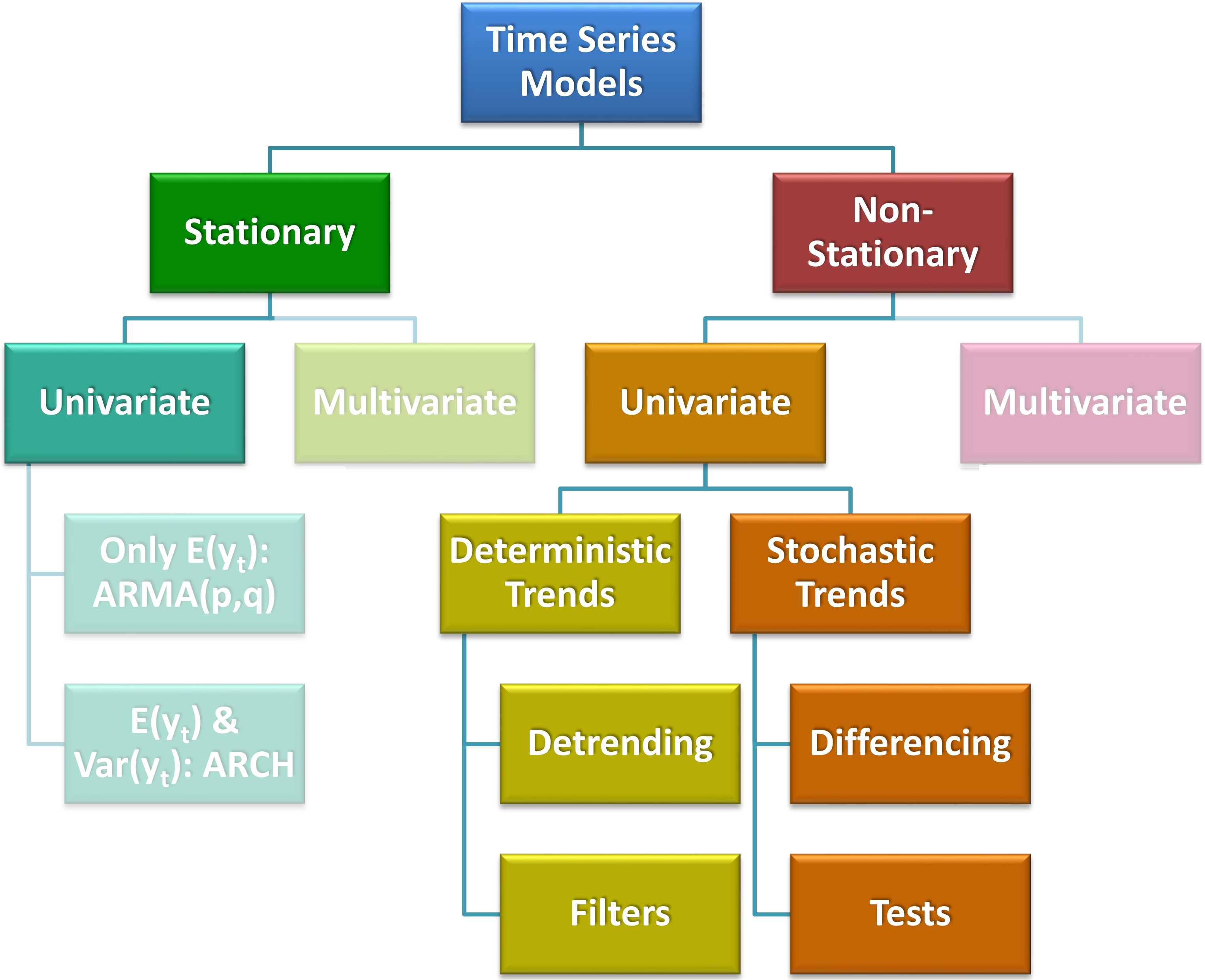
Econometrics 871
Time Series

TOPIC 4

Trends and Unit Roots

Plan

- Examples of the concern
- The two types of non-stationarity
  - Definition
  - Simple Cases
  - Dangers
  - Removing non-stationarity: Detrending, Differencing, Filtering
- Testing for unit roots
  - ACF, PACF
  - Dickey Fuller Tests and Variants


Time Series Models
Stationary
Non-Stationary
Univariate
Multivariate
Only $E(y_t)$: ARMA(p,q)
$E(y_t)$ & $Var(y_t)$: ARCH
Univariate
Multivariate
Deterministic Trends
Stochastic Trends
Detrending
Filters
Differencing
Tests

Definitions

- Weakly stationary process
  - Time independent expectation, variance and covariance
- Trends and/or Unit roots:
  - Cases of time dependent moments

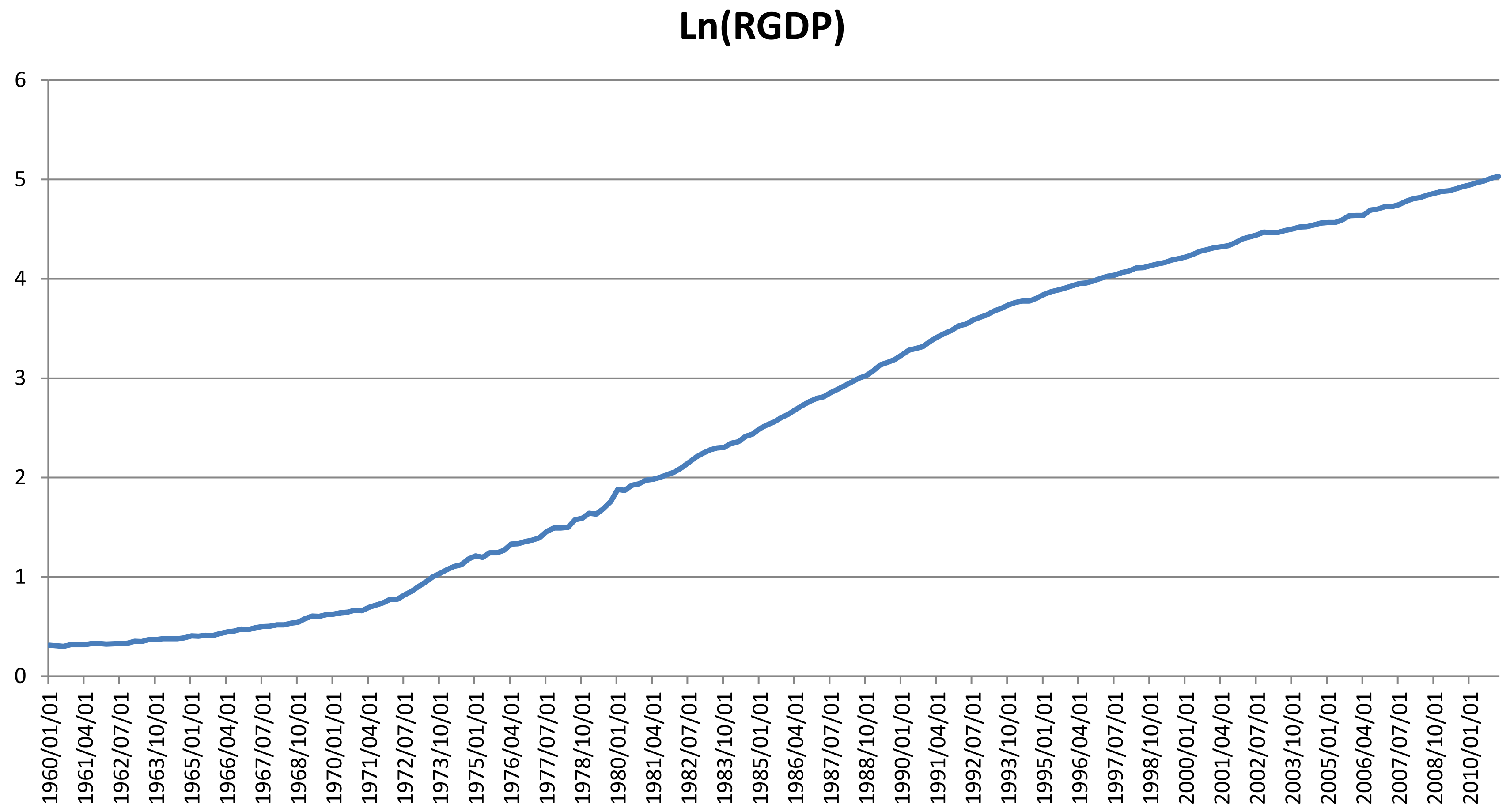
Key Questions:

- How to identify non-stationarity:
  - Specifically, the *type* of non-stationarity
- Given some type of non-stationarity:
  - How to correctly “deal with” the non-stationarity
  - Remove it? Or
  - Model it?

