1 Today’s Agenda

1. Next time: Review (2 hrs.) + Final (2 hrs.)
   Today:

2. Finish Spectral Representation

3. Model Testing and Misspecification
   (a) Testing
   (b) Joint Tests
   (c) Breaks (using OLS and GMM)

4. Bootstrap and Simulations

5. Non-Stationary Time Series
   (a) “Very persistent” series (prices, volume, volatility?)
   (b) Augmented Dickey-Fuller test
   (c) Dangers of spurious correlation
   (d) Cointegration
2 Finish Spectral Representation

• In the time domain, we can decompose a series $Y_t$ as

$$Y_t = \mu + \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}$$

where $\theta_j$ are parameters and $\{\varepsilon_t\}$ is a sequence of white noise, or the error that one would make in the forecasting of $Y_t$, using past $Y_t$'s.

• In the frequency domain, we can decompose $Y_t$ as

$$Y_t = \int_0^\infty \alpha(\omega) \cos(\omega t) d\omega + \int_0^\infty \beta(\omega) \sin(\omega t) d\omega$$
• Recall that for a covariance stationary random variable $Y_t$ with autocovariances $\rho(h)$, we have the autocovariance generating function:

$$g_y(z) = \sum_{h=-\infty}^{\infty} \rho(h) z^h$$

• Second, the Fourier transform of $\rho(h)$ is

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \rho(h) e^{-i\omega h}$$

$$= \frac{1}{2\pi} g_y(e^{-i\omega})$$

$$= \frac{1}{2\pi} \left[ \rho(0) + 2 \sum_{h=1}^{\infty} \rho(h) \cos(\omega h) \right]$$
CONCLUSION:

- If we know the autocovariances (and thus the autocovariance generating function), we can calculate $f_x(\omega)$ for any value of $\omega$.

\[
f_x(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \rho(h) e^{-i\omega h}
\]

- IF we know $f_x(\omega)$ for any value of $\omega$, we can find all the autocovariances of the process. This follows because $f(\omega)$ and $\rho(h)$ are Fourier pairs, and we can write

\[
\rho(h) = \int f_x(\omega) e^{i\omega h} d\omega
\]

\[
= \int f_x(\omega) \cos(\omega h) d\omega
\]

- Thus, the spectrum and the autocovariances contain EXACTLY the same information about the time series process.
Q: OK, but why is the spectrum useful? We have $\rho(h)$ and it is quite intuitive

Note: For $h = 0$, we have

$$\rho(0) = \text{Var}(X_t) = \int_{-\pi}^{\pi} f_x(\omega) d\omega$$

In other words, the variance of $X_t$ is the area under the spectrum between $-\pi$ and $\pi$.

More generally,

$$\int_{-s}^{s} f_x(\omega) d\omega$$

is the fraction of the variance of $X_t$ between frequencies $-s$ and $s$!
3 Model Testing and Misspecification

- Thus far, we have done the following things:
  - Specify a model
    \[ y_t = f(x_t | \theta) + \varepsilon_t \]
  - Estimate \( \theta \) using various methods.
  - Test simple hypotheses, i.e. whether single elements of \( \theta \) are equal to a certain value.
\[ y_t = f(x_t|\theta) + \varepsilon_t \]

- The world is WAY more complicated
- List of things we might have done wrong:
  - Misspecification of model: \( g(z_t|\gamma) \) instead of \( f(x_t|\theta) \).
    * “Omitted variables” and “too many variables.”
    * Incorrect functional form.
  - Estimation might be flawed (inconsistent) if \( x_t \) is non-stationary, even if the model is right.
  - The parameters might be time-varying, \( i.e. \)
    \[ \theta_t = \phi \theta_t + u_t. \]
  - The moment, or model, might be time-varying:
    \( f_t(x_t|\theta) \).
  - Or both: \( f_t(x_t|\theta_t) \). Distinguishing variations in \( f_t \) versus variations in \( \theta_t \) is close to hopeless? Why?
  - The least of our worries: \( \varepsilon_t \) is not Niid. Why?
Q: Which is worse: omitting relevant information from $f$ or including too many variables?

