme events? More? Less?
  – With credit derivatives, suppose that the credit rating of the underlying (Bond) changes.

• Or, can we model the dependence non-linearly?

• Would we see skewness, kurtosis, etc.?

• There is a lot of interest in the profession in finding adequate (alternative) measures of dependence that go beyond simple correlations.
There are two (three) main ways to go:
- Copulas
- Tail dependence
- Quantile Regressions

These are different approaches at modeling the non-linear dependence between the distributions $F_{RS}$ and $F_{RF}$.
1.1 Copulas

The problem is as follows. Find a general way of relating the two distributions, \( F_{R_S} \) and \( F_{R_F} \) in order to find the joint distribution \( F_{R_S,R_F} \).

- In other words, we want to do the following
  \[
  F_{R_S,R_F} = C(F_{R_S}, F_{R_F})
  \]
  where \( C(., .) \) is some appropriately chosen function.

- The goal is to choose the right \( C(., .) \).
Note: We cannot choose any $C(.,.)$. There are some restrictions that are imposed by the fact that we are dealing with distributions.

- Simple example; $C(x, y) = x\cdot y$
  - Q: When is this copula appropriate?

- Another example: $C(x, y) = \min(x, y)$

- Another example: $C(x, y) = N_{\rho}(N^{-1}(x), N^{-1}(y))$
  where $\rho$ is the correlation.
Two widely used copulas:

– Gumbel’s bivariate exponential:
\[ C_\theta(x, y) = x + y - 1 + (1 - x)(1 - y) e^{-\theta \ln(1-x)(1-y)} \]
where \( \theta \) measures the non-linear dependence.

For \( \theta = 0 \), we have
\[ C_{\theta=0}(x, y) = x \cdot y \]

– AMH Copulas (Ali-Mikhail-Haq):
\[ C_\theta(x, y) = \frac{xy}{1 - \theta (1 - x)(1 - y)} \]
where \( \theta \) measures the non-linear dependence.

For \( \theta = 0 \), we have
\[ C_{\theta=0}(x, y) = x \cdot y \]
In general, if you work with copulas:
- You want to find a flexible form for $C(.,.)$
- You want to be able to estimate all the dependence parameters.
1.2 Tail dependence

The main question is as follows. What is the probability that we will have an extreme (tail) realization of $R_F$ given that we observe an extreme tail realization of $R_S$?

• Note: We are not looking at typical co-movements (covariance) but at tail co-movements?

• A natural definition of tail dependence is:
  – On the positive side:
    \[
    \lambda_+ = \Pr(R_S > \text{extreme positive value} \mid R_F > \text{extreme positive value})
    = \lim_{u \to 1^-} \Pr(R_S > F_{R_S}^{-1}(u) \mid R_F > F_{R_F}^{-1}(u))
    \]
  – On the negative side:
    \[
    \lambda_- = \Pr(R_S < \text{extreme negative value} \mid R_F > \text{extreme negative value})
    = \lim_{u \to 0} \Pr(R_S < F_{R_S}^{-1}(u) \mid R_F < F_{R_F}^{-1}(u))
    \]
• I.e., what is the chance of observing a very positive (negative) return in the underlying asset, given that we have observed a very positive (negative) return in the hedging asset.

• Instead of average dependence, we are looking at tail dependence.

• Intuitively, copulas and tail dependence are related.

• There are also mathematical connections between the two:

\[
\lambda_+ = \lim_{u \to 1} \frac{1 - 2u + C(u, u)}{1 - u}
\]

\[
\lambda_- = \lim_{u \to 0} \frac{C'(u, u)}{u}
\]

• For instance, for the Gumbel’s copula

\[
\lambda_+ = 1 - 2^\theta
\]

• Note: \( \lambda_+ = 0 \) means no tail dependence (when \( \theta = 0 \)). But the distributions might be otherwise dependent.
1.3 Quantile regression:

- Model the conditional quantiles of the distribution
- Instead of

\[ R_F = \alpha + \beta R_S + \varepsilon \]

where we model the conditional mean, we want to model the conditional quantiles.

- We do that by specifying:

\[ Q_{R_F}(\tau) = \alpha(\tau) + \beta(\tau) R_S + \varepsilon \]

for \( \tau \) between 0 and 1 where \( Q_{R_F}(\tau) \) is the \( \tau \) quantile of \( R_F \).

- Recall the definition of a quantile: (inverse of distribution).

- This is called a quantile regression, because we are modeling the quantiles of the distribution.
  - For \( \tau = 0.5 \), we are estimating how the median of the distribution of \( R_F \) depends on \( R_S \).
  - Similarly, for \( \tau = 0.05 \) or \( \tau = 0.99 \), we are estimating the extreme quantiles of the distribution.

- Quantile regressions are easy to estimate (similar to OLS).
• Example:
Liquidity is like... love. Everybody knows what it is, but it is hard to explain.