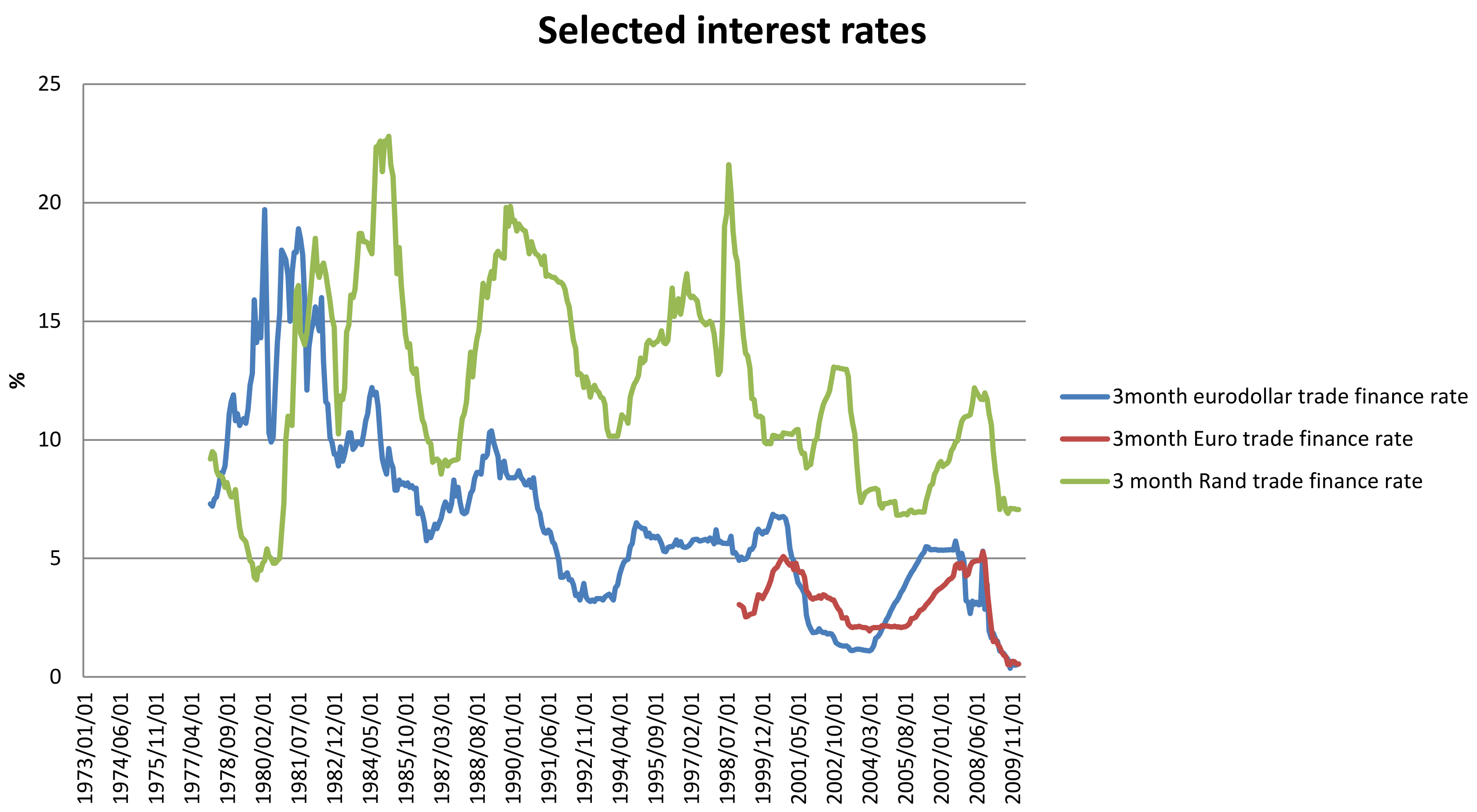
Non-Stationarity

- Many macro series
  - Increase exponentially in levels/linearly in logs
    OR
  - Appear to wander without apparent mean reversion

The Issue


Ln(RGDP)
6
5
4
3
2
1
0
1960/01/01
1961/04/01
1962/07/01
1963/10/01
1965/01/01
1966/04/01
1967/07/01
1968/10/01
1970/01/01
1971/04/01
1972/07/01
1973/10/01
1975/01/01
1976/04/01
1977/07/01
1978/10/01
1980/01/01
1981/04/01
1982/07/01
1983/10/01
1985/01/01
1986/04/01
1987/07/01
1988/10/01
1990/01/01
1991/04/01
1992/07/01
1993/10/01
1995/01/01
1996/04/01
1997/07/01
1998/10/01
2000/01/01
2001/04/01
2002/07/01
2003/10/01
2005/01/01
2006/04/01
2007/07/01
2008/10/01
2010/01/01

The Issue



Selected interest rates
25
20
15
10
5
0
%
1973/01/01
1974/06/01
1975/11/01
1977/04/01
1978/09/01
1980/02/01
1981/07/01
1982/12/01
1984/05/01
1985/10/01
1987/03/01
1988/08/01
1990/01/01
1991/06/01
1992/11/01
1994/04/01
1995/09/01
1997/02/01
1998/07/01
1999/12/01
2001/05/01
2002/10/01
2004/03/01
2005/08/01
2007/01/01
2008/06/01
2009/11/01
3month eurodollar trade finance rate
3month Euro trade finance rate
3 month Rand trade finance rate

Non-Stationarity

- Many macro series
  - Increase exponentially in levels/linearly in logs
    OR
  - Appear to wander without apparent mean reversion

- This has a dramatic impact on the properties of estimators and statistical tests
  - Standard classical results apply to stationary processes *asymptotically*
    - But do very poorly in small samples when series are highly persistent
    - "do poorly" here means that the size of the test is not what is predicted
  - Very different statistical results apply in small and large samples if processes are non-stationary, specifically if they have unit roots

Two Types of Non-Stationarity

- “Trend Stationary”
  - Stationary process around a deterministic trend
  - Trend is deterministic function of time
    - ARMA + f(t)

- “Difference Stationary”
  - Trend is stochastic
    - Random Walk
  - Trend has a stochastic and a deterministic component
    - Random Walk with Drift
    - Random Walk with Drift and Noise
  - ARIMA(p,d,q) with d=1 or 2 (usually)
  - These are called “integrated processes” or processes with a unit root

- Philosophical considerations:
  - What types of non-stationarity make sense as an economic concept?
    - Forever? remember, most hypothesis tests are only valid asymptotically

“Dealing with” the Trend

- “Dealing with” = getting rid of
  - Mostly applicable Univariate situation
  - Strongly related to the univariate tests we will consider
  - Risky
    - Can throw away information
    - Can “create” false information

Methods/Approaches:

- Detrending
- Differencing
- Filtering (later) e.g.
  - Hodrick-Prescott filter
  - Band-Pass filter
  - Kalman filter

“dealing” with a trending series

Univariate approaches:

- Deterministic trend = trend stationary process
  - Detrending is the correct approach
  - What happens if you difference a process with a deterministic trend?
- Stochastic trend = difference stationary process
  - Differencing is the correct approach
  - What happens if you detrend a process with a stochastic trend?

Multivariate approach

- Model the trend as part of the system with an economic, theoretical motivation as to the origin of the trend
  - In my view, the only correct approach if the purpose is using structural economic insights – topic 5: Cointegration

Trend Stationary Processes

- Consider an MA(q) process with a well-defined $C(L)$ and a linear deterministic trend:

$$y_t = a_0 t + C(L)\varepsilon_t$$

  - By assumption, the non-trend part of this process is stable

- The first moment depends on time, so the process is not covariance stationary:

$$E(y_t) = a_0 t$$

Trend Stationary Processes

- The deviation from expectation (i.e. deviations from the trend) will be stationary, however, so the second moments will be constants:

$$y_t - E(y_t) = a_2 t + C(L)\varepsilon_t - a_2 t = C(L)\varepsilon_t$$

- This suggests an approach to model this series:
  - First regress y on a vector of time periods (or dates) to get an estimate of the deterministic trend
  - Subtract the estimate of the time trend from y
  - Model further as an ARMA process

- This is called *detrending* (for historical reasons – jargon)
  - "Trend stationary" = stationary around a deterministic trend
  - "Detrending" = removing the deterministic trend, leaving only the stationary part

Trend Stationary Processes

- What happens if we difference a trend stationary process?
- Consider the simple process with $|a_1| < 1$:
$$y_t = a_0 t + a_1 y_{t-1} + \varepsilon_t$$
- Its first lag:
$$y_{t-1} = a_0(t-1) + a_1 y_{t-2} + \varepsilon_{t-1}$$
- Subtracting the second from the first:
$$\Delta y_t = a_0 + a_1 \Delta y_{t-1} + \varepsilon_t - \varepsilon_{t-1}$$
- Is this process stationary?

Trend Stationary Processes

- What happens if we difference a trend stationary process?
- Consider the simple process with $|a_1| < 1$:

$$y_t = a_0 t + a_1 y_{t-1} + \varepsilon_t$$

- Its first lag:

$$y_{t-1} = a_0(t-1) + a_1 y_{t-2} + \varepsilon_{t-1}$$

- Subtracting the second from the first:

$$\Delta y_t = a_0 + a_1 \Delta y_{t-1} + \varepsilon_t - \varepsilon_{t-1}$$

- Is this process stationary?
  - Yes – it is an ARMA(1,1) process – check by finding it's first two moments
  - It is, however, non-invertible there is a unit root in the moving average part, so the Box-Jenkins method cannot be used.
- Moreover, if the deterministic trend is not linear, components of it will remain

Trend Stationary Processes

- One can easily extend this to higher order deterministic trends:

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_1 t^n + C(L)\varepsilon_t$$

- This can be detrended by adding an appropriately high order polynomial in time to the estimation equation
  - Not empirically very effective – multi-collinearity between even/odd terms
  - Other functional forms work better
  - Orthogonal polynomials work very well (e.g. Chebyshev or Fourier polynomials)
- The problems: Does a (non-linear) deterministic trend make sense?
  - Philosophical: what does a deterministic trend mean in economics? What theoretical model predicts a permanent deterministic trend?
  - Practical: by the Taylor approximation theorem, any continuous function can be arbitrarily well approximated by a sufficiently high order polynomial. How do you know you have the "right" trend and the "right" stationary part? More on this later.