A: Suppose the true model is:

$$y_t = \beta x_t + \phi q_t + \varepsilon_t$$

But we omit $q_t$, and estimate

$$y_t = \beta x_t + u_t$$

Then

$$\hat{\beta}_{ols,mle,gmm} = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{\sum x_t (\beta x_t + \phi q_t + \varepsilon_t)}{\sum x_t^2} = \frac{\beta \sum x_t^2}{\sum x_t^2} + \frac{\phi \sum x_t q_t}{\sum x_t^2} + \frac{\sum x_t \varepsilon_t}{\sum x_t^2} \rightarrow^p \beta + \phi \frac{\text{cov}(x_t, q_t)}{\text{var}(x_t)}$$

Then $\hat{\beta} \rightarrow^p \beta$.

If we have only two variables, the direction of the bias is known. But if we have more than two variables, things get very complicated.
• A: Suppose the true model is $y_t = \beta x_t + \varepsilon_t$,

• But we include an extra variable $q_t$, and estimate $y_t = \beta x_t + \phi q_t + u_t$

• Then

$$E \left( \left[ \begin{array}{c} \hat{\beta} \\ \hat{\phi} \end{array} \right] \right) = \left[ \begin{array}{c} \beta \\ 0 \end{array} \right]$$

• But $VAR(\hat{\beta}) > Var(\beta \text{ estimated without } q_t)$.

• Conclusion: Omitting relevant information seems to be worse than including useless variables. Trade-off between mean and variance.

• Why: Think forecasting?

• In practice: People start with a k-variable regression and start testing sequentially for useless variables.

• Some of the variables might be non-linear:

$$r_t = \beta_1 r_{t-1} + \beta_2 r_{t-1}^2 + \beta_3 r_{t-1}^3 + \beta_4 r_{t-1}^4$$

• As long as the function is linear in the parameters, we can estimate with OLS, which is equivalent to GMM (for some moments). Under the further assumption of normal errors, OLS is equivalent to MLE.

• If the residuals are serially correlated or heteroskedastic, then use some of that information.
4 Testing

● General Idea: Test whether our model is true or not.
● We need three concepts:
  – Null hypothesis ($H_0$)
  – Alternative hypothesis ($H_a$)
  – “Ultimate Truth”
● The null hypothesis is not the ultimate truth!
● GOAL: Design a statistic $t(x_t)$ that will help us decide whether or not the data comes from a hypothesized population (specified by the null).

● Since the sample is random, so is the statistic. Different samples, from the same DGP can lead to different conclusions.

● Two types of errors:
  – **Type I error**: Reject the null, when it is true. This is known as the size of the test, or $\alpha$, or the level of significance.
  – **Type II error**: Fail to reject the null, when it is false. This is denoted as $\beta$.

● Note: We can control the level of significance (by changing the rule), but we can’t control type II error.
• The statistic $t(x_t)$ has a certain distribution (think $N(0,1)$), under the null hypothesis.

• We can find out the testing rule as: Reject if the outcome is in the “tails” of the distribution.

• Q: What do we mean by tails?

• A: Depends on the level of significance we want to impose.

• Suppose $\alpha = 0.05$. Then, even if the null is true, if we simulate the DGP under the null 1000 times, we will incorrectly reject the model 50 times.

• Suppose we let $\alpha = 0.01$. Better.

• Level of significance is under control. But there is a trade-off.

• Power($t(x_t)$)=1-Prob(Type II error)=is the probability that $t(x_t)$ will lead to a correct rejection of a false null hypothesis.

• We want to have a test that has as great a power as possible, given the size.

• In other words, we want to maximize power (minimize $\beta$), subject to the size constraint.
• Think of testing in the following way: Does \( t \) come from:
  
  – The distribution suggested by the null (black)?
  – Some other distribution, suggested by the alternative (red, green)?