Ability to buy or sell significant quantities of a security quickly, anonymously, and with minimal to no price impact.

Liquidity is very important for a well-functioning financial market.

Market-makers: provide liquidity by taking the opposite side of a transaction.
- If an investor wants to buy, the market-maker sells and vice versa.

In exchange for this service, market-makers buy at a low bid price $P^b$ and sell at a higher ask price $P^a$.

This ability to buy low and sell high insures that the market-makers will make some profits.

The difference $P^a - P^b$ is called the bid-ask spread.
• The bid-ask spread complicates things a bit, since we don’t observe the true price.
  – We have three prices: The bid, the ask, and the true price.
  – The true price is often between the bid and the ask, although it need not be.
  – How do we define returns: From bid to bid, from ask to ask, from bid to ask...
  – How is the bid-ask spread determined?

• It is fairly intuitive that the bid-ask spread has an effect on returns.

• Roll (1984) provides a particularly simple and appealing model of how the bid-ask spread might impact the time-series properties of returns.

• This model provides most of the intuition and the framework on how financial economists think about the bid-ask spread.
The observed market price is
\[ P_t = P^*_t + I_t s \]
- \( P^*_t \) is the fundamental price in a frictionless economy
- \( s \) is the bid-ask spread
- \( I_t \) is an iid index variable that takes values of 1 with probability 0.5 (if the trade is initiated by a buyer) and -1 with probability 0.5 (if the trade is initiated by a seller).

Note that \( E (I_t) = 0 \) and \( Var (I_t) = 1 \).

For simplicity assume that \( P^*_t \) does not change.

The change in price is
\[ \Delta P_t = \Delta P^*_t + \frac{s}{2} I_t - \frac{s}{2} I_{t-1} \]

Its variance, covariance, and correlation are:
\[ Var (\Delta P_t) = 2 \frac{s^2}{4} = \frac{s^2}{2} \]
\[ Cov (\Delta P_t, \Delta P_{t-1}) = -\frac{s^2}{4} \]
\[ Cov (\Delta P_t, \Delta P_{t-k}) = 0, k > 1 \]
\[ Corr (\Delta P_t, \Delta P_{t-1}) = -\frac{1}{2} \]
• The fundamental value is fixed, but there is variation from $s$.

• The bid-ask spread induces negative correlation in returns even in the absence of other fluctuations!

• This is quite intuitive, right?

• The variance and serial correlation depend on the magnitude of the bid-ask spread.

• This particular example induces a first-order serial correlation.

• We can also express the spread as a function of the covariance, or
$$ s = 2 \sqrt{-\text{Cov}(\Delta P_t, \Delta P_{t-1})} $$

• However, in practice $\text{Cov}(\Delta P_t, \Delta P_{t-1}) > 0$ is not uncommon.

• Roll (1984) defines the spread as
$$ s = -2 \sqrt{|\text{Cov}(\Delta P_t, \Delta P_{t-1})|} $$

• One can also take $s = 2 \sqrt{-\text{Cov}(\Delta P_t, \Delta P_{t-1})}$ to be a testable implication. The fact that we observe $\text{Cov}(\Delta P_t, \Delta P_{t-1}) > 0$ implies that the model is misspecified (Glosten and Harris (1988) and Stoll (1989)).
• Roll proposes to estimate the effective spread as

\[ s = -2\sqrt{\text{Cov}(\Delta P_t, \Delta P_{t-1})} \]

• But why estimate a quantity that is observable. After all, we know the bid and the ask prices, and the spread?

• Roll argues that the quoted (or observed) spread is different from the effective spread. Sometimes transactions occur at prices within the bid-ask spread because
  – market-makers do not update bid-ask quotes in a timely manner.
  – provide discount to customers that are trading for reasons other than private info (Glosten and Milgrom (1985), Goldstein (1993)).
  – inventory rebalancing.

• The assumption that \( P_t^* \) is fixed might be relaxed. As long as it is independent of \( I_t \), we can express

\[
\text{Corr}(\Delta P_t, \Delta P_{t-1}) = -\frac{s^2}{4} \frac{1}{s^2 + \text{Var}(\Delta P_t^*)}
\]
• Roll’s (1984) model is designed to illustrate how the spread can induce negative serial correlation in returns.

• However, the serial correlation is a function of the spread.

• But the spread is set exogenously.

• The question remains: What determines the bid-ask spread?

• The spread is very important for market-makers.

• The spread is unlikely to be independent of the price $P_t^*$. 
What are the fundamental forces that might have an impact on the bid-ask spread?
- Order-processing costs—basic setup and operation costs.
- Inventory costs—holding an undesired security that is subject to risk.
- Adverse selection costs—some investors are better informed than the market maker about the stock. The market-maker has no way of distinguishing between informed and uninformed traders and must be compensated for the added risk.