Difference Stationary Processes

- The simplest difference stationary process is called a random walk:

$$y_t = y_{t-1} + \varepsilon_t$$

- A key feature of this process (and all processes called *martingales*) is that its current value is the best forecast for all of time:

$$y_{t+1} = y_t + \varepsilon_{t+1}$$

$$E(y_{t+s}|y_t) = y_t \ \ \forall s > 0$$

- $E(y_{t+s}|y_t)$ is the expectation of y_{t+s} *conditional* on the observed value y_t
  - Also denoted $E_t(y_{t+s})$ where the subscript is read as "expectation conditional on all information up to period t", where information means observed values of variables/processes
  - Conditional expectations form a large an important part of statistics

Difference Stationary Processes

- The simplest difference stationary process is called a random walk:

$$y_t = y_{t-1} + \varepsilon_t$$

- Using recursive substitution:

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_t \\ &= y_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &= y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &\vdots \end{aligned}$$

$$y_t = y_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j}$$

Difference Stationary Processes

- The simplest difference stationary process is called a random walk:

$$y_t = y_{t-1} + \varepsilon_t$$

- Using recursive substitution:

$$y_t = y_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j} = y_0 + \sum_{s=1}^{t} \varepsilon_s$$

- The effects of any shock is **permanent**
  - The process retains *all* shocks over time – infinite memory
  - This process is said to have a *stochastic trend* $\sum_{s=1}^{t} \varepsilon_s$

Difference Stationary Processes

- Taking as y_0 as given, then

$$var(y_t|y_0) = E[(\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1)^2] = t\sigma^2$$

$$\begin{aligned}
\mathbb{E}[(y_t - y_0)(y_{t-j} - y_0)] &= \mathbb{E}[(\varepsilon_t + \varepsilon_{t-1} + \ldots + \varepsilon_1)\ldots \\
&\quad (\varepsilon_{t-j} + \varepsilon_{t-j-1} + \ldots + \varepsilon_1)] \\
&= \mathbb{E}[(\varepsilon_{t-j})^2 + (\varepsilon_{t-j-1})^2 + \ldots + (\varepsilon_1)^2] \\
&= (t-j)\sigma^2
\end{aligned}$$

- So, the process is not covariance stationary, as the second moments depends on time

Difference Stationary Processes

- This type of process is labelled *difference stationary* as the first difference is a stationary process:

$$y_t - y_{t-1} = \varepsilon_t$$
$$\Delta y_t = \varepsilon_t$$

- Another term for these processes are *integrated processes*

  - y_t is I(1) or first order integrated
  - Δy_t is I(0) or not integrated

  - A process can also be integrated of a higher order
  - An I(d) process is one that is stationary in d^{th} difference

Classes of Difference Stationary Processes

- Pure Random walk (Random Walk without *drift*):

$$y_t = y_{t-1} + \varepsilon_t$$

  - changes to the level are purely random: $\Delta y_t = \varepsilon_t$
  - the solution is $y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$

- Random walk with drift:

$$y_t = a_0 + y_{t-1} + \varepsilon_t$$

  - changes to the level are a constant plus a random term: $\Delta y_t = a_0 + \varepsilon_t$
  - the solution is $y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i$

- Random walk with noise
  - the solution is $y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i + \eta_t$

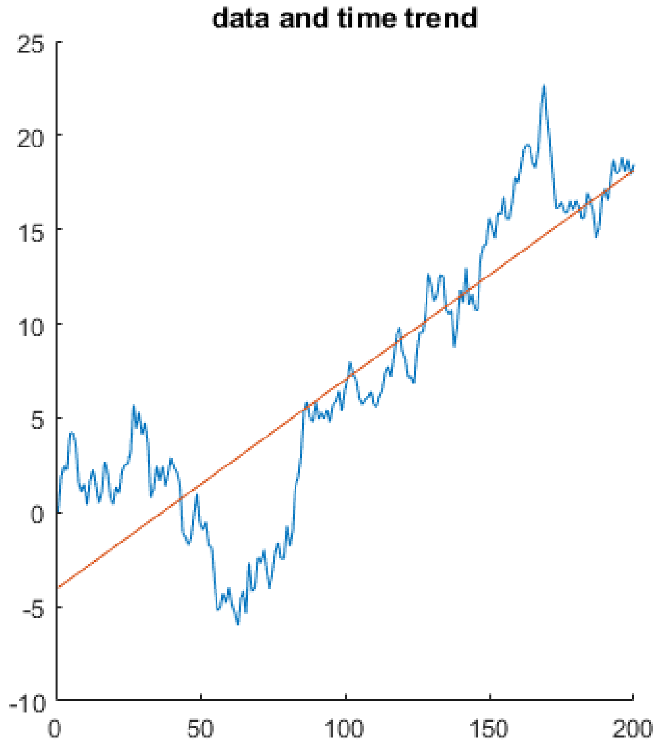
Economics

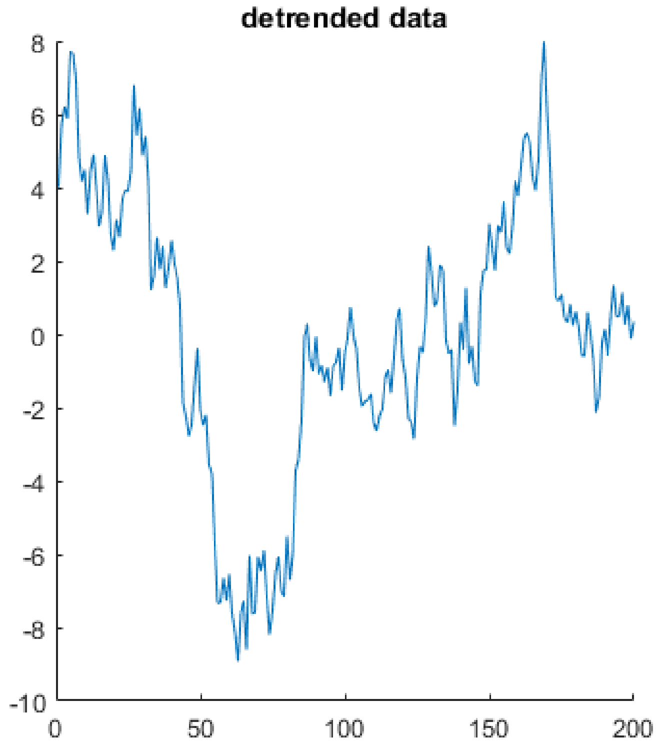
- From economic theory, what predicts
  - A deterministic trend?
    - Linear?
    - Higher order?
  - A stochastic trend?

Differencing vs Detrending

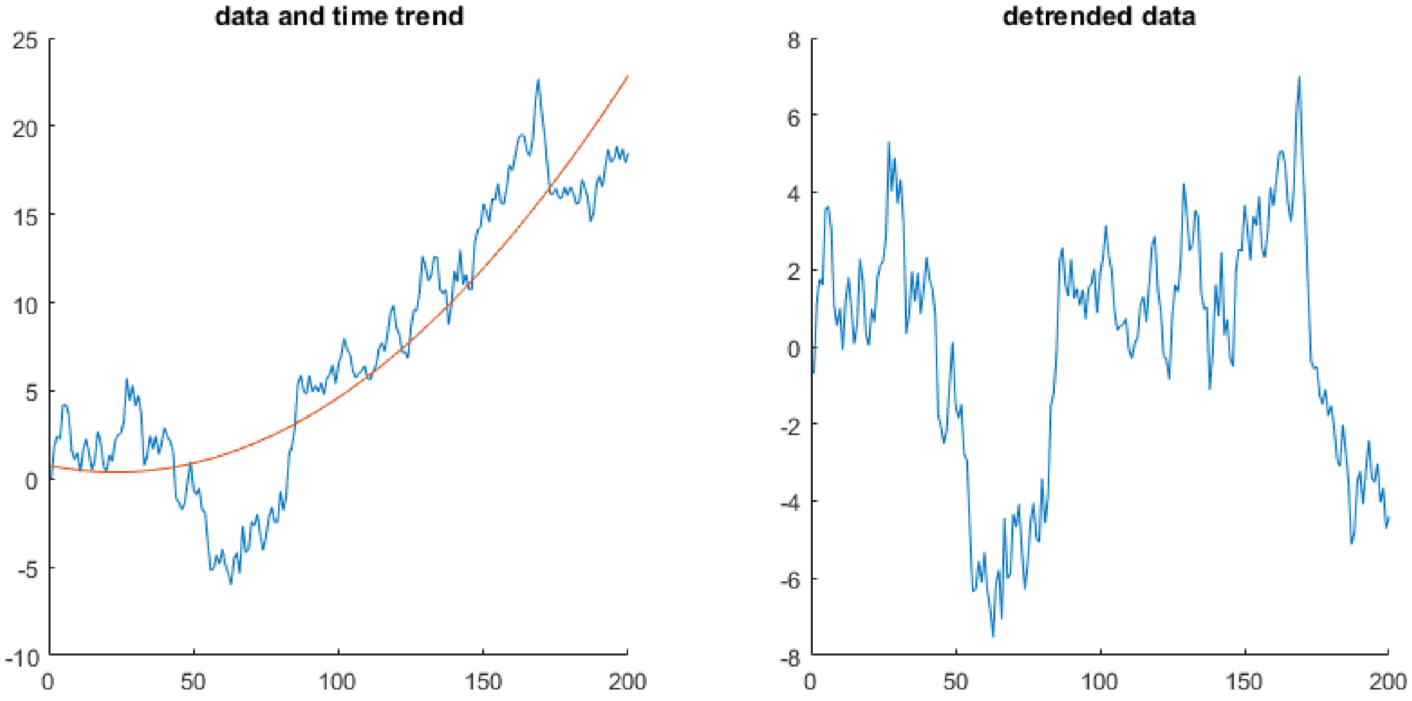
- We’ve shown that differencing a trend-stationary process (with a linear trend) yields a stationary process, although this is not the correct approach
- What happens if we detrend a difference stationary process?