• Power\( (H_a) \): The further the null is from the alternative (all else equal), the higher the power of the test.
- We can graph the power function, Power($H_a$):
- Example:
  \[ r_t = \phi r_{t-1} + \varepsilon_t \]
- $H_0 : \phi = 0$
- $H_a : \phi \neq 0$ (two-sided test).
- As $\phi$ diverges from 0, it will be “easier” for the test to differentiate between the null and the alternative.
- The t-test is:
  \[ t = \frac{\hat{\phi} - 0}{se(\hat{\phi})} \]
- Asymptotically, $t \rightarrow N(0, 1)$
- But, we can define the following test:
  \[ W = t^2 = \frac{(\hat{\phi} - 0)^2}{var(\hat{\phi})} \]
- How is $W$ distributed, asymptotically? Why?
- Note: Both tests test the same null hypothesis versus the same alternative hypothesis.
- Q: Which test is “better”?
- I.E.: Which test is more powerful?
- A: ....
Why is this important?

Suppose the true process is

\[ r_t = \phi r_{t-1} + \varepsilon_t \]

where \( \varepsilon_t \) is i.i.d. with \( E(\varepsilon_t^2) = \sigma^2 \).

Suppose that \( \sigma^2 \) is very large. The variance of \( \hat{\phi} \) is very large.

It will be hard to distinguish the null hypothesis from the alternative.

The more noise there is in the system, the less powerful the test will be.

Whenever we have returns, there is a lot of noise (as there should be if markets are to be close to efficient). Therefore, in a forecasting relation, we will not have power to distinguish the null from an alternative, especially a close alternative.
Joint Hypothesis Testing

- Suppose we have the regression:
  \[ r_t = \beta_1 r_{t-1} + \beta_2 r_{t-1}^2 + \beta_3 r_{t-1}^3 + \beta_4 r_{t-1}^4 + \varepsilon_t \]
  \[ = x_{t-1}' \beta + \varepsilon_t \]

- Suppose you want to test the hypothesis that the last two variables are equal, i.e. that skewness and kurtosis have the same effect on future returns.

- You can do two \( t \)-tests: \( \beta_3 = 0 \) and \( \beta_4 = 0 \). Not quite our hypothesis.

- Problem: Suppose \( r_{t-1}^3 \) and \( r_{t-1}^4 \) are correlated. Then the two \( t \)-tests will be correlated? Why?

- So, the tests cannot be done independently, without taking into account the joint distribution of the variables.

- We want to test: \( \beta_3 = \beta_4 \) jointly!
• Recall that
\[
\hat{\beta} \sim N \left( \beta, \sigma^2 \left( \sum x_{t-1}x'_{t-1} \right)^{-1} \right)
\]
\[
\sim N \left( \beta, V_{\hat{\beta}} \right)
\]
• Let’s write: \( R = [0 \ 0 \ 1 \ -1], q = 0 \)
• Then, the null is
\[
R\beta = q
\]
\[
\begin{bmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4
\end{bmatrix}
= 0
\]
\[
\beta_3 = \beta_4
\]
• Now, if the null is true
\[
0 \approx R\hat{\beta} - q = d
\]
\[
E (d) = R\beta - q = 0
\]
\[
Var (d) = Var \left( R\hat{\beta} \right) = RVar \left( \hat{\beta} \right) R'
\]
\[
d \sim N \left( R\beta, \sigma^2 R \left( \sum x_{t-1}x'_{t-1} \right)^{-1} R' \right)
• Then
\[ d'Var (d)^{-1} d \sim \chi^2 (1) \]

• But we have to estimate \( \sigma^2 \).
\[
F = \frac{\left( R\hat{\beta} - q \right)' \left[ R \left( \sum x_{t-1}x'_{t-1} \right)^{-1} R' \right]^{-1} \left( R\hat{\beta} - q \right) / 1}{\sum e_i^2 / (T - 4)} \sim F (1, \ldots).
\]
Another Example: Suppose we want to test \( \beta_3 = \beta_4 = 0 \). We specify
\[
R = \begin{bmatrix}
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
q = \begin{bmatrix}
0 \\
0
\end{bmatrix}
\]