Glosten (1987) has a nice model where the adverse selection costs are modelled explicitly.
2.1 Information Content of Stock Trades: Hasbrouck (1988, 1991)

- Premise: People trade because of new information. Then larger (measured by volume) trades—trades that reflect the revelation of ‘lots’ of new information—must have a larger impact on prices than smaller trades.

- Hasbrouck (1991) conducts a VAR analysis and finds that there is such a price impact and that in fact it is quite large.
2.2 Transactions Data

- Recent datasets such as the TAQ (Trades and Quotes) or the Optionmetrics databases have given economists access to a lot of new information.
- The possibilities and the challenges are great.
- Such datasets are often tick-by-tick, all transactions of every stock are recorded.
- The transactions are not evenly spaced.
- The IID assumption fails.
- Most economists are only now starting to develop tools to handle such datasets.
- Discreteness must be taken into account.
- In many instances, people aggregate or smooth the data.
2.3 Buyer- and Seller-Initiated Trades

- When we observe a trade, we observe:
  - $P^*$, the price at which the transaction has occurred
  - $Q^*$, the number of trades bought or sold
- It would be nice if we could observe whether the trade were buyer- or seller-initiated.
- Why?

- Here is an algorithm to do that:
  - We also observe bid and ask prices: $P^a$ and $P^b$
    (What happens when $P^a < P^b$?)
  - Find midpoint as $P^m = \frac{P^b + P^a}{2}$
  - If $P^* > P^m$, buyer-initiated trade
  - If $P^* < P^m$, seller-initiated trade
- Justification: Mark and Ready (1992) algorithm
3 Bootstrap

- Sometimes:
  - The normality assumption is not adequate. But we need an adequate assumption in order to have accurate tests.
  - There are peculiarities of the data that cannot be handled with asymptotic theory
  - The CLT might not provide good approximation (it is only an asymptotic result).

- Example: Short rate might be bi-modal.
- Example: Extreme overlap in data
- Example: Missing values
- Q: What to do?
- IDEA: Treat the sample as if it were the population.
- Resample WITH replacement in order to create many pseudo-samples.
- Treat the pseudo-samples as if they are realizations from the true population.
- Those samples will give you an idea about the true distribution (provided the sample is representative)
- CAREFUL: This works only if we have a representative sample. Sample selection problems abound.....
Example: Suppose we have a sample \( \{x_1, x_2, \ldots, x_n\} \). We believe that there is heteroskedasticity in the sample, but we don’t know how to model it.

- From the sample, we compute the mean \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \).

- Now the question is to find the dispersion around the true value of \( \mu \) in order to conduct tests.

- Draw a sample of \( n \) observations WITH REPLACEMENT from \( \{x_1, x_2, \ldots, x_n\} \). Name this sample \( \{x_1^1, x_2^1, \ldots, x_n^1\} \).

- Draw a second sample of \( n \) observations WITH REPLACEMENT from \( \{x_1, x_2, \ldots, x_n\} \). Name this sample \( \{x_1^2, x_2^2, \ldots, x_n^2\} \).

- Draw \( J \) more samples, where the \( j \)th sample is denoted by \( \{x_1^j, x_2^j, \ldots, x_n^j\}_{j=1}^{J} \).

- For each pseudo-sample, \( j \), we can compute its mean, \( \bar{x}^j = \frac{1}{n} \sum_{i=1}^{n} x_i^j \).

- We can plot the histogram of the means \( \{\bar{x}^j\}_{j=1}^{J} \). This is the empirical distribution of \( \bar{x} \).

- We can form confidence intervals and tests based on the above distribution.
• The same principle can be applied to any statistic. In the previous example, we could have focused on the t-statistic, instead of the mean.

• From the sample, we compute the t statistic \( t = \frac{\bar{x} - 0}{se} \).
  – Now the question is to find the dispersion around the true value of \( t \) in order to conduct tests.
  – Draw a sample of \( n \) observations WITH REPLACEMENT from \( \{x_1, x_2, \ldots, x_n\} \). Name this sample \( \{x_1^1, x_2^1, \ldots, x_n^1\} \).
  – Draw a second sample of \( n \) observations WITH REPLACEMENT from \( \{x_1, x_2, \ldots, x_n\} \). Name this sample \( \{x_1^2, x_2^2, \ldots, x_n^2\} \).
  – Draw \( J \) more samples, where the \( j \text{th} \) sample is denoted by \( \{x_1^j, x_2^j, \ldots, x_n^j\}_{j=1}^J \).
  – For each pseudo-sample, \( j \), we can compute its t statistic, \( t^j = \frac{1}{n} \sum_{i=1}^n x_i^j \).
  – We can plot the histogram of the t statistics \( \{t^j\}_{j=1}^J \). This is the empirical distribution of \( t \).
  – We can form confidence intervals and tests based on the above distribution.
3.1 Bootstrap in OLS–Conditional vs Unconditional Bootstrap