Detrending a pure random walk: linear trend

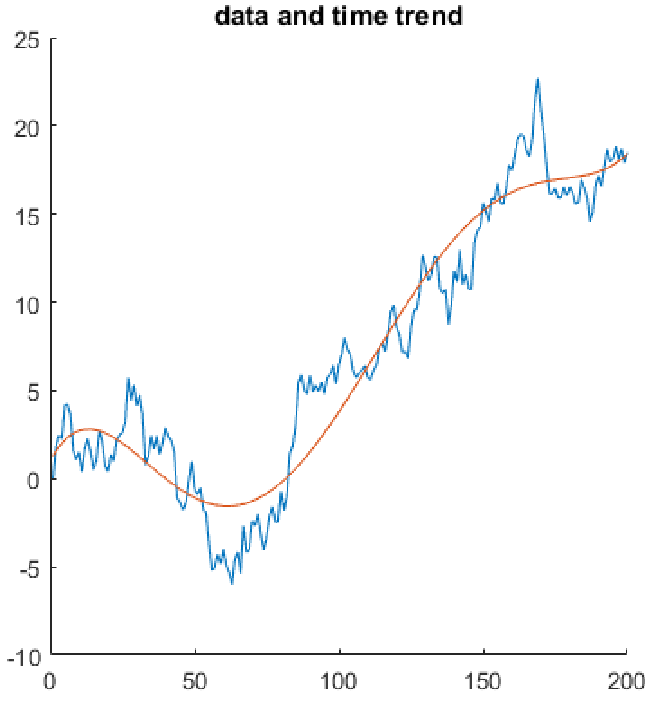


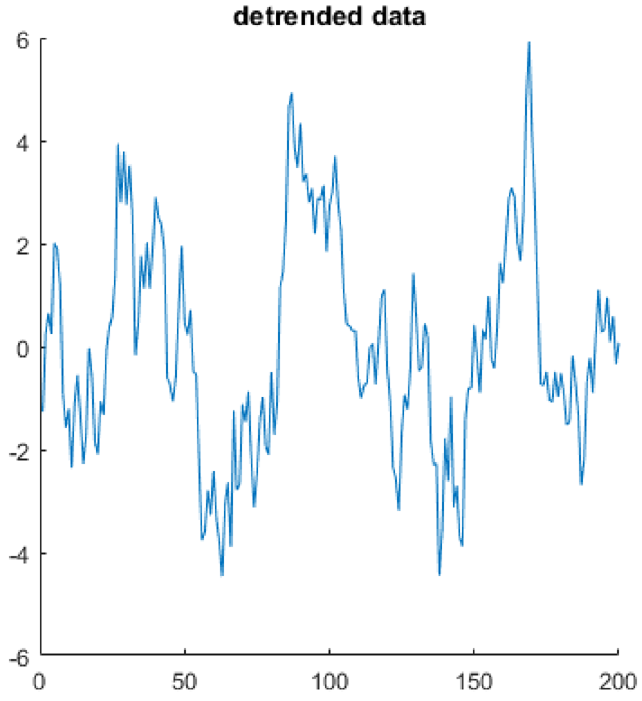


Detrending a pure random walk: quadratic trend



Detrending a pure random walk: 5^{th} order trend





Differencing vs Detrending

- What happens if we detrend a difference stationary process?
- Typically, it will not remove the unit root
  - However, removing a polynomial in time of high enough order will yield a series that will *test* as if it is stationary, for any empirical test
  - This will change the properties of the data into something we manufactured
  - This is particularly problematic in studies of the business cycle: if the true cause of cyclicality is a stochastic trend, detrending (or deterministic filtering) may induce "spurious cycles"

Random Walk with drift

- An integrated process with a constant term is a random walk with drift:

$$y_t = \mu + y_{t-1} + \varepsilon_t$$

- Using recursive substitution:

$$\begin{aligned}
y_t &= \mu + y_{t-1} + \varepsilon_t \\
&= \mu + (y_{t-2} + \mu + \varepsilon_{t-1}) + \varepsilon_t \\
&= 2\mu + (y_{t-3} + \mu + \varepsilon_{t-2}) + \varepsilon_{t-1} + \varepsilon_t \\
&\vdots \\
y_t &= \mu \cdot t + \sum_{j=0}^{t-1} \varepsilon_{t-j}
\end{aligned}$$

Random Walk with drift

- An integrated process with a constant term is a random walk with drift:

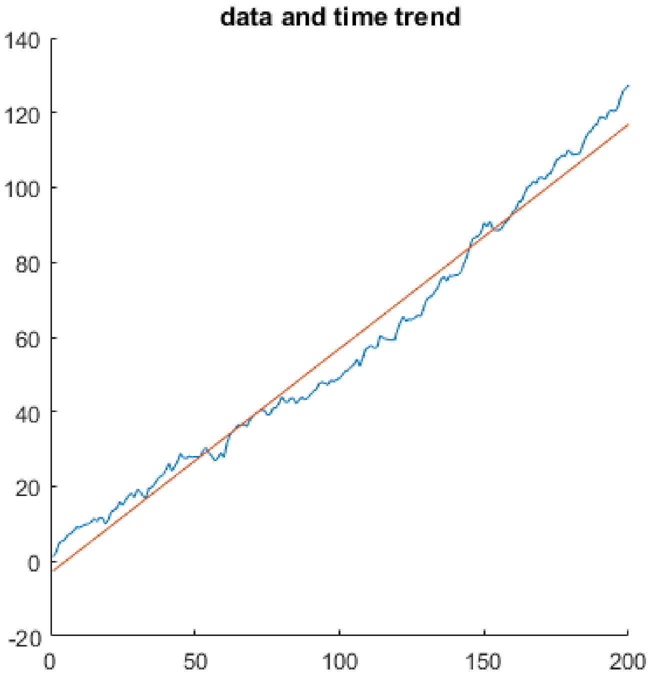
$$y_t = \mu + y_{t-1} + \varepsilon_t$$

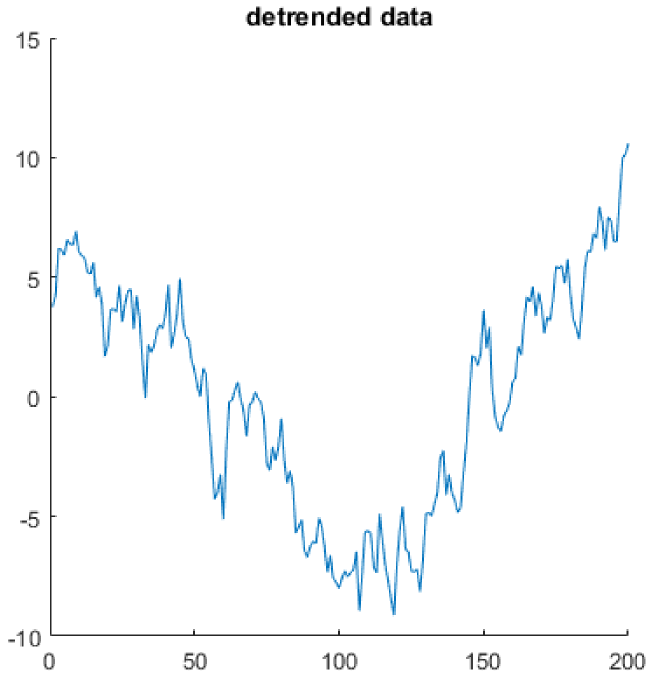
- Using recursive substitution:

$$y_t \quad = \quad \mu \cdot t + \sum_{j=0}^{t-1} \varepsilon_{t-j}$$

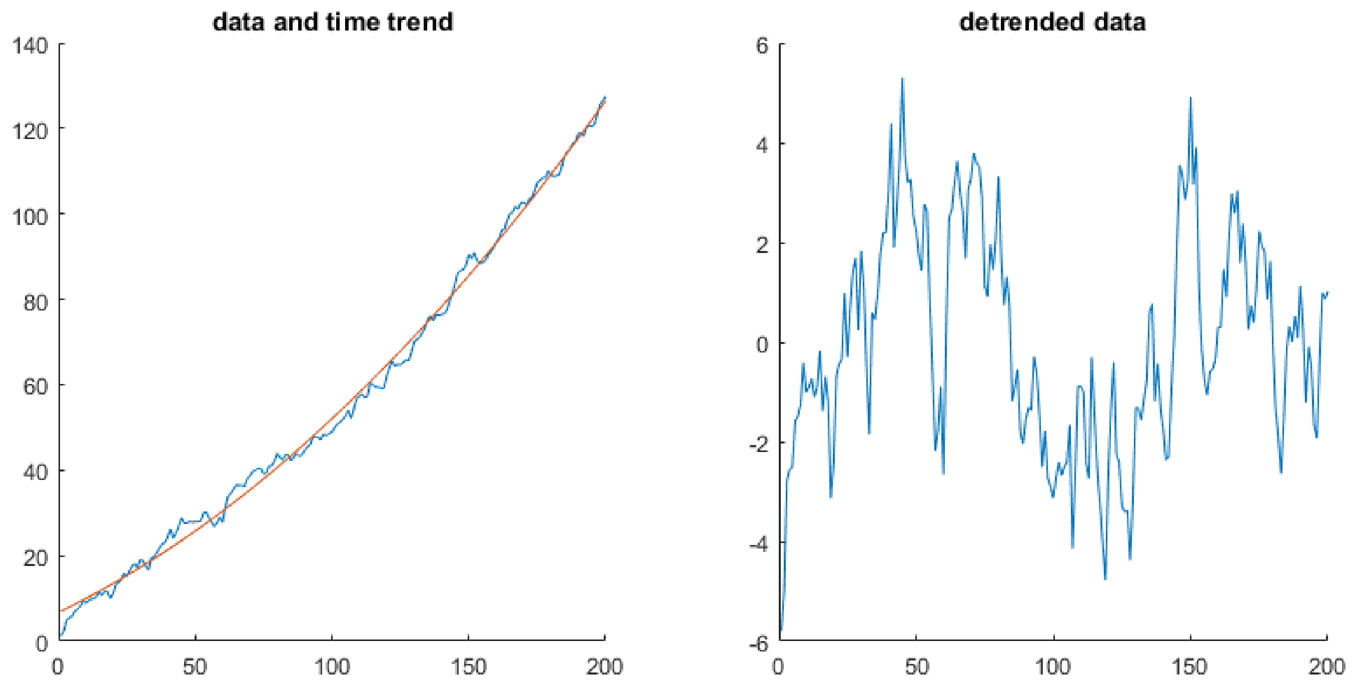
- I.e. this process has both a deterministic and a stochastic trend
  - Deviations from the trend, however, are not stationary

Detrending a random walk with drift

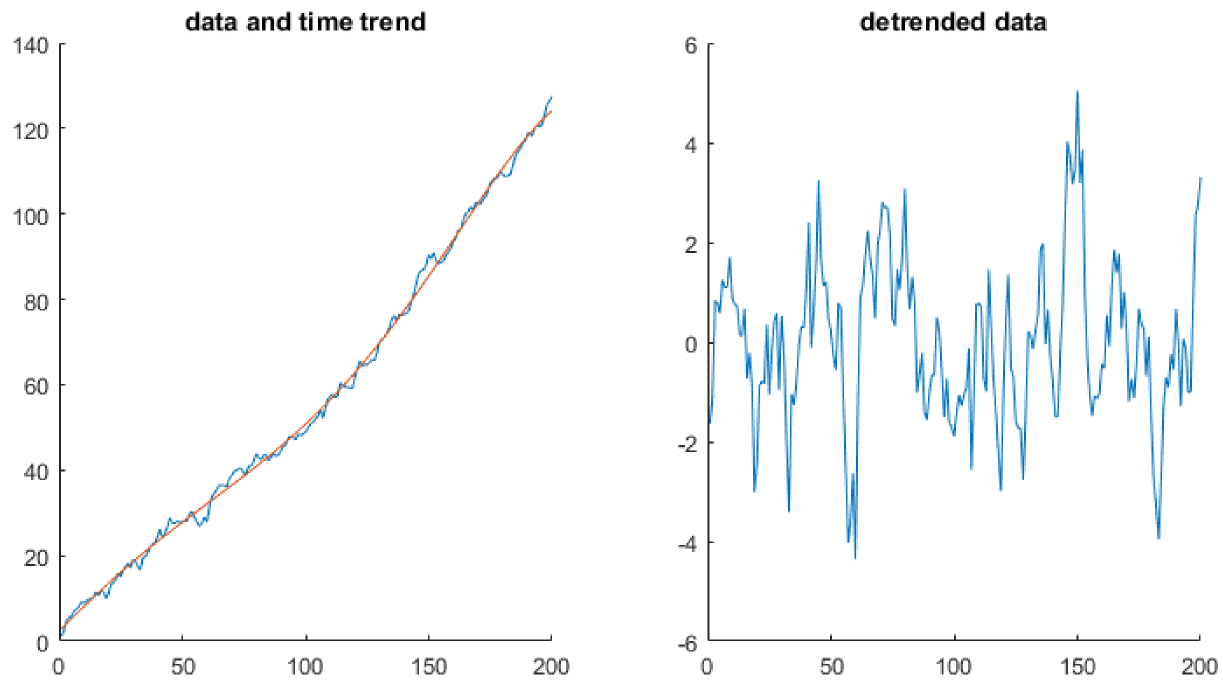




Detrending a random walk with drift



Detrending a random walk with drift



Detrending a random walk with drift

- **Linearly** detrending a random walk with drift is unlikely to yield an empirically stationary process
  - Although, depending on the exact realization of shocks, it might

- A high enough order polynomial in time will *necessarily* yield an empirically stationary series
  - Various approximation theorems show that a function (process) can be arbitrarily well approximated by a high enough polynomial in the basis function

- Given this, "polynomials in time" in econometrics is at best treated for completeness.
  - I have rarely seen it used in an empirical study of a real structural economic problem
  - At least not one I would cite…

ACF of a random walk
(conditional on an initial condition)

$$y_t = y_{t-1} + \varepsilon_t$$

$$\mathbb{E}[(y_t - y_0)(y_{t-j} - y_0)] \quad = \quad (t-j)\sigma^2$$

- Standard deviation:

$$\sqrt{\text{var}(y_t)} = \sqrt{t\sigma^2} \qquad \sqrt{\text{var}(y_{t-j})} = \sqrt{(t-j)\sigma^2}$$

- Autocorrelation:

$$\begin{aligned} \rho_j &= (t-j)\sigma^2 / \sqrt{(t-j)\sigma^2}\sqrt{(t)\sigma^2} \\ &= (t-j)/\sqrt{(t-j)t} \\ &= \sqrt{(t-j)/t} \quad < 1 \end{aligned}$$

ACF of a random walk

- The *empirical* ACF of a process with a unit root will decay, but slowly
- So will the ACF of a process with a deterministic trend
- So will the ACF of an AR(p) process with roots close to one
- Thus, we cannot use an ACF to distinguish between these cases

Detecting non-stationarity

- Simplest case: Unknown AR(1) process:

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

  – If $|a_1| < 1$, the process is stationary with finite, constant variance
    - regression yields a consistent estimate of a_1 with classical asymptotic distributional results
  – If $a_1 = 1$, the process is non-stationary with infinite variance:

$$\lim_{t\to\infty} var(y_t|y_0) = \lim_{t\to\infty} t\sigma^2 = \infty$$

    - regression estimate of a_1 is inconsistent, biased downwards, with a non-standard distribution
- Thus, it is tricky to set up a null and alternative hypothesis with a single known distribution

Tests for non-stationarity

- A variety of tests have been developed:
  - Dickey and Fuller (1979, 1981)
  - Phillips and Perron (1988)
  - ADF-GLS (Elliot, Rothemberg and Stock, 1996)
  - Ng and Perron (1995, 2001)
  - Kwiatkowski, Phillips, Schmidt and Shin (1992)
- The first four are all extensions/variants of the DF test
  - They test the **null of unit root** against **the alternative of no unit root**
  - The last tests the **null of no unit root** against the **alternative of a unit root**
- Note: Tests are only as good as the data is representative
  - If there are structural breaks, then a test that does not include this option will give inaccurate results
  - We will consider extensions to these tests that allow for one or more structural break under the non-linear topic

Dickey Fuller test

- Subtracting y_{t-1} from both sides yields:

$$\Delta y_t = (a_1 - 1)y_{t-1} + \varepsilon_t$$
$$= \gamma y_{t-1} + \varepsilon_t$$

- Now the hypotheses are not problematic:
  - If $|a_1| < 1 \Leftrightarrow \gamma < 0$,
    - y_t is stationary, thus so is Δy_t
    - a regression yields a consistent estimate of γ
  - If $a_1 = 1 \Leftrightarrow \gamma = 0$
    - Δy_t is stationary but y_t is I(1)
    - However: the I(1) term falls out of the regression at the null of a unit root
  - The distribution of γ is non-standard, obtained by Monte Carlo methods

Dickey Fuller test

- Dickey Fuller Test equation:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$$

- A standard t-test statistic from an OLS regressions is used: $t_{DF} = \frac{\hat{\gamma}}{\widehat{se}(\gamma)}$
  - The **null hypothesis** is:
    - The series contains a unit root, or equivalently that $\gamma = 0$
  - But the distribution (critical values) constructed via simulation

- This is a one-sided test
  - The **alternative hypothesis** is that the series is stationary, or that $\gamma < 0$
  - We use the simulated distribution to determine whether a specific estimate $\hat{\gamma}$ is small enough to be statistically significantly *smaller* than zero
  - An aside: Tests for bubbles are also build on this idea, but test whether $\gamma > 0$ in parts of the time path (i.e. locally explosive)

Dickey Fuller test

How comfortable are you with the concepts of a null and alternative hypothesis?