Then, we can form the F-test
\[
F = \frac{(R\hat{\beta} - q)' \left[ R \left( \sum x_{t-1}x_{t-1}' \right)^{-1} R' \right]^{-1} (R\hat{\beta} - q) / 2}{\sum e_t^2 / (T - 4)} \sim F(2, T - 4)
\]

In general,
\[
R_{J \times K} \beta = q_{K \times 1}
\]
where we have \( J \) restrictions, \( K \) parameters. Then, we can use
\[
F \sim F(J, T - K)
\]
• Note that so far, we have looked at “nested” models.
• Here is another way of testing (equivalent) the hypothesis $\beta_3 = \beta_4 = 0$:
  
  – First, run the unrestricted regression:
    \[ r_t = \hat{\beta}_1 r_{t-1} + \hat{\beta}_2 r_{t-1}^2 + \hat{\beta}_3 r_{t-1}^3 + \hat{\beta}_4 r_{t-1}^4 + \hat{\epsilon}_t^u \]
  
  – Second, run the restricted regression (imposing $\beta_3 = \beta_4 = 0$):
    \[ r_t = \hat{\beta}_1 r_{t-1} + \hat{\beta}_2 r_{t-1}^2 + \hat{\epsilon}_t^r \]

• If the restrictions are true, we are not going to lose too much in fit by suppressing them.
• We should find a way to compare the residuals.
• It turns out that:
  \[ F = \frac{(\hat{\epsilon}_t^r \hat{\epsilon}_t^r - \hat{\epsilon}_t^u \hat{\epsilon}_t^u) / J}{\hat{\epsilon}_t^u \hat{\epsilon}_t^u / (T - K)} \sim F(J, T - K) \]

• Note: It can be proven that the two ways of computing the $F$ test are (numerically) equivalent.
5 Parameter Change

- Suppose that from $t = 1, ..., T_1$, we have the DGP:
  \[ y_t = \beta X_t + \varepsilon_t \]
- and from $T_1 + 1, ..., T$, we have a sudden change in parameters and the DGP is:
  \[ y_t = \gamma X_t + u_t \]

- First, is this plausible?
- Yes? Lucas (1972) Critique. People are not atoms. They might change their behavior.

- How do we deal with this problem:
  \[
  \begin{bmatrix}
  y_t \\
  y_t
  \end{bmatrix}
  =
  \begin{bmatrix}
  X_t & 0 \\
  0 & X_t
  \end{bmatrix}
  \begin{bmatrix}
  \beta \\
  \gamma
  \end{bmatrix}
  +
  \begin{bmatrix}
  \varepsilon_t \\
  u_t
  \end{bmatrix}
  \]

- This is nothing but two separate regressions. So we can run them separately. This is the unrestricted model with $\hat{\beta}^u_t = \hat{\varepsilon}_t + \hat{u}_t$.

- The null hypothesis is that there is no change between the two samples, or $\gamma = \beta$. 
• The restricted model is:
\[
\begin{bmatrix}
  y_t \\
  y_t
\end{bmatrix} =
\begin{bmatrix}
  X_t \\
  X_t
\end{bmatrix}
\begin{bmatrix}
  \beta \\
  \beta
\end{bmatrix} +
\begin{bmatrix}
  e_t \\
  e_t
\end{bmatrix}
\]
which is nothing but estimating the regression over the entire sample, \( t = 1, ..., T \).

• The restricted SS is:
\[
\hat{\varepsilon}_t^r = \hat{\varepsilon}_t
\]

• Therefore, we can use the F-test above to test the restriction.

• There are many ways of specifying breaks in the parameters.
• But this is the simplest way of modelling time-variation in the model.

• Suppose that from \( t = 1, ..., T_1 \), we have the DGP:
  \[
y_t = \beta X_t + \varepsilon_t
  \]

• and from \( T_1 + 1, ..., T \), we have a sudden change in the DGP as:
  \[
y_t = \gamma X_t + \eta Z_t + \delta Z_t^{1-\alpha} + u_t
  \]

• We have a change not only in the parameters (from \( \beta \) to \( \gamma \)), but also in the functional form.