3.1.1 Conditional Bootstrap

• We have the regression
  \[ y_i = \beta x_i + \varepsilon_i \]

• Estimate it using OLS, to obtain \( \hat{\beta} \), its t-statistic, and
  \( \{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n\} \)

• Conditional on the \( x_i \)'s, we can simulate the \( y_i \)'s as follows:
  – Draw with replacement from \( \{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n\} \) J
    samples of \( n \) observations each.
  – Construct the samples as: \( \hat{y}_i^j = \hat{\beta} x_i + \hat{\varepsilon}_i^j \)
  – Note that the randomness comes only from resampling the residuals.
  – For each sample compute \( \hat{\beta}^j \) and/or \( t^j \)
  – Form tests as before.
3.1.2 Unconditional Bootstrap

- In the conditional bootstrap, we held the $x’s$ fixed. In other words, we were really bootstrapping the marginal distribution of $y|x$.

- Again, suppose we have

$$y_i = \beta x_i + \varepsilon_i$$

- Estimate it using OLS, to obtain $\hat{\beta}$, its t-statistic, and $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n\}$

- We also have $\{x_1, x_2, \ldots, x_n\}$ which are independent of $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n\}$ by construction

- Therefore, we can simulate the $y_i’s$ as follows:
  - Draw with replacement from $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n\}$ $J$ samples of $n$ observations each.
  - Draw with replacement from $\{x_1, x_2, \ldots, x_n\}$ $J$ samples of $n$ observations each.
  - Construct the samples as: $y_i^j = \hat{\beta} x_i^j + \varepsilon_i^j$
  - Note that the randomness comes not only from the residuals but also from the explanatory variables.
  - For each sample compute $\hat{\beta}^j$ and/or $t^j$
• The unconditional bootstrap test is more conservative.

• Test for spurious correlation.

• CAUTION: The bootstrap does not work if the data is serially dependent. Why?
  – Block bootstrap: Works pretty well, depending how dependent the data is.
3.2 Bootstrap in GMM context

3.2.1 Conditional

- Recall the moment condition
  \[ E \left[ \left( 1 - \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + R_{t+1}) \right) x_t \right] = 0 \]

- Draw with replacement from \( \{c_1, c_2, \ldots\} \) and \( \{R_1, R_2, \ldots\} \) \( J \) samples of \( T \) observations each. Do not sample independently.

- Condition on the \( x_t \).

- Run the GMM for each sample, producing estimates \( \{\hat{\gamma}^1, \hat{\gamma}^2, \ldots, \hat{\gamma}^J\} \).

- Plot the empirical distribution and test if the estimate \( \hat{\gamma} \) is in the tails.
3.2.2 Unconditional

- Follow the same procedure, but also sample $\{x_t\}_{t=1}^T$ independently from the other two series.
- Form the same tests.
- This test can have the interpretation to test for spurious correlation.
3.3 Simulations

- Not to be confused with bootstrap.
- Neither better nor worse, just different.
- We have the regression
  \[ y_i = \beta x_i + \varepsilon_i \]
- Estimate it using OLS, to obtain \( \hat{\beta} \) and its t-statistic.
- Assume that the errors \( \varepsilon_i \) have a certain distribution, say \( \varepsilon_i \sim T(15) \)
- Conditional on the \( x_i \)'s, we can simulate the \( y_i \)'s as follows:
  - Simulate from, say, \( T(15) \), J samples of \( n \) observations each.
  - Construct the samples as: \( y_i^j = \hat{\beta} x_i + \varepsilon_i^j \)
  - Note that the randomness comes only from the simulated residuals.
  - For each sample compute \( \hat{\beta}^j \) and/or \( t^j \)
  - Form tests as before.
We can also assume dependence between the residuals, as in $\varepsilon_i = \phi \varepsilon_{i-1} + \nu_t$, where $\nu_t$ is NID(0,1).

We can also simulate the $x's$.

IMPORTANT: In the simulations, we have to make crucial assumptions about the underlying distribution of the $\varepsilon's$, the very assumption we want to circumvent in the bootstrap exercise.

IMPORTANT: The simulations exercise can handle dependent data.
Filtering is about transforming a time series with certain “characteristics” or properties into another time series with different characteristics.

The concept of filtering is very general and we will see it now in the time domain and later in the frequency domain.

Filtering is best explained by examples.

Define the lag operator $L$ to be such that, for a time series $X_t$,

$$LX_t = X_{t-1}$$

(sometimes you will see $L$ denoted as $B$, which people call the “backward” operator)

The lag operator will be treated as a “number”, i.e. we can define all the usual math operations on it, such as addition, multiplication, inversion, etc.