If not, review and come talk to me

- Dickey Fuller Test equation:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$$

- A standard t-test statistic from an OLS regressions is used: $t_{DF} = \frac{\hat{\gamma}}{\widehat{se}(\gamma)}$
  - The **null hypothesis** is:
    - The series contains a unit root, or equivalently that $\gamma = 0$
  - But the distribution (critical values) constructed via simulation

- This is a one-sided test
  - The **alternative hypothesis** is that the series is stationary, or that $\gamma < 0$
  - We use the simulated distribution to determine whether a specific estimate $\hat{\gamma}$ is small enough to be statistically significantly *smaller* than zero
  - An aside: Tests for bubbles are also build on this idea, but test whether $\gamma > 0$ in parts of the time path (i.e. locally explosive)

Dickey Fuller Test

- The previous version makes sense if there is no obvious trend in the series
  - When there is an obvious trend, we would like to test for a unit root with drift against an alternative of a stationary process around a deterministic trend

- The extended test equations are:

$$\Delta y_t = \beta_1 + (a_1 - 1)y_{t-1} + \varepsilon_t$$

$$\Delta y_t = \beta_1 + \beta_2 t + (a_1 - 1)y_{t-1} + \varepsilon_t$$

  - Critical values of the test differ in these cases, and are separately simulated and tabulated

  - Also, the interpretation of the coefficients is very different in null and alternative hypotheses:

    - Under H_0: β_1 is the rate of drift of a unit root process
    - Under H_1: β_1 is part of the constant of a stationary process ($\frac{\beta_1}{1-a_1}$)

Augmented Dickey Fuller Tests

- The above assumed we knew it was an ARIMA(1,1,0) or ARIMA(1,0,0) process, not some other ARIMA(p,d,q) process
  - I.e. it does not allow for higher order AR parts, or any MA parts
- For the Dickey Fuller tests to be valid, the estimators of the coefficients must be consistent
- I.e. the test regression has to encompass the DGP, for which a minimum requirement is white noise residuals
  - No autocorrelation (no ARMA behaviour)
  - Constant variance
- If the stationary part of the process is a higher order ARMA(p,q) process, the DF test equation above will not yield consistent estimates
  - because it will leave systematic information in the residuals
- The ADF test deals with the autocorrelation concern, but not the constant variance concern

Augmented Dickey Fuller Tests

- For an AR(2) process the test equation can be derived as follows:

$$y_t = \beta_1 + \beta_2 t + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t$$

- Add and subtract $a_2 y_{t-1}$:

$$y_t = \beta_1 + \beta_2 t + (a_1 + a_2) y_{t-1} + a_2 (y_{t-2} - y_{t-1}) + \varepsilon_t$$

$$y_t = \beta_1 + \beta_2 t + (a_1 + a_2) y_{t-1} - a_2 \Delta y_{t-1} + \varepsilon_t$$

- Subtract y_{t-1} from both sides:

$$\Delta y_t = \beta_1 + \beta_2 t + (a_1 + a_2 - 1) y_{t-1} - a_2 \Delta y_{t-1} + \varepsilon_t$$

Augmented Dickey Fuller Tests

- For more complicated ARIMA processes?
  - i.e. with MA parts

- ARIMA(p,1,0) can approximate ARIMA(p’,1,q) well

Augmented Dickey Fuller Tests

- The most general test equation becomes:

$$\Delta y_t = \beta_1 + \beta_2 t + \gamma_1 y_{t-1} + \sum_{i=2}^{p} \gamma_i \Delta y_{t-i+1} + \varepsilon_t$$

- Where the ADF test is based on the t statistic for γ_1
  – The choice of lags is usually done by the usual information criteria

- The tabulated critical values depend on the deterministic components
  – They also may depend on sample size, which is not easy to tabulate…

- Dickey and Fuller also tabulate the critical values for F tests of the joint test for unit root and deterministic parts

ADF tests: how to choose the appropriate version

- In previous years, I followed the textbook and how it often applied in papers
  - Test all versions, argue for the best
  - Kevin Kotze argues for a general to specific approach, but Enders warns against this as
- Last year, I realized that this was problematic as many of the tests have nonsensical alternative hypotheses
- In my opinion, there are only two versions of the test that are well motivated, and it depends entirely on the time series patterns of the data:
  - Specifically, does the process have an obvious trend or not
- We will explore this carefully in the tutorial session
  - It is critical that you go through the basic ideas of the test very carefully before that session

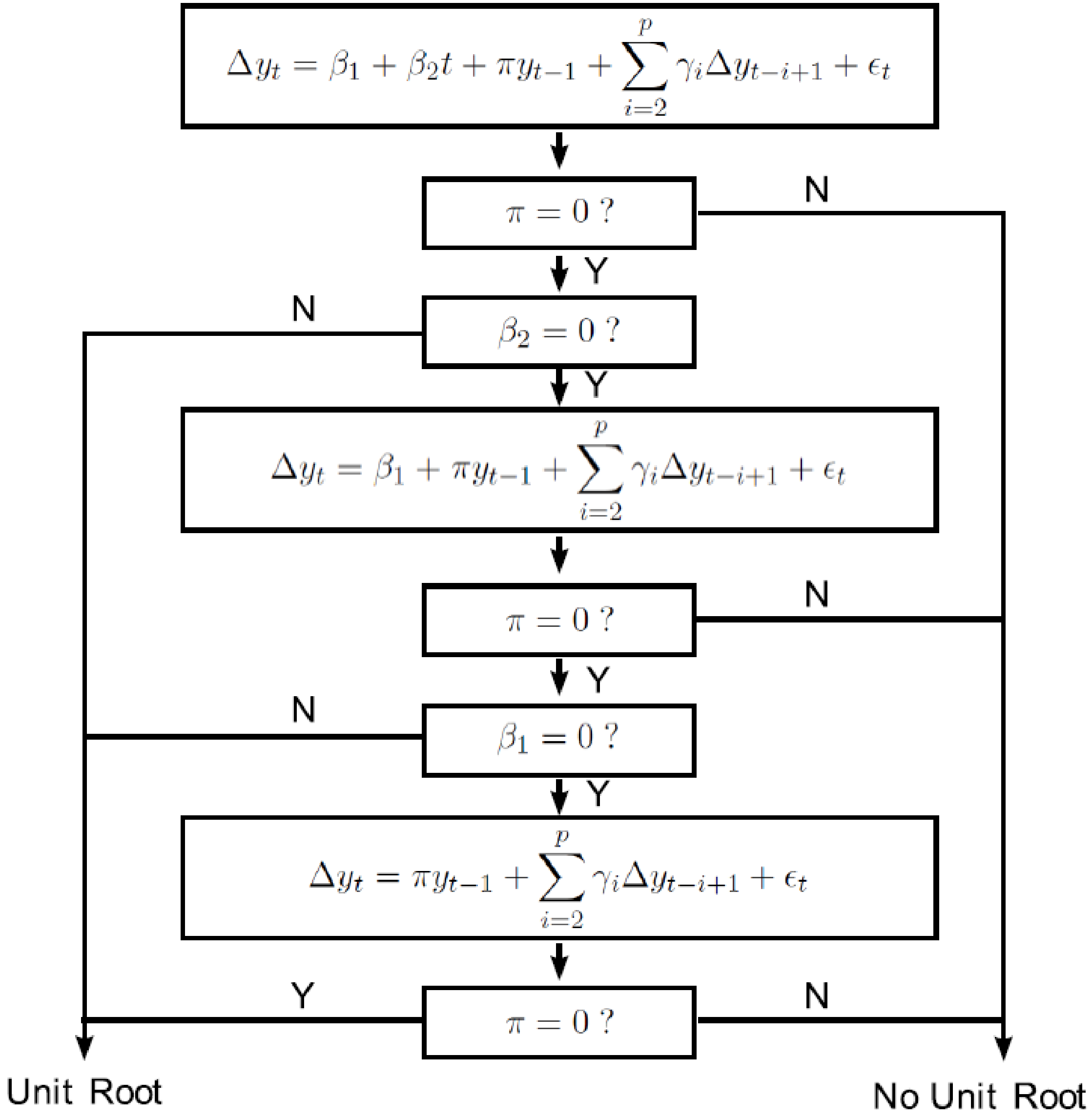
ADF tests: how to determine the order of integration

- The statistical properties of estimators and tests depend on the order of integration of the process involved
- A process that is integrated of order d must be differenced d times to be stationary
  - Denoted $I(d)$
  - A stationary process is $I(0)$ because it does not have to be differenced to be stationary
  - If a process y_t is $I(2)$,
    - Its first difference $x_t = \Delta y_t$ is $I(1)$
    - Its second difference $z_t = \Delta x_t = \Delta(\Delta y_t)$ is $I(0)$

ADF tests: how to determine the order of integration

- This yields an iterative scheme when confronted with real world data
  - Start by testing whether the process has a unit root in levels
    - If the test rejects, conclude that process it is $I(0)$ and continue with the rest of the analysis
    - If a unit root is not rejected, conclude that it is at least $I(1)$
  - If the level is at least $I(1)$, test for a unit root in the first difference
    - If the test rejects a unit root in the first difference, conclude that original process is $I(1)$ and continue with the rest of the analysis
    - If a unit root is not rejected in the first difference, conclude that the original series is at least $I(2)$
  - And so on…
- Mostly, we work with at most $I(2)$ process,
  - I've found that prices (e.g. CPI) and/or money supply sometimes test as $I(2)$ processes
  - Critically, one must have a clear economic reason why a process would have a unit root

A General to Specific testing approach



Source: Kevin Kotze
Economodel.com

Size vs Power of a statistical test
(jargon to know)