• The problem is even worse. Often one would detect an instability in the parameters, but it would be attributed only to a parameter change (previous example).

• People have a tendency to cling on to a particular model.

• But the model changes (Think APT with factors coming in and out and in different forms!)

• Is there hope of dealing with such uncertainty?
• GMM might offer a way of differentiating between parameter variation and model variation.
• As a starting point, we can break the orthogonality restrictions into those that identify and those that over-identify the parameters
  \[ E \left( \begin{array}{c} h_1 (w_t; \theta_0) \\ h_2 (w_t; \theta_0) \end{array} \right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \]
• Some people have suggested to see how the estimates would change as we add more restrictions to \( h_2 \) (say, starting from no over-identifying restrictions, and adding progressively).
• This set-up has also yielded insights into the stability properties of the moments and (or versus) the estimates.
• Suppose we have $K$ moments and $R$ parameters, where $R < K$.

• Suppose we use the first $K$ moments to estimate the parameters over two subsamples. If the parameters change over the sub-samples, it must be parameter change.

• But if the model is stable, the over-identifying restrictions must be close to 0 in both subsamples. In other words, we must have

\[
E\left(\begin{pmatrix} h_2 \left( w_t; \hat{\theta}_1 \right)_{t=1}^{T_1} \\ h_2 \left( w_t; \hat{\theta}_1 \right)_{t=T_1} \end{pmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\]

• However, if in one sub-sample, the over-identifying restrictions hold, but in the other they do not, this implies that the model has also changed.

• But in practice, it is a question of power. Do we have enough information to differentiate between all hypotheses?
6 Bootstrap

- Sometimes, the normality assumption is not adequate. But we need an adequate assumption in order to have accurate tests.
- In some cases, the CLT might not provide good approximation (it is only an asymptotic result).
- Example: Short rate might be bi-modal.
- Q: What to do?
- IDEA: Treat the sample as if it were the population.
- Resample without 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 \( jth \) sample is denoted by \( \{x_1^j, x_2^j, \ldots, x_n^j\} \) \( j=1 \).
- 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 \( \{\bar{x}^j\} \) \( j=1 \). 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 \( t = \frac{\bar{x} - \mu_0}{se} \).

- 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 \( j^{th} \) 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.
Another Example: Cumulative Abnormal Returns (CAR):

Idea: There is a major event that, we think, will impact the returns of an asset. But the extent of the impact and its horizon is uncertain.

For instance: Mergers, IPO’s, etc.

We want to measure the returns of the asset (or portfolio of assets) that is subject to the event relative to a “normal return”.

What is a normal return?

CAPM or APT will tell us.

We define:

\[
R^i_t - R_f = \alpha + \beta_1 (R^M_t - R_f) + \beta_2 R^2_t + \beta_3 F^3_t + \varepsilon^i_t
\]

Estimate the regression and

\[
AR^i_t = \hat{\varepsilon}^i_t
\]

To look at the cumulative effect of the event, we look at:

\[
CAR^i_{t,t+k} = \sum_{j=1}^{k} \log \left(1 + \hat{\varepsilon}^i_{t+j}\right)
\]

and plot it over time, or as a function of \(k\).

CARs are like prices. We expect a big jump at the beginning and nothing afterward.
- The errors of CARs are notoriously difficult to estimate.
- There are all sorts of biases.
- We usually have a small sample of assets (CLT approx. is not adequate).
- Q: What to do?
- A: Bootstrap the abnormal returns $\{\hat{\varepsilon}_{i+j}^i\}$, since the $\varepsilon_i'$s must be i.i.d.
- CONCLUSION: The bootstrap is a very good way to construct standard errors in complicated situations.
6.1 Bootstrap in OLS–Conditional vs Unconditional Bootstrap

6.1.1 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
  \( \{\hat{\varepsilon}_1, \hat{\varepsilon}_2, ..., \hat{\varepsilon}_n\} \)