For example, we can write

$$L^2 X_t = LLLX_t = X_{t-2}$$

$$\frac{X_t}{(1 - L)} = \sum_{j=0}^{\infty} X_{t-j} L^j$$
• Example: We can represent the familiar AR(1) process \( Y_t = \phi Y_{t-1} + \varepsilon_t \) as:
  \[
  Y_t = \phi LY_t + \varepsilon_t
  \]
  or
  \[
  Y_t - \phi LY_t = \varepsilon_t \\
  (1 - \phi L) Y_t = \varepsilon_t
  \]

• Example: We can represent an AR(2) process \( Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \) as:
  \[
  Y_t = \phi_1 LY_t + \phi_2 L^2 Y_t + \varepsilon_t
  \]
  or
  \[
  Y_t - \phi_1 LY_t - \phi_2 L^2 Y_t = \varepsilon_t \\
  (1 - \phi_1 L - \phi_2 L^2) Y_t = \varepsilon_t
  \]
• In the AR(1) case, suppose that $\phi = 0.95$. Then the process is persistent.

• But suppose we have an AR(2) process with $\phi_1 = 0.95$ and $\phi_2 = -0.2$.

• Q: Is the process persistent?
• A:

• Q: What exactly do we mean by a “persistent” process?
• A: We want to know:  

$$ \frac{\partial y_{t+k}}{\partial \varepsilon_t} $$

• In other words, if there is a shock, or “news,” how long would the news impact the process and to what extent.

• In the AR(1) case,

$$ y_{t+k} = \phi y_{t+k-1} + \varepsilon_{t+k} $$

$$ = \varepsilon_{t+k} + \ldots + \phi^k \varepsilon_t + \phi^{k+1} y_{t-k-1} $$

$$ \frac{\partial y_{t+k}}{\partial \varepsilon_t} = \phi^k $$

• $\frac{\partial y_{t+k}}{\partial \varepsilon_t}$ (as a function of $k$) is known as an “impulse response” function.

• But what about in the AR(2) or AR(k) case?
• We can write an AR(2) process as a vector AR(1) process:

\[
\begin{bmatrix}
    y_t \\
    y_{t-1}
\end{bmatrix} =
\begin{bmatrix}
    \phi_1 & \phi_2 \\
    1 & 0
\end{bmatrix}
\begin{bmatrix}
    y_{t-1} \\
    y_{t-2}
\end{bmatrix} +
\begin{bmatrix}
    \varepsilon_t \\
    0
\end{bmatrix}
\]

\( Y_t = \Phi Y_{t-1} + \varepsilon_t \)

• Then the dynamics of \( Y_t \) are driven by \( \Phi \), because

\( Y_{t+k} = \varepsilon_{t+k} + \ldots + \Phi^k \varepsilon_t + \Phi^{k+1} Y_{t-1} \)

• So, we have to characterize the behavior of \( \Phi^k \).

• But \( \Phi \) is a matrix and everything can get pretty messy.
• But there is an easy way out. Write:
  \[ \Phi = T \Lambda T^{-1} \]
  where \( T \) is the matrix of eigenvectors and \( \Lambda \) is a diagonal matrix with the eigenvalues along the diagonal.

• Then:
  \[ \Phi^2 = \Phi \Phi = T \Lambda T^{-1} T \Lambda T^{-1} = T \Lambda^2 T^{-1} \]

• Then:
  \[ \Phi^k = T \Lambda^k T^{-1} \]

• So the eigenvalues in \( \Lambda \) will govern the behavior of \( \Phi \). In particular, the highest eigenvalue of \( \Lambda \) will be the most important to consider.
• Recall that the eigenvalues of $\Phi$ (in the AR(2) case) are the solution to:

$$\begin{vmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{vmatrix} - \lambda \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 0$$

• Or:

$$\lambda^2 - \phi_1 \lambda - \phi_2 = 0$$
We can view \((1 - \phi_1 L - \phi_2 L^2)\) as a polynomial in the lag operator. Therefore, we can look for roots, or numbers \(\lambda_1\) and \(\lambda_2\) such that
\[
(1 - \phi_1 L - \phi_2 L^2) = (1 - \lambda_1 L) (1 - \lambda_2 L) = 0
\]
\[
(1 - (\lambda_1 + \lambda_2) L + \lambda_1 \lambda_2 L^2) = 0
\]
Therefore, matching powers of \(L\), we can find the roots by solving the equations \((\lambda_1 + \lambda_2) = \phi_1\) and \(\lambda_1 \lambda_2 = -\phi_2\).
• Example: Suppose $\phi_1 = 0.6$ and $\phi_2 = -0.08$, so that $Y_t = 0.6LY_t - 0.08L^2Y_t + \varepsilon_t$. You can check that 
\[(1 - 0.6L + 0.08L^2) = (1 - 0.4L)(1 - 0.2L)\]
• But we want to be able to find $\lambda_1$ and $\lambda_2$ in general situations. How to do that?
• Think of a generic polynomial in $z$, (we use $z$ because the $L$ operator has a particular meaning, i.e. it is the lag operator), as above 
\[(1 - \phi_1z - \phi_2z^2) = (1 - \lambda_1z)(1 - \lambda_2z)\]
• Now, we ask, what values of $z$ will set the right hand side to zero?
• The answer is: $z = 1/\lambda_1$ or $z = 1/\lambda_2$.
• But why is it important?
• Well, the same $z$ must also set the left hand side to zero.
• But we also know that the $z$ that sets $\lambda_1z + 4\phi_2z \quad 0$ can be found by 
\[z_{1,2} = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}\]
• In other words, $z_1 = 1/\lambda_1$ and $z_2 = 1/\lambda_2$. 
• So, can we always find the $\lambda'$s from the $z'$s by inverting?