- The **size of a test** is the probability that a true null hypothesis will be rejected
  - If we say a coefficient is statistically significantly different from zero at 5%, we state that:

    IF the null is true (coefficient is zero), then we will mistakenly conclude that it is not zero in 5% of cases due to random sample variation

- The **power of a test** is the probability that a false null hypothesis will be rejected against a specific alternative
  - If the test has a null hypothesis $\alpha_1 = 1$ against an alternative of $\alpha_1 = 0.9$ and a power of 90%, then:

    If the null hypothesis is false and the alternative true, then we will correctly reject the null in 90% of cases

    Equivalently: 10% of cases we will incorrectly *fail to reject* the false null hypothesis and mistakenly conclude that $\alpha_1 = 1$

- Via Monte Carlo exercises it has been established:
  - ADF tests have low power against close alternatives with finite data (e.g. roots around 0.9) – i.e. tend to suggest a unit root when there isn't one
  - Additionally, they are likely to suggest a unit root when there are structural breaks in an otherwise stationary process

Other tests:

Phillips and Perron (1988)

- Derive asymptotic distributions of the basic AR(1) test equation statistic under very general assumptions on the residuals
  - Allowing for both autocorrelation *and* heteroscedasticity (non constant variance)
  - I.e. the residuals may have autocorrelated levels *and* variances (GARCH behaviour)
- This is analytically very dense and relies on functional analysis (i.e. the asymptotic convergence of functions)
- They construct adjusted **DF** test statistics (i.e. only up to AR(1) + deterministic parts) that are asymptotically consistent
- Similar to "small sample corrections" elsewhere in econometrics
- Simulations show that can improve on ADF power
  (see simulations for T=100, AR coeff 0.85)

A pause for context

- Up to this point is what I expect you to focus on in technical detail
  - The purpose is to
    - Understand what an integrated process is, and
    - Have a systematic way to test for the order of integration of an observed time series
  - The ADF is the simplest and most common application
    - It has shortcomings that later work have improved on
- What comes after is to expose you to the extensions that have since been designed
  - I expect you to know the gist of the issues, not any mathematical detail
  - For professional econometric work, you will have to delve a little bit into each of the following to know when and why to use the different tests

Other tests:

Phillips and Perron (1988):

- Constructed the exact, theoretical asymptotic distribution of the levels equation under weak assumptions
- Does not perform as well in small samples (tentative: see Enders)

- Requires deep measure theory to follow:

(i) $E(u_t) = 0$ for all t;
(ii) $\sup_t E|u_t|^{\beta+\varepsilon} < \infty$ for some $\beta > 2$ and $\varepsilon > 0$;
(iii) as $T \to \infty$, $\sigma^2 = \lim E(T^{-1}S_T^2)$ exists and $\sigma^2 > 0$, where $S_t = u_1 + \ldots + u_t$;
(iv) $\{u_t\}$ is strong mixing with mixing coefficients α_m that satisfy $\Sigma\, \alpha_m^{1-2/\beta} < \infty$, where the sum is over $m = 1, \ldots, \infty$.

Other tests:

ADF-GLS (Elliot, Rothemberg and Stock, 1996)

- Consider the general ADF test equation:

$$\Delta y_t = \beta_1 + \beta_2 t + \gamma_1 y_{t-1} + \sum_{i=2}^{p} \gamma_i \Delta y_{t-i+1} + \varepsilon_t$$

- Suppose $\beta_1, \beta_2 > 0$:
  - If $\gamma_1 < 0$, the process is stationary around a deterministic trend
  - If $\gamma_1 = 0$, the process has a stochastic trend, drift (i.e. a linear deterministic trend) due to β_1 *and* a quadratic deterministic trend due to $\beta_2 t$
  - The joint test for $\beta_2 = \gamma_1 = 0$ goes some way towards dealing with this

- The ADF-GLS test deals explicitly with a trend before unit root testing
  - It uses a GLS approach to consistently remove whatever deterministic trend there might be, and then does the standard ADF test
  - This leads to different critical values

Other tests:

Ng and Perron (1995, 2001)

- Combines the ideas of Phillips and Perron (1988) and the ADF-GLS approach of Elliot, Rothemberg and Stock (1996)
  - Take the ADF-GLS approach to detrend before testing
  - Then derive the asymptotic distribution of test statistics as in Phillips and Perron
  - This yields new small sample adjusted test statistics that are asymptotically consistent under very general error processes
  - Not yet available in many packages

Other tests:

Kwiatkowski, Phillips, Schmidt and Shin (1992)

- The previous tests all build on DF: testing the null of a unit root against the alternative of no unit root
- KPSS reverses this: The null is no unit root, the alternative is a unit root
- The intuition is remarkably simple. If a process has a unit root, it can be written as:

$$y_t = d_t + r_t + \varepsilon_t$$
$$r_t = r_{t-1} + u_t$$

  - Where d_t contains any stationary ARMA part and deterministic trends
  - r_t is the pure unit root/random walk part of the process.
  - If $\text{var}(u_t) = 0$, then r_t is a constant, and there is no unit root
- The null hypothesis is thus $\text{var}(u_t) = 0$ against alternative $\text{var}(u_t) > 0$

Other tests:

Kwiatkowski, Phillips, Schmidt and Shin (1992)

- The null hypothesis is thus $\mathrm{var}(r_t) = 0$ against the alternative $\mathrm{var}(r_t) > 0$
- They derive the following test statistic for the null hypothesis:

$$KPSS = \frac{\sum_{t=1}^{T} s_t^2}{\widehat{\mathrm{var}}\,(\varepsilon_t)}$$

- Where $s_t = \sum_{i=1}^{t} \hat{\varepsilon}_i$
- Again, the critical values for this test were constructed via simulation

Multivariate Danger: Spurious Regression

- 4 multivariate cases:
  - y and z are both stationary
    - normal regressions valid
  - y is I(b) and z is I(d) with b>d
    - Levels regression meaningless
  - y and z are I(1) with independent stochastic trends
    - Levels regression meaningless, in differences, valid
    - But often find significant coefficients
      = spurious regression results
  - y and z are I(1) with *common* stochastic trend
    - Levels regression valid and super consistent
    - Variables are "co-integrated"

Multivariate Danger: Spurious Regression

- y and z are I(1) with independent stochastic trends

$$y_t = y_{t-1} + \varepsilon_{y,t}$$
$$z_t = z_{t-1} + \varepsilon_{z,t}$$

  – Where $\varepsilon_{y,t}$ and $\varepsilon_{z,t}$ independent, normal random variables

- Consider the regression:

$$y_t = b_0 + b_1 z_t + u_t$$

  – By construction: $b_0 = b_1 = 0$
  – How often do we expect a normal significance test (at 95% confidence) to reject the hypothesis: $\mathrm{H}_0: b_1 = 0$?
    - If the classical results hold: 5% of the time
    - Granger and Newbold (1974) show that the rejection rate is 75%!

- Intuition?
  – A random walk meanders without pattern over its range. Highly likely that several pairs of independent random walk happen to meander in the same direction for part of the sample, which would lead to a spurious correlation and a significant regression coefficient

My (non-text-book) summary

- Pure deterministic trends make no sense in any economic model that I have ever seen
  - At best it can be a “local”/small sample solution
  - I am truly sceptical of any such “fixes”
  - However, I’m in a very small minority in the larger literature…

- A stochastic trend in any series does not always make economic “common sense”
  - Use your economic understanding to judge all cases

- Detrending OR differencing
  - At best a crude way to summarize data in a specific sample
  - Almost certainly strips out informative (long run) correlations with other variables that are more interesting to an economist
  - There are situations in which you can extract “real” economic conclusions from “pre-differencing” but they should be motivated on economic grounds, not empirical grounds.

- Always use co-integration analysis when you encounter a set of variables that
  - Seem to be individually integrated (and tests as such)
  - Can theoretically consistently be considered to be integrated
  - Should be in a joint equilibrium relationship based on economic theory

Next week: Stationary Multivariate Models

- Read chapter 5 of Enders
- Brush up on Linear Algebra and Matrix operations

Econometrics 871
Time Series

TOPIC 4: TUTORIAL

Replicating the Dickey-Fuller distribution

Tutorial: Monte Carlo Experiment:

1. Create data according to DGP of interest

2. Do candidate estimation

3. Compare distribution of residuals to standard predicted distribution

4. Use simulated distribution to obtain “true” critical values

Review of asymptotic results and hypothesis testing

- Consider the simplest linear regression on a sample of n observations:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad with \quad \varepsilon_i \sim (0, \sigma^2)$$

- OLS estimate:

$$\hat{\beta}_{1,OLS} = \frac{cov(y,x)}{var(x)}$$

- If the standard OLS assumptions hold then OLS is consistent:

$$\lim_{n\to\infty} \hat{\beta}_{1,OLS} = \beta_1$$

- In a finite sample, $\hat{\beta}_{1,OLS}$ is a random variable
  - If ε_i is i.i.d. normal, then $\hat{\beta}_{1,OLS}$ is also normal with $var\left(\hat{\beta}_{1,OLS}\right) = \frac{\sigma^2}{n\, var(x)}$
  - Even if ε_i is i.i.d. but not normal, the central limit theorem proves:

$$\lim_{n\to\infty} \sqrt{n}\left(\hat{\beta}_{1,OLS} - \beta_1\right) \sim N\left(0, \frac{\sigma^2}{n\, var(x)}\right)$$

Review of asymptotic results and hypothesis testing

- the central limit theorem proves:

$$\lim_{n\to\infty}\sqrt{n}\left(\hat{\beta}_{1,OLS}-\beta_1\right)\sim N\left(0,\frac{\sigma^2}{n\,var(x)}\right)$$

- Thus we might use the small sample **approximation** for hypothesis tests:

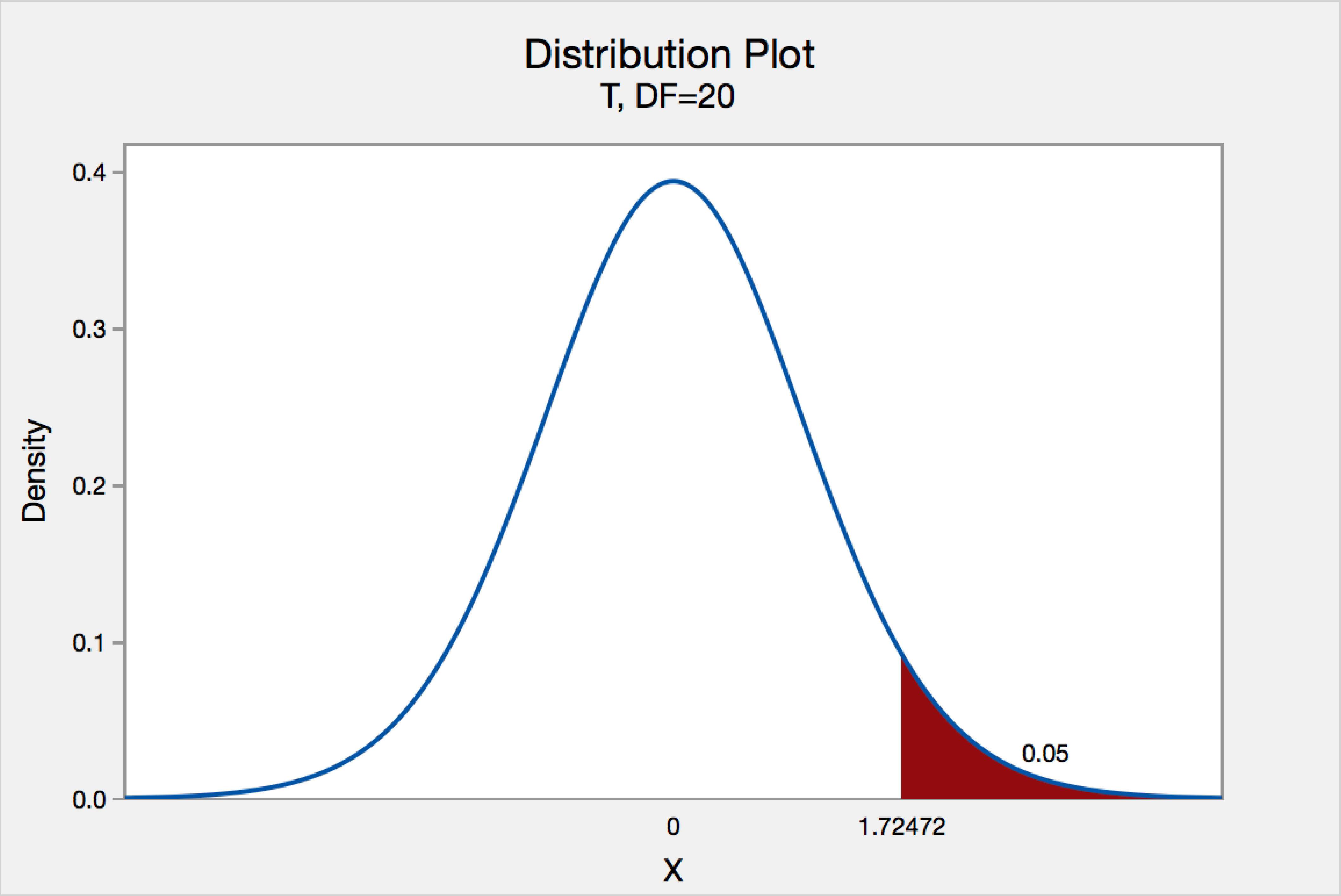
$$\frac{\left(\hat{\beta}_{1,OLS}-\beta_1\right)}{\sqrt{var\left(\hat{\beta}_{1,OLS}\right)}}\sim N(0,1)$$

- However, we do not know σ^2 - it has to be estimated with

$$\mathrm{s}^2=\frac{1}{\mathrm{n}}\sum\varepsilon_i^2$$

- Then we use the standard t-distribution

$$\frac{\left(\hat{\beta}_{1,OLS}-\beta_1\right)}{\sqrt{\frac{s^2}{n\,var(x)}}}\sim t(n-k)$$


Distribution Plot
T, DF=20
Density
0.4
0.3
0.2
0.1
0.0
0.05
0
1.72472
X

Fundamental Setting

Given an unknown AR(1) process:

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

- If $|a_1| < 1$
  - The process is stationary
  - An OLS regression of y_t on y_{t-1} yields a consistent (but biased) estimate of a_1
  - Let the sample be of size T
  - Biased means: $E(\hat{a}_1) \neq a_1$
  - Consistent means: $\lim_{T\to\infty} \hat{a}_1 = a_1$

- If $a_1 = 1$
  - The process is non-stationary
  - An OLS regression of y_t on y_{t-1} yields an inconsistent estimate of a_1: $\lim_{T\to\infty} \hat{a}_1 \neq a_1$
  - In this setting: $\lim_{T\to\infty} \hat{a}_1 < a_1$

Test equation:

- Subtracting y_{t-1} from both sides yields the test equation:

$$\Delta y_t = (a_1 - 1)y_{t-1} + \varepsilon_t$$

$$= \gamma y_{t-1} + \varepsilon_t$$

  - If $|a_1| < 1 \Leftrightarrow \gamma < 0$,
    - y_t is stationary, thus so is Δy_t
    - a regression of Δy_t on y_{t-1} yields a consistent estimate of γ, with standard distributional results (i.e. $\mathrm{t}_\gamma = \frac{\hat{\gamma}_{OLS} - \gamma}{s.e.(\hat{\gamma})}$ has an asymptotic t-distribution centred at zero)
    - Consistency: $\lim\limits_{T \to \infty} \hat{\gamma}_{OLS} = \gamma$
    - However, in a small sample $\hat{\gamma}_{OLS}$ will be biased because y_{t-1} is not exogenous with respect to ε_t: I.e. the condition $\mathrm{E}(y_t \varepsilon_s) = 0 \forall t, s$ does not hold

Test equation:

- Subtracting y_{t-1} from both sides yields the test equation:

$$\Delta y_t = (a_1 - 1)y_{t-1} + \varepsilon_t$$

$$= \gamma y_{t-1} + \varepsilon_t$$

- If $a_1 = 1 \Leftrightarrow \gamma = 0$
  - the I(1) term, y_{t-1}, falls out of the regression at the null of a unit root, so the regression is valid, but $\hat{\gamma}_{OLS}$ has a non-standard distribution
  - We will show that the *mode* of the distribution of $\hat{\gamma}_{OLS}$ is equal to γ, but the mean and median are not, so $\lim_{T\to\infty} \hat{\gamma}_{OLS} \neq \gamma$
  - Moreover, the distribution is non-standard ($\mathrm{t}_\gamma = \frac{\hat{\gamma}_{OLS}-\gamma}{s.e.(\hat{\gamma})}$ does not have a t-distribution)
  - Thus the critical values of the hypothesis test are different from those of a t-distribution at the null hypothesis of a unit root (i.e. $\mathrm{H}_0: \gamma = 0$)

Exercise for the day:

- Construct a Monte Carlo simulation that reconstructs the Dickey Fuller distribution and critical values for the t-test of a null of a unit root
- We will do a general simulation, for any value of γ (unit root and no unit root)
- We will show that:
  - If $\gamma < 0$, $\hat{\gamma}_{OLS}$ is on average correct/consistent and the distribution of the test of H_0: $\hat{\gamma}_{OLS} = \gamma$ has a an approximate t-distribution *only if* the sample of observations T is large enough
    - This raises a subtle point not often discussed: for near-unit root processes, small sample test statistics can be misleading
  - If $\gamma = 0$, $\hat{\gamma}_{OLS}$ is on average incorrect/inconsistent and the distribution of the test of H_0: $\hat{\gamma}_{OLS} = 0$ does not have a t-distribution *no matter* how large the sample of observations is

Monte Carlo Simulation

- For a process defined by a given AR coefficient a_1:
$$y_t = a_1 y_{t-1} + \varepsilon_t$$
  - Generate N different time-paths of length T
  - For each time-path $i \in N$,
    - do the OLS regression of the test equation:
$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$$
    - Store $\hat{\gamma}_{OLS}$ and $\frac{\hat{\gamma}_{OLS} - \gamma}{s.e.(\hat{\gamma})}$
  - Approximate the density function of $\hat{\gamma}_{OLS}$ and $\frac{\hat{\gamma}_{OLS} - \gamma}{s.e.(\hat{\gamma})}$
  - Compare the density function of $\frac{\hat{\gamma}_{OLS} - \gamma}{s.e.(\hat{\gamma})}$ to that of a standard t-distribution
  - Compute the empirical critical t-statistic and compare to the theoretical t-statistic of an α significance level

- Study the impact of varying a_1 and T