- Conditional on the \( x_i \)'s, we can simulate the \( y_i \)'s as follows:
  - Draw with replacement from \( \{\hat{\varepsilon}_1, \hat{\varepsilon}_2, ..., \hat{\varepsilon}_n\} \) \( J \) 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 resampling the residuals.
  - For each sample compute \( \hat{\beta}^j \) and/or \( t^j \)
  - Form tests as before.
6.1.2 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 $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, ..., \hat{\varepsilon}_n\}$

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

- Therefore, we can simulate the $y_i$'s as follows:
  - Draw with replacement from $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, ..., \hat{\varepsilon}_n\}$ $J$ samples of $n$ observations each.
  - Draw with replacement from $\{x_1, x_2, ..., x_n\}$ $J$ samples of $n$ observations each.
  - Construct the samples as: $y^j_i = \hat{\beta} x^j_i + \varepsilon^j_i$
  - 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?
7 Bootstrap in GMM context

7.1 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.
7.2 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.
8 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.
9 Non-Stationary Time Series

• Thus far, we have been looking at stationary time series. We have focused on \( r_t, \sigma_t^2 \), assuming that they are stationary. For some series, this assumption is more tenable than for others.

• But suppose you want to work with non-stationary time-series, i.e. prices, volume, number of investors in a particular fund, number of funds, etc. Those processes are inherently non-stationary.

• Let \( p_t \) be the log-price. We know that

\[
p_t = p_{t-1} + \varepsilon_t
\]

or \( p_t \) is an AR(1) process with an unit-root.

• This process is non-stationary. We cannot apply the CLT.

• But we are still interested in testing the null \( \phi = 1 \) versus \( \phi < 1 \).

• Problem. Under the null, the process is non-stationary.

• Under the alternative, the process is stationary.
• It turns out that (FCLT)
\[
\frac{1}{\sqrt{T}}P_t = \frac{1}{\sqrt{T}} \sum_{s=1}^{t} \varepsilon_s \\
= \frac{1}{\sqrt{T}} \sum_{s=1}^{[rT]} \varepsilon_s \Rightarrow W(r)
\]
where \(W(r)\) is a Brownian motion on \([0, 1]\).

• Q: Can’t we standardize the non-stationary processes by a power of \(T\) in order for them to converge.

• A: Yes.

• Let’s get a “flavor” of how things work:
Recall that
\[ \hat{\phi} = \frac{\sum p_t p_{t-1}}{\sum p_{t-1}^2} = \frac{\sum p_{t-1} (\phi p_{t-1} + \varepsilon_t)}{\sum p_{t-1}^2} \]
\[ = \phi + \frac{\sum p_{t-1} \varepsilon_t}{\sum p_{t-1}^2} \]

- If \( \phi < 1 \), we had
\[ \hat{\phi} = \phi + \frac{1}{T} \frac{\sum \varepsilon_t p_{t-1}}{\sum p_{t-1}^2} \rightarrow_p \phi \]
\[ \sqrt{T} \left( \hat{\phi} - \phi \right) \sim N \left( 0, \sigma_{\hat{\phi}}^2 \right) \]

- But if \( \phi = 1 \), the results do not hold. But
\[ \hat{\phi} = \phi + \frac{\sum \varepsilon_t p_{t-1}}{\sum p_{t-1}^2} \]
\[ T \left( \hat{\phi} - \phi \right) \Rightarrow O_p(1) \]

- In other words, \( \hat{\phi} \) is super-consistent.

Q: But since we don’t know the distribution of \( \hat{\phi} \), can we use this result for testing?