• Yes, except for the case when $\phi_1^2 + 4\phi_2 < 0$, because then we will have trouble evaluating $\sqrt{\phi_1^2 + 4\phi_2}$.

• In such a case $z_1$ and $z_2$ are complex numbers (and complex conjugates).

• We can write $z_1 = a + bi$, where $i = \sqrt{-1}$. In our example, you can check that $a = \phi_1/2$ and $b = \frac{1}{2} \sqrt{-\phi_1^2 - 4\phi_2}$.

• We can write the complex number $z$ in the polar coordinate as

$$z_1 = R[\cos \theta + i \sin \theta]$$

where

$$R = \sqrt{a^2 + b^2}$$

$$\cos \theta = a/R$$

$$\sin \theta = b/R$$

• Now, we will use DeMoivre’s theorem to write

$$z_1 = R[\cos \theta + i \sin \theta] = R.e^{i\theta}$$

• Similarly

$$z_2 = R[\cos \theta - i \sin \theta] = R.e^{-i\theta}$$
Therefore
\[ \lambda_1 = z_1^{-1} = R^{-1}e^{-i\theta} \]
\[ \lambda_2 = z_2^{-1} = R^{-1}e^{i\theta} \]

The possibility of having complex roots is really cool. We can see that an AR(2) process with complex roots can generate a process that has a very pronounced cyclical component (sin and cos behavior).

Q: Can we have a cyclical behavior in an AR(1) process?

We will come back to those same ideas when we get into the frequency domain representation of a series.
As a final remark, the lag operator is extremely useful to manipulate time series. For instance, we can find the mean and the variance of an AR(1) process simply as:

\[ Y_t = c + \phi LY_t + \varepsilon_t \]

Hence, denoting \( E(Y_t) = \mu \), we obtain

\[
\begin{align*}
\mu & = c + \phi L \mu \\
\mu & = c + \phi \mu \\
\mu & = \frac{c}{1 - \phi}
\end{align*}
\]

Similarly for the variance. Denoting the variance \( Var(Y_t) = \gamma \), we can write

\[
\begin{align*}
\gamma & = \phi^2 L^2 \gamma + \sigma^2 \\
\gamma - \phi^2 \gamma & = \sigma^2 \\
\gamma & = \frac{\sigma^2}{1 - \phi^2}
\end{align*}
\]
Now that we are comfortable with the lag operator and polynomials in the lag operator, here is a filtering example:

Suppose that $\varepsilon_t$ is an i.i.d. series with variance $\text{Var}(\varepsilon_t) = \sigma^2$.

We are going to filter, or transform the series $\varepsilon_t$ into another series, $X_t$.

First, define the filter

$$F(L) = (1 - \theta L)$$

Second, we obtain the $X_t$ series as

$$X_t = F(L) \varepsilon_t$$
$$= (1 - \theta L) \varepsilon_t$$
$$= \varepsilon_t - \theta \varepsilon_{t-1}$$

The series $X_t$ is called a moving average of order 1, or MA(1), process.
- Note that the properties of $X_t$ are different from those of $\varepsilon_t$
  - $\text{Var}(\varepsilon_t) = \sigma^2$, $\text{Cov}(\varepsilon_t\varepsilon_{t-k}) = 0$
  - $\text{Var}(X_t) = \sigma (1 + \theta^2) > \text{Var}(\varepsilon_t)$
  - $\text{Cov}(X_tX_{t-1}) = \text{Cov}([\varepsilon_t - \theta\varepsilon_{t-1}][\varepsilon_{t-1} - \theta\varepsilon_{t-2}]) = -\theta\sigma^2$
  - $\text{Cov}(X_tX_{t-j}) = 0$, $j > 1$.

- So, the filtered series $X_t$ induces a slight (one period) serial correlation.
• Suppose we define another filter, \( G(L) = F(L)^{-1} = 1/(1 - \theta L) \)

• We will filter \( \varepsilon_t \) to obtain a series \( Y_t \), such that
  \[ Y_t = G(L)\varepsilon_t \]

• Note that \( X_t \) and \( Y_t \) are obtained from the same \( \{\varepsilon_t\} \) process, but by applying different filters.

• We will investigate the properties of \( Y_t \)
  \[ Y_t = \frac{\varepsilon_t}{(1 - \theta L)} \]
  \[ Y_t (1 - \theta L) = \varepsilon_t \]
  \[ Y_t = \theta Y_{t-1} + \varepsilon_t \]

• \( Y_t \) is our familiar AR(1) process. We know that
  - \( \text{Var} (Y_t) = \frac{\sigma^2}{1 - \theta^2} \)
  - \( \text{Cov} (Y_t Y_{t-k}) = \frac{\sigma^2 \theta^k}{1 - \theta^2} \)

• Therefore, the filter \( G(L) \) induces a persistence that lasts longer than only one period.
• The $H(L) = (1 - L)$ filter.

• Recall that if we have a series that is non-stationary, my advice was to take first-difference, and the new series would be stationary. We will illustrate this advice with a (quite general) example.

• Suppose we have a series, $Y_t$, composed of a stationary and a non-stationary series:

\[
Y_t = Z_t + X_t \\
Z_t = Z_{t-1} + e_t \\
X_t = \phi X_{t-1} + u_t, \quad |\phi| < 1
\]

• Note that we can rewrite $Z_t (1 - L) = e_t$ and $X_t (1 - \phi L) = u_t$. 
Therefore, we can write
\[ Y_t = Z_t + X_t \]
\[ Y_t = \frac{e_t}{1 - L} + \frac{u_t}{1 - \phi L} \]
\[ (1 - L)(1 - \phi L)Y_t = e_t (1 - \phi L) + u_t (1 - L) \]
\[ (1 - L)(1 - \phi L)Y_t = \eta_t \]

But, from here we can immediately see that one of the roots of the polynomial in \( L \) is 1. Therefore, this process will be non-stationary.

Q: Why don’t we like non-stationary processes?
• We will stationarize $Y_t$, by filtering it using $H(L)$.

• Let’s call the filtered series $\tilde{Y}_t = H(L)Y_t$

\[
\tilde{Y}_t = (1 - L) \left[ \frac{e_t}{1 - L} + \frac{u_t}{1 - \phi L} \right]
\]

\[
= e_t + \frac{u_t (1 - L)}{1 - \phi L}
\]

\[
= e_t + \frac{u_t}{1 - \phi L} - \frac{u_{t-1}}{1 - \phi L}
\]

\[
= stationary + stationary + stationary
\]

\[
= stationary
\]

• Therefore, the filter $H(L)$ transformed the non-stationary series $Y_t$ into a stationary series $\tilde{Y}_t$. 
To summarize:
- We have defined the lag operator $L$
- Using the lag operator, we can define a filter function $F(L)$
- $F(L)$ modifies the properties of a series $Y_t$
- It is useful to modify the properties of $Y_t$ (think stationarity, seasonality, etc.)

Filtering is used everywhere
- Stereos
- Cell phones
- Search for extra-terrestrials
5 Examples of filters

• We have already given a few examples of familiar filters. Here we discuss some more:

• Intuitively, we know that when we take a moving average of a series, we effectively smooth the series, i.e. we iron out any non-linearities.

• Example:
In the above example, we have created a series 
\[ Y_t = Z_t + X_t. \]

- \[ Z_t = 0.9Z_{t-1} + e_t \] and \[ X_t = 0.1X_{t-1} + u_t. \]

- We think of \( Z_t \) as the strong "signal" and \( X_t \) is the almost-iid noise.

- In other words, \( Z_t \) is responsible for the big swings in \( Y_t \), whereas the jagged peaks are due to \( X_t \).
We want to smooth out the noise and be left with the signal.

We can do that, as in the figure, by applying a filter 
\[ F(L) = \frac{1}{3}L + \frac{1}{3}L^0 + \frac{1}{3}L^{-1} \] to \( Y_t \).

The filtered series is:
\[
\tilde{Y}_t = F(L)Y_t \\
= \left( \frac{1}{3}L + \frac{1}{3}L^0 + \frac{1}{3}L^{-1} \right)Y_t \\
= \frac{1}{3}Y_{t-1} + \frac{1}{3}Y_t + \frac{1}{3}Y_{t+1}
\]

This is nothing but a moving average of \( Y_t \) and its two adjacent values.

This is a centered filter...Not very useful for forecasting, but good for analytical work.

In a similar fashion, we can smooth the series even more by taking a longer moving average.

Example: 
\[ F(L) = \frac{1}{7}L^3 + \frac{1}{7}L^2 + \frac{1}{7}L + \frac{1}{7}L^0 + \frac{1}{7}L^{-1} + \frac{1}{7}L^{-2} + \frac{1}{7}L^{-3} \]
• The second filter, the MA(7) filter induces an even smoother behavior in the filtered series.

• So, by filtering the series, we have effectively removed the components of \( Y_t \) that are moving fast, or that induce the mostly unpredictable behavior in \( Y_t \).

• If you have used some of the financial websites (Yahoo!), you know that they offer the option of displaying a price using a MA filter.
But \( F(L) = \frac{1}{i}L^3 + \frac{1}{i}L^2 + \frac{1}{i}L + \frac{1}{i}L^0 + \frac{1}{i}L^{-1} + \frac{1}{i}L^{-2} + \frac{1}{i}L^{-3} \)

is a special filter. It is:
- Two sided
- Symmetric
- Equally-weighted

The design of the filter depends on the application.

We also have to understand the properties of the filters:
- Amplitude
- Phase
• The MA filter is the single most widely used filter in applied work (justifiably so, since it is simple and does the job)

• Riskmetrics (forecasts of market and credit risk measures)

• However, suppose we want to filter out a specific feature of the data.
  – For example, corporate taxes are due quarterly. Therefore, any series that is tightly related to taxes will have a quarterly-frequency component to it. In other words, every time taxes are due, the series will exhibit a certain pattern.
  – We might not want to have this quarterly pattern in taxes influence our results. In other words, we want to seasonally adjust (or filter) the series so that quarterly fluctuations are not part of the new series. We can do that with a quarterly filter.
  – There are prices, such as prices of commodities, prices of utilities that have a seasonal component. An entire industry of analysts is trying to forecast such fluctuations. For example, it is widely believed that the price of crude (and refined) oil has a 6 months cycle, peaking in the summer (travel season) and the winter (heating).
– Although this is usually (historically) true, we should not forget that we are dealing with random variables. For example, last year, the price of oil DECREASED during the summer months.
• So, the question is, how can we isolate the features of the data that happen every 3 or every 6 months?
• Can we decompose a series into periodic components, where each component has different periodicity?
• Yes, but for that we will need a great deal of tools (next time).
6 Bayesian Estimation

- Premise: We are not estimating parameter values, but rather always updating and sharpening our subjective beliefs about the state of the world.

- The centerpiece of the Bayesian methodology is Bayes’ theorem:

\[
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}
\]

- In terms of the applications at hand, we are interested in the value of the parameters, given the data.

- We can write

\[
P(\text{Parameters}|\text{Data}) = \frac{P(\text{Data}|\text{Parameters})P(\text{Parameters})}{P(\text{Data})}
\]
We will view the data as not depending on the parameters, or we can further write

\[ P(\text{Parameters}|\text{Data}) \propto P(\text{Data}|\text{Parameters}) \, P(\text{Parameters}) \]

- The terms are:
  - \( P(\text{Data}|\text{Parameters}) \) : Density of the data, given the parameters
  - \( P(\text{Parameters}) \) : Prior density of the parameters—Prior belief of the econometrician
  - \( P(\text{Parameters}|\text{Data}) \) : Posterior density of the parameters, given the data—It is a mixture of the prior and the “current information” from the data.

- We must specify the prior and the form of the data in order to get the posterior.

- Once we get more data, the posterior becomes the prior and we update again.
The calculations involved in Bayesian analysis might become quite burdensome, but here we will present a simple example.

Suppose that the data is
\[ f(y|\beta, \sigma, X) = N(X\beta, \sigma^2) \]

We also assume an “uninformative” prior
\[ f(\text{Parameters}) \propto \text{constant} \]

For simplicity, we assume for now that \( \sigma \) is known.

We can write
\[
\begin{align*}
    y - X\beta &= y - Xb - X(\beta - b) \\
    b &= (X'X)^{-1} X'y
\end{align*}
\]
Then, the posterior is
\[
f(\beta|\beta, \sigma, X) = h(\sigma^2) \times (2\pi \sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \right\} \\
= (2\pi \sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (y - Xb)'(y - Xb) \right\} \\
- \frac{1}{2\sigma^2} (\beta - b)'X'X(\beta - b) \\
\propto \sigma^2 \left| (X'X)^{-1} \right|^{-1/2} \exp \left\{ -\frac{1}{2} (\beta - b)' \left( \sigma^2 X'X \right)^{-1} (\beta - b) \right\}
\]

- In other words, the density is normal with mean $\beta$ and variance matrix $\sigma^2 (X'X)^{-1}$.

- We can do the similar calculations when we impose another prior on $\sigma$.

- The results would change. The distribution of $\beta$ would be a multivariate-t which is another familiar result.

- Note: We had to specify the priors and the distribution of the data. If we change any of these two assumptions, the results would change.

- Note: We obtain exact results, because we make distributional assumptions on the data.

- We did not need such assumptions in the frequentist world. Why?