A: Yes, if we simulate the distribution.
• Dickey-Fuller (DF) Test:
  \[ H_0: \phi = 1 \]
  \[ H_a: \phi < 1 \]
• The test is: \[ t = \frac{\hat{\phi} - 1}{se(\hat{\phi})} \]
• Suppose that \( \varepsilon_t \) follows an AR(p) process. The distribution of the DF test is influenced by those parameters. Not good.
• To get rid of those parameters, we run the following regression:
  \[ p_t = \phi p_{t-1} + \zeta_1 \Delta p_{t-1} + \zeta_2 \Delta p_{t-2} + \ldots + \zeta_k \Delta p_{t-k} + \upsilon_t \]
• Then, \[ t = \frac{\hat{\phi} - 1}{se(\hat{\phi})} \]
• This is called the Augmented DF, or ADF test.
• Summary of ADF test–Testing for a unit root:
  - Regress $p_t$ on $p_{t-1}, \Delta p_{t-1}, \Delta p_{t-2}, \ldots, \Delta p_{t=k}$
  - Form: $t = \frac{\hat{\phi} - 1}{se(\hat{\phi})}$
  - Get the critical value from simulations.
• So, is working with non-stationary variables that easy?
• NO:
• Suppose $p^1_t$ and $p^2_t$ are the prices of the same asset traded on two markets. Then, it must be the case that

$$p^1_t = p^2_t$$

• Empirically, this is almost true. We find

$$p^1_t - p^2_t = \varepsilon_t$$

where $\varepsilon_t$ is almost iid, and $E(\varepsilon_t) = 0$.

• How do we take advantage of this?
• Regress:

$$p^1_t = \gamma p^2_t + \varepsilon_t$$

• If

$$\varepsilon_t > 0$$

$$p^1_t > \gamma p^2_t$$

• The asset is “too expensive” in market 1.
• Easy, right?
• Wrong!
• Suppose we have two assets, with prices $p_t$ and $q_t$. One might be tempted to look for arbitrage strategies as in

$$p_t = \gamma q_t + \varepsilon_t$$

• If $\gamma > 0$, there is a relationship, and we can trade.
• No.
• Suppose

$$p_t = p_{t-1} + u_t$$
$$q_t = q_{t-1} + v_t$$
$$\text{cov}(u_t v_t) = 0$$

• Note: The two log-prices represent two independent discrete Brownian motions.
• But:
\[ \hat{\gamma} = \frac{\sum_{t=1}^{T} p_t q_t}{\sum_{t=1}^{T} q_t^2} = \frac{\sum_{t=1}^{T} (\sum_{s=1}^{t} v_s) (\sum_{s=1}^{t} u_s)}{\sum_{t=1}^{T} (\sum_{s=1}^{t} v_s)^2} \]
\[ = \frac{1}{T^2} \sum_{t=1}^{T} (\sum_{s=1}^{t} v_s) (\sum_{s=1}^{t} u_s) = O_p(1) \]

• Similarly:
\[ t = \frac{\hat{\gamma}}{se(\hat{\gamma})} \]

is not consistent.

• The $R^2$ does not converge to 0, as $T \to \infty$.

• Illustration: spurious.m
Co-integration:

• Suppose that $p_t$ and $q_t$ are unit root (integrated) processes, but there is a linear combination of $p_t$ and $q_t$ that is stationary. That is, there exists a vector $\gamma = \begin{bmatrix} 1 & -\gamma_1 \end{bmatrix}$ such that

$$p_t - \gamma q_t = \varepsilon_t$$

and $\varepsilon_t$ is a stationary process. Then, $p_t$ and $q_t$ are said to be cointegrated.

• There are formal tests for cointegration, but they have low power against the alternative. Why? (Think spurious correlation).

• Cointegration is only between contemporaneous variables. I.e. $p_t - \gamma q_t$ is a cointegrating vector. But $\Delta p_t$ is not. However, the latter is also a way of stationarizing the process.

• Cointegration occurs “naturally” in economics. It is dictated by theory.
• Examples:
  – Prices $p_t^1$ and $p_t^2$ of the same asset traded on two markets.
  – Dividend price ratio is cointegrated, or:
    \[ d_t - p_t \]
    must be stationary
  – The long and the short rate must be cointegrated
  – Consumption and GDP must be cointegrated, etc.
  – Lettau & Ludvigson (2001), the “cay” ratio:
    \[ c_t - \gamma_1 a_t - \gamma_2 y_t \]
1.

Quick Illustration: