Class 4: Non-stationary Time Series

Macroeconometrics - Fall 2011

Jacek Suda, BdF and PSE

November 23, 2011

Outline

Outline:

1. Unit Root
    - Dickey-Fuller Test
    - Phillips-Perron
    - Stationarity Tests
    - Variance Ratio Test
2. Structural Breaks

UNIT ROOT

Autoregressive Unit Root Tests

- ARMA(p, q) process:

$$\phi(L)y_t = \theta(L)\varepsilon_t, \qquad \varepsilon_t \sim WN$$

- Consider hypothesis:
  $H_0 : \phi(z) = 0$, where $\phi(z)$ is characteristic equation.

$$\begin{aligned} \phi(L) &= 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p \\ \phi(z) &= 0 \Rightarrow 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = 0 \\ &\Rightarrow (1 - \frac{1}{\lambda_1} z)(1 - \frac{1}{\lambda_2} z) \cdots (1 - \frac{1}{\lambda_p} z) = 0 \end{aligned}$$

- $H_0 : \phi(z) = 0$ has (at least) one root on unit circle.

Unit Root in ARMA

- Then

$$y_t = y_{t-1} + u_t,$$

- and, given y_0,

$$y_t = y_0 + \sum_{j=0}^{t-1} u_{t-j} \sim I(1).$$

Shocks do not die out.

- An alternative for H_0 is

$$\begin{aligned} \mathrm{H}_1 : \quad & \phi(z) = 0 \text{ has all roots outside unit circle} \\ & y_t = \phi(L)^{-1}\theta(L)\varepsilon_t \\ & y_t = \Psi(L)\varepsilon_t = u_t \sim I(0) \end{aligned}$$

Shocks will die out over time.

Dickey-Fuller: Case 1

Consider AR(1):

$$\begin{aligned} y_t &= \phi y_t + \varepsilon_t, \qquad \varepsilon_t \sim WN \\ & \phi(L) = 1 - \phi L \end{aligned}$$

Then

$$\begin{array}{llll} H_0: & \phi(z) = 0 & \text{has unit root} & \Leftrightarrow \quad \phi = 1 \\ H_1: & \phi(z) = 0 & \text{has roots outside unit circle} & \Leftrightarrow \quad |\phi| < 1 \end{array}$$

Standard test statistics:

$$\hat{t}_\phi = \frac{\hat{\phi} - \phi}{\hat{SE}(\hat{\phi})},$$

where $\hat{\phi}$ comes from OLS on $y_t = \hat{\phi} y_{t-1} + \hat{\varepsilon}_t$.

Dickey-Fuller Result

Testing for any $\phi \neq 1$

$$t_{\phi=0.9} = \frac{\hat{\phi} - 0.9}{\hat{SE}(\hat{\phi})} \sim t - distribution$$

Testing for $\phi = 1$:

$$t_{\phi=1} = \frac{\hat{\phi} - 1}{\hat{SE}(\hat{\phi})} \sim DF$$

$$DF \xrightarrow{d} \frac{\int_0^1 W(r)dW(r)}{(\int_0^1 W(r)^2 dr)^{1/2}}$$

- It is based on continuous time random walk process
- Both numerator and denominator are functions of r, W
- It is theoretical result: the distribution can be found numerically by simulation

Brownian Motion

A Wiener process (Brownian motion) $W(\cdot)$ is a continuous-time stochastic process, associating each date $r \in [0, 1]$ a scalar random variable $W(r)$ that satisfies:

1. $W(0) = 0$
2. For any dates $0 \leq t_1 \leq \ldots \leq t_k \leq 1$, the changes $W(t_2) - W(t_1), W(t_3) - W(t_2), \ldots, W(t_k) - W(t_{k-1})$ are independent normal with
$$W(s) - W(t) \sim N(0, (s - t))$$
3. $W(s)$ is continuous in s.

Intuition: A Wiener process is the scaled continuous time limit of a random walk.

Dickey-Fuller distribution

Dickey-Fuller distribution

- Does not have a closed form representation.
- Is not centered around 0.
- 5% critical value for 1-side test is -1.94 (-1.65 for Normal)
- 1% critical value for 1-side test is -2.57 (-2.32 for Normal)
- Note: -1.65 is the 9.45% quantile of the DF distribution.
- It has less power under H_0, but higher size adjusted power.

Additionally,

- $\hat{\phi}$ is super-consistent; that is, $\hat{\phi} \xrightarrow{p} \phi$ at rate T instead of the usual rate $T^{1/2}$.

Nuisance parameter

- Assume

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

- For $H_0: \quad \phi = 0$,

$$t_{\phi=0} = \frac{\phi - 0}{\hat{SE}(\hat{\phi})} \overset{A}{\sim} N(0,1)$$

  Asymptotically, the distribution is always $N(0,1)$, no matter what c and σ^2 are.
- If the test statistics does not depend asymptotically on other parameters (nuisance parameter) it is pivotal.
- Note: It may not be pivotal for small sample; for example, for $t = 100$ it may depend on c and/or σ^2.

Nuisance parameter in DF

- DF statistics, even asymptotically, depends on c — c is a nuisance parameter.
- Dickey and Fuller shows that if $\phi = 1$ nuisance parameters are important not only in small sample but also in asymptotic distribution (it's no longer pivotal testing).
- The small-sample distribution for DF converges to asymptotic distribution much faster than in normal case: even for $t = 100$ it will look very much like asymptotical distribution; if $\phi = 0.9$ it will require a lot of observations to get to normal distribution.

Dickey-Fuller: Case 2

Constant + AR(1):

$$(y_t - \mu) = \phi(y_{t-1} - \mu) + \varepsilon_t, \qquad \varepsilon_t \sim iid\ WN$$

- Under $H_0 : \phi = 1$

$$y_t = y_0 + \sum_{j=1}^{t} \varepsilon_j \sim I(1), \qquad y_0 = \mu,$$

- i.e. ε, shock, never dies out - its effect will be forever present in the series.
- Alternatively, under $H_1 : |\phi| < 1$

$$y_t = c + \phi y_{t-1} + \varepsilon_t \sim I(0), \qquad c = \mu(1 - \phi),$$

with shocks dying out over time.

Dickey-Fuller distribution

- The $t-$statistics is

$$t^{\mu}_{\phi=1} = \frac{\hat{\phi} - 1}{\hat{SE}(\hat{\phi})}$$

- from OLS regression $y_t = \hat{c} + \hat{\phi} y_{t-1} + \hat{\varepsilon}_t$,
- Dickey-Fuller shows that, under $H_0 : \ \phi = 1$, it is

$$t^{\mu}_{\phi=1} \xrightarrow{d} DF^{\mu} = \frac{\int_0^1 W^{\mu}(r) dW(r)}{\left(\int_0^1 W^{\mu}(r)^2 dr\right)^{1/2}},$$

- with

$$W^{\mu}(r) = W(r) - \int_0^1 W(r) dr$$

the "de-meaned" Wiener process, $\int_0^1 W^{\mu}(r) = 0$.

Remarks

- If $y_0 = \mu \neq 0$, it converge to DF^{μ}, if $\mu = 0$ then DF.
- It doesn't matter what value of y_0 is.
- The asymptotic distributions of these test statistics are influenced by the presence (but not the value) of the constant in the test regression
- The inclusion of a constant pushes the distributions of $t^{\mu}_{\phi=1}$ to the left:
  - 5% critical value for 1-side test is -2.86 (-1.65 for Normal)
  - 1% critical value for 1-side test is -3.43 (-2.32 for Normal)
  - 1.65 is the 45.94% quantile of the DF^{μ} distribution!

Dickey-Fuller: Case 3

The test regression is

$$y_t - c - \beta \cdot t = \phi(y_{t-1} - c - \beta \cdot (t-1)) + \varepsilon_t$$

and includes a constant and deterministic time trend to capture the deterministic trend under the alternative.

Hypothesis

- $H_0 : \ \phi = 1$:

$$y_t = c + \beta \cdot t + \sum_{j=1}^{t} \varepsilon_j \sim I(1) \text{ with drift,}$$

  where $c + \beta \cdot t$ denotes deterministic component, and $\sum_{j=1}^{t} \varepsilon_j$ the random walk component.

- $H_1 : \ |\phi| < 1$:

$$\begin{aligned} y_t &= c + \beta \cdot t + \phi y_{t-1} + \varepsilon_t \sim \text{Trend stationary} \\ y_t &- \beta \cdot t \sim I(0) \end{aligned}$$

Test statistics

- Test statistics

$$t^{\beta}_{\phi=1} = \frac{\hat{\phi} - 1}{\widehat{SE}(\hat{\phi})}$$

where $\hat{\phi}$ is from OLS regression

$$y_t = \hat{c} + \hat{\beta} \cdot t + \hat{\phi} y_{t-1} + \varepsilon_t.$$

- Both β and c are nuisance parameters.
- Under $H_0: \ \phi = 1$

$$t^{\beta}_{\phi=1} \xrightarrow{d} DF^{\mu} = \frac{\int_0^1 W^{\beta}(r) dW(r)}{(\int_0^1 W^{\beta}(r)^2 dr)^{1/2}},$$

with

$$W^{\beta}(r) = W^{\mu}(r) - 12\left(r - \frac{1}{2}\right)\int_0^1 \left(s - \frac{1}{2}\right) W(s) ds,$$

Test statistics

- De-meaned and detrended Wiener process.
- The inclusion of a constant and trend in the test regression further shifts the distribution of $t^{\beta}_{\phi=1}$ to the left.
  - 5% critical value for 1-side test is -3.41 (-1.65 for Normal)
  - 1% critical value for 1-side test is -3.96 (-2.32 for Normal)
  - 1.65 is the 77.52% quantile of the DF^{β} distribution!
- Test DF$^{\mu}$ has more power than DF$^{\beta}$ if $\beta = 0$.

Extending DF

- The previous unit root tests are valid if the time series y_t is well characterized by an AR(1) with white noise errors.
- Many economic and financial time series have a more complicated dynamic structure than is captured by a simple AR(1) model.
- Said and Dickey (1984) augment the basic autoregressive unit root test to accommodate general ARMA(p, q) models with unknown orders and their test is referred to as the augmented Dickey-Fuller (ADF) test

Hypothesis

- Basic AR(p) model

$$\begin{aligned} \phi(L)y_t &= \varepsilon_t, \qquad \varepsilon_t \sim WN \\ \phi(L) &= 1 - \phi_1 L - \ldots - \phi_p L^p \end{aligned}$$

- Hypothesis

$$\begin{aligned} H_0: &\quad \phi(z) = 0 \ \text{ has one unit root} \\ &\quad \phi(z) = (1-z)\phi^*(z), \qquad \phi^*(z) \ \text{ has no unit root.} \\ H_1: &\quad \phi(z) = 0 \ \text{ has all roots outside unit circle.} \end{aligned}$$

Transformation

- Dickey-Fuller transformation of $\phi(L)$

$$y_t = \rho y_{t-1} + \phi_1^* \Delta y_{t-1} + \phi_2^* \Delta y_{t-2} + \ldots + \phi_{p-1}^* \Delta y_{t-p-1} + \varepsilon_t,$$

where

$$\begin{aligned} \rho &= \phi_1 + \phi_2 + \ldots \phi_p \\ \phi_j^* &= -\sum_{k=j+1}^{p} \phi_k \end{aligned}$$

- It's just different representation of AR(p) process.
- Example: AR(2):

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \\ &= \phi_1 y_{t-1} + \phi_2 y_{t-1} - \phi_2 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \\ &= (\phi_1 + \phi_2) y_{t-1} - \phi_2 \Delta y_{t-1} + \varepsilon_t \\ &= \rho y_{t-1} + \phi_1^* \Delta y_{t-1} + \varepsilon_t. \end{aligned}$$

Hypothesis restated

- The hypothesis can be simply restated as

$$\begin{aligned} H_0 : \quad & \rho = 1 \Leftrightarrow \text{ unit root} \\ H_1 : \quad & |\rho| < 1 \Leftrightarrow I(0) \end{aligned}$$

in equation

$$y_t = \rho y_{t-1} + u_t, \qquad u_t \sim I(0)$$

with u_t containing lagged differences to capture serial correlation in u_t.

ADF

- Augmented Dickey-Fuller Test (ADF Test)

$$\hat{t}_{\rho=1} = \frac{\hat{\rho} - 1}{\widehat{SE}(\hat{\rho})}$$

from OLS regression

$$y_t = \hat{\rho} y_{t-1} + \hat{\phi}_1^* \Delta y_{t-1} + \ldots + \hat{\phi}_{p-1}^* \Delta y_{t-p-1} + \varepsilon_t.$$

- The distribution of t-statistics is

$$\begin{aligned} \hat{t}_{\rho=1} &\xrightarrow{d} DF \\ \hat{t}^{\mu}_{\rho=1} &\xrightarrow{d} DF^{\mu} \\ \hat{t}^{\beta}_{\rho=1} &\xrightarrow{d} DF^{\beta}. \end{aligned}$$

Intuition

Re-parameterize AR(2) model

$$
\begin{aligned}
y_t &= \rho y_{t-1} + \phi_1^* \Delta y_{t-1} + \varepsilon_t \\
\rho &= \phi_1 + \phi_2 \\
\phi_1^* &= -\phi_2
\end{aligned}
$$

- $y_{t-1} \sim I(1) \ \Rightarrow \ \hat{\rho}$ has a non-normal, asymptotic "unit root" distribution;
- $\Delta y_{t-1} \sim I(0) \Rightarrow \ \hat{\phi}_1^*$ has an asymptotic normal distribution

Remarks

Remarks:

- If $\phi(L)y_t = \theta(L)\varepsilon_t$, ADF works asymptotically as p grows with sample size at rate $T^{1/3}$.
- If p unknown: choose large enough p to eliminate serial correlation in u_t in $y_t = \rho y_{t-1} + u_t$.
- If p is too small then the remaining serial correlation in the errors will bias the test.
- If p is too large then the power of the test will suffer.
- Monte Carlo experiments suggest it is better to error on the side of including too many lags.
- Choose max lag (e.g. 12 for monthly data). Test last lag with $|t_{\phi^*}| > 1.645$
- Backward selection procedure.

Phillips-Perron Unit-Root Test

Model

$$\Delta y_t = \rho y_{t-1} + u_t, \qquad u_t\text{– serially correlated residuals}$$

- We do not specify how it is correlated, do not put any parametric approach.
- If $\sum \phi^*$ is close to -1, ADF has terrible size.
- Phillips-Perron addresses this issue

Phillips-Perron Unit-Root Test

- The PP tests correct for any serial correlation and heteroskedasticity in the errors u_t of the test regression.
- It directly modifies the test statistics $t_{\rho=0}$:

$$
\begin{aligned}
Z_t &= \left(\frac{\hat{\sigma}^2}{\hat{\lambda}^2}\right)^{1/2} t_{\rho=0} - \frac{1}{2}\left(\frac{\hat{\lambda}^2 - \hat{\sigma}^2}{\hat{\lambda}^2}\right)\left(\frac{T \cdot \widehat{SE}(\hat{\rho})}{\hat{\sigma}^2}\right) \\
t_{\rho=0} &= \frac{\hat{\rho}}{\widehat{SE}(\hat{\rho})}
\end{aligned}
$$

Phillips-Perron Unit-Root Test

- Terms $\hat{\sigma}^2$ and $\hat{\lambda}^2$ are consistent estimates of the variance parameters

$$
\begin{aligned}
\sigma^2 &= \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E(u_t^2) \\
\lambda^2 &= \lim_{T\to\infty} \sum_{t=1}^{T} E(T^{-1} S_T^2) = \text{“long run variance”} \\
S_T &= \sum_{t=1}^{T} u_t .
\end{aligned}
$$

- Result: Under the null hypothesis that $\rho = 0$, the PP Z_t statistic has the same asymptotic distributions as the ADF t-statistic.

Phillips-Perron Unit-Root Test

- The sample variance of the least squares residual $\hat{u}_t$ is a consistent estimate of σ^2.
- The Newey-West long-run variance estimate of u_t using $\hat{u}_t$ is a consistent estimate of λ^2.

$$
\begin{aligned}
\hat{\lambda}^2 &= \hat{\gamma}_0 + 2\sum_{j=1}^{m}\left[1 - \frac{j}{m+1}\right]\hat{\gamma}_j^* \\
\hat{\gamma}_0 &= \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 \\
\hat{\gamma}_j^* &= \frac{1}{T}\sum_{t=j+11}^{T}\hat{u}_t\hat{u}_{t-j}
\end{aligned}
$$

Stationarity Tests

- Nelson-Plosser find that all macro variables have unit roots.
- They failed to reject H_0 of the presence of unit root (we may reject it because of the low power of the test).
- H_0 in unit root test is that there is unit root and N-P fails to reject it.
- We want to then flip the problem and test if the series is stationary so that unit root would make that we reject stationary H_0.
- Kwiatkowski, Phillips, Schmidt and Shin (KPSS), 1992 JoE - non parametric approach;
- Leybourne and McCabe (1994, JEBS) - parametric approach.

Unobserved Components Model

Unobserved Components Model

$$y_t = \mu_t + \varepsilon_t,$$

where

$$\begin{array}{rcl} \mu_t &=& \mu_{t-1} + u_t \qquad\qquad u_t \sim iid(0, \sigma_u^2), \qquad \mu_0 = constant, \\ \varepsilon_t &\sim& I(0) \qquad\qquad (i.e.\ \ \phi(L)\varepsilon_t = \theta(L)\eta_t) \end{array}$$

μ_t is unobserved component.

- it takes both cases of unit root and stationarity
- $\mu_t =$ local mean $+$ unobserved component: it changes every time.
- Even though we don't observe this shock we can still recover it.

Stationarity Tests (KPSS)

Hypotheses:

$$
\begin{aligned}
H_0: &\quad \sigma_u^2 = 0 \implies y_t \sim I(0) \\
H_1: &\quad \sigma_u^2 > 0 \implies y_t \sim I(1)
\end{aligned}
$$

- Test is equivalent to testing for unit MA root in Δy_t.
- How to estimate variance $\sigma^2 = 0$, in model that is already difficult to estimate? It is equivalent to testing for unit MA root in Δy_t.

Unit MA root

- So far we talked about AR unit root: $\sigma_u^2 > 0$.
- There is alternative approach MA unit root to difference which corresponds to the general notion of overdifferencing.

Unit MA root:

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t \qquad\qquad \text{apply } (1-L) \\
(1-L)y_t &= (1-L)\mu_t + (1-L)\varepsilon_t \\
\Delta y_t &= u_t + \varepsilon_t - \varepsilon_{t-1}
\end{aligned}
$$

- We never observe the shock (unless $u_t = 0$).
- Under $H_0: \ \sigma_u^2 = 0$, $u_t = 0$ so if there is a shock today then tomorrow will exactly offset it => no permanent shock to accumulate of the process.
- So unit MA root implies no permanent effect of shock.

Granger representation

If $\varepsilon_t \sim iid$, Granger representation theorem implies that

$$\Delta y_t = e_t + \theta e_{t-1},$$

where e_t is unobservable forecast error.
If $cov(u_t, \varepsilon_t) = 0$

$$cov(\Delta y_t, \Delta y_{t-1}) = cov(u_t + \varepsilon_t - \varepsilon_{t-1}, u_{t-1} + \varepsilon_{t-1} - \varepsilon_{t-2}) = -\sigma_\varepsilon^2$$

and

$$cov(\Delta y_t, \Delta y_{t-j}) = 0, \qquad \forall j > 1$$

- The same autocovariance structure as in MA(1) process.

MA(1) representation

$$\Delta y_t = e_t + \theta e_{t-1},$$

- For $\sigma_{\varepsilon u} = 0$

$$\theta = \frac{-(q+2) + \sqrt{q^2 + 4q}}{2}, \qquad q = \frac{\sigma_u^2}{\sigma_\varepsilon^2}.$$

  q is the signal-to-noise ration.

- As $q \to 0$ we have unit MA root $\to$ we have permanent shock but they are very very small.
- If $\sigma_u^2 = 0$ then

$$q = 0 \Longrightarrow \theta = \frac{-2}{2} = -1$$

  so $\Psi^*(L) = 1 + \theta L$ has unit root.

KPSS Test

Testing

- Regress Δy_t on MA(1) process and see if $\theta = -1$.
- Not easy to do -> problem with power.
- KPSS proposes one-sided LM statistics for hypothes

$$
\begin{aligned}
H_0 : &\quad \sigma_u^2 = 0 &\quad \text{no random walk component, just constant} \\
H_1 : &\quad \sigma_u^2 > 0 &
\end{aligned}
$$

- LM statistics depends on process for y_t

KPSS: Case 1

- Case 1: constat term only

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t \\
\mu_t &= \mu_{t-1} + u_t, \qquad\qquad \mu_0 = constant
\end{aligned}
$$

- Test regression

$$
y_t = \alpha + \varepsilon_t \Longrightarrow \hat{\varepsilon}_t = y_t - \bar{y}
$$

- LM test:

$$
\hat{\eta}_\mu = \frac{1}{T^2} \sum_{t=1}^{T} \frac{S_t^2}{\Lambda^2},
$$

  where
  - $S_t^2 = \sum_{j=1}^{t} \hat{\varepsilon}_j$ is a partial sum over time of residuals, and
  - Λ^2, spectral density at frequency 0

- Sum up sample residual over time $\longrightarrow$ under H_0 they should not be a big number, they should cancel out. Otherwise, (under alternative) they should get larger and larger.

KPSS: Case 1

- Under H_0 : $\sigma_u^2 = 0$

$$\eta_\mu \xrightarrow{d} \int_0^1 V(r)^2 dr$$

where

$$V(r) = W(r) - rW(1) \quad = \text{standard Brownian Bridge}$$

- Reject at 5% if $\hat{\eta}_\mu > 0.463$.
- To do it need to estimate Λ^2, spectral density at frequency 0.
- KPSS proposes against Bartlet kernel approach:
  - depending how you choose the bandwidth you get different statistics depending on what bandwidth you choose -> big sample size distortions.
  - Some people suggest using parametric approach to construct test statistics.

KPSS: Case 2

- Case 2: constat + trend

$$
\begin{aligned}
y_t &= \tau \cdot t + \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim I(0) \\
\mu_t &= \mu_{t-1} + u_t, \qquad u_t \sim iid(0, \sigma_u^2)
\end{aligned}
$$

- Test regression

$$y_t = \alpha + \tau \cdot t + \varepsilon_t$$

- LM test:

$$\hat{\eta}_\mu = \frac{1}{T^2} \sum_{t=1}^{T} \frac{S_t^2}{\Lambda^2},$$

- Reject H_0 at 5% if $\hat{\eta}_\tau > 0.146$.

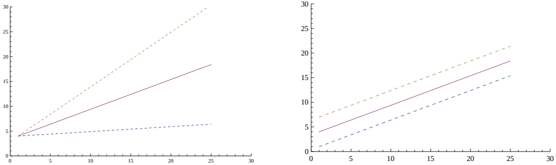
Testing

- Now we have unit root and stationarity test: apply both.
  1. Unit root test: you can't reject H_0; KPSS test: reject H_0. If put together, they imply that series has unit root.
  2. If we can't reject both test: data give not enough observations.
  3. Reject unit root, reject stationarity: both hypothesis are component hypothesis – heteroskedasticity in series may make a big difference; if there is structural break it will affect inference.
- Power problem: if there is small random walk component (small variance σ_u^2), we can't reject unit root and can't reject stationarity.
- Economics: if the series is highly persistence we can't reject H_0 (unit root) – highly persistent may be even without unit root but it also means we shouldn't treat/take data in levels.
- If we want to quantify how important the unit root is, we should use Variance Ratio Test.

Idea

- Cochrane, 1988
- non-paramteric measure of "economic" importance of unit root

Idea



- Variance of random walk grows linearly with horizon (no unconditional variance)
- Variance of trend stationary process is finite
  (it may grow over short horizon but it will finally settle down)
- Test the behavior of variance.

Variance

- Let
$$V_k = (1/k)var(Y_{t+k} - Y_t - k\mu)$$
  - V_k the variance of k^{th} period difference,
  - μ is deterministic trend that does not affect the variance.
- If there is no random walk it should converge to zero as variance is converging to constant and k is growing.
- Rewrite it as
$$\begin{aligned} V_k &= \frac{1}{k}E\left[(\Delta y_{t+1} - \mu) + (\Delta y_{t+2} - \mu) + \ldots + (\Delta y_{t+k} - \mu)\right]^2 \\ &= \gamma_0^* + 2\sum_{j=1}^{k-1}\left(\frac{k-j}{k}\right)\gamma_j^* \\ &= \text{weighted average of auto-covariances } \gamma_j^* = cov(\Delta y_t, \Delta y_{t+j}) \end{aligned}$$
- Auto-covariances are important because everything about the covariance-stationary series is in auto-covariance generating function.
- If model reduced to covariance can do analysis non-parametrically – don't have to specify the model when using sample auto-covariances.

Variance ratio

- To make the test better suited to compare series:
  - Variance ratio

$$VR_k = \frac{V_k}{V_1}, \qquad V_1 = \gamma_0^*$$

- <u>Result</u>

$$\lim_{k\to\infty} V_k = \sum_{k=-\infty}^{\infty} \gamma_k^* = \Lambda^2 = \text{spectral density at frequency 0 for } \Delta y$$

Example: Random Walk

- Random walk

$$y_t = y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

- Then

$$\begin{aligned} y_t &= y_0 + \sum_{j=1}^{t} \varepsilon_j, \qquad y_{t+k} = y_0 + \sum_{j=1}^{t+k} \varepsilon_j \\ y_{t+k} - y_t &= \sum_{j=t+1}^{t+k} \varepsilon_j \end{aligned}$$

- Variance

$$\begin{aligned} var(y_{t+1} - y_t) &= \sigma_\varepsilon^2, \qquad var(y_{t+k} - y_t) = k\sigma_\varepsilon^2 \\ V_1 &= \sigma_\varepsilon^2; \qquad V_k = \frac{1}{k} var(y_{t+k} - y_t) = \sigma_\varepsilon^2 \qquad \forall k \end{aligned}$$

- Variance ratio:

$$VR_k = \frac{V_k}{V_1} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2} = 1, \qquad \forall k.$$

- shock today has effect on series today and on series in the future.

Example: White Noise

- White noise

$$y_t = \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

- Variance

$$\begin{aligned} var(y_{t+1} - y_t) &= 2\sigma_\varepsilon^2 \\ var(y_{t+k} - y_t) &= 2\sigma_\varepsilon^2 \\ V_k &= \frac{2\sigma_\varepsilon^2}{k} \qquad \forall k \end{aligned}$$

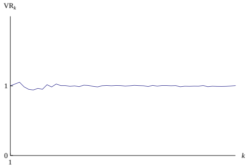
- Variance ratio

$$VR_k = \frac{1}{k} \to 0 \qquad \text{as } k \to \infty$$

- So for different type of process the variance ratio behaves differently.
- In practice, we have to estimate VR_k.
- Cochrane estimates $\widehat{VR}_k$ using Newey-West $\hat{\Lambda}_{NW}^2, \hat{\gamma}_0^*, \hat{\gamma}_j^*$.

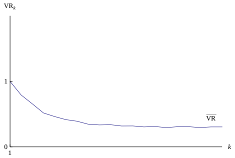
Implications

1. If $\widehat{VR}_k \to 1$



   - random walk, all shocks are permanent

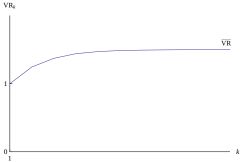
2. If $\widehat{VR}_k \to 0 < \overline{VR} < 1$



   - "mean reverting" I(1) process

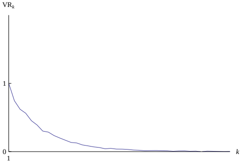
Implications

3. If $\widehat{VR}_k \to \overline{VR} > 1$



   - "mean averting" I(1) process
   - possible serial correlation in differences

4. If $\widehat{VR}_k \to 0$



   - trend stationary I(0) process

Implications

- If k too large you get spurious "mean reversion".
- In sample it always the case that

$$\hat{\gamma}_0 + \sum_{h=1}^{T-1} \hat{\gamma}_h^* = 0$$

  so mean reversion has to appear.
- Which k to use? : balance the two effects.
- Non parametric approach makes problem in small sample.
- At long horizon GDP has neither $\widehat{VR} \to 0$ not $\widehat{VR} \to 1$, it's between. With standard errors, though, you can't reject any of them.

Structural Breaks

- Failure to reject unit root could reflect structural breaks, not unit root
- Issue: frequency of permanent shocks
  how often do you have permanent shock to the mean or drift

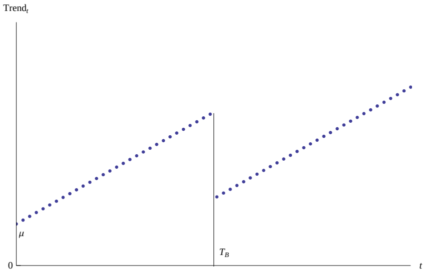
Trend stationary	$\Longrightarrow$	never
TS with one structural break	$\Longrightarrow$	once
TS with n structural breaks	$\Longrightarrow$	n times
TS with T structural breaks	$\Longrightarrow$	Unit root (difference stationary)

Perron, 1989 *Econometrica*,
"The Great Crash, The Oil Price Shock, and the Unit Root Hypothesis"

- Given a level a level break in 1929, the unit root can be rejected for 11 of 14 Nelson-Plosser (1982) series including at 10% real GNP and nominal stock prices. (annual data).
- Given the break in growth in 1973, Perron rejects unit root for postwar quarterly GDP.

Model A: The "Great Crash" Model

Model A: The "Great Crash" Model

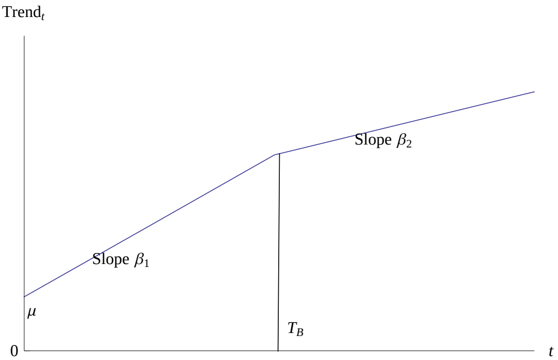


- Model of trend fluctuations of GDP.
- Trend is deterministic, when removed one gets the covariance stationary series.

$$
\begin{aligned}
Trend_t &= \mu_1 + \beta \cdot t + (\mu_2 - \mu_1) DU_t + e_t \\
DU_t &= \begin{cases} 1 & \text{if } t > T_B \\ 0 & \end{cases}
\end{aligned}
$$

Model B: The "Oil Shock" Model

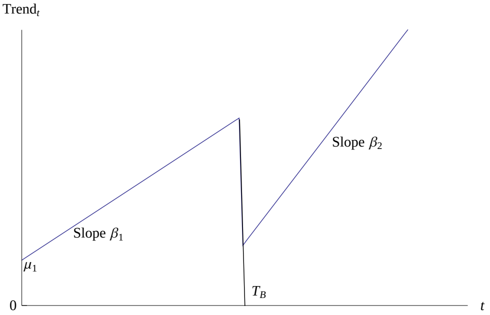
Model B: The "Oil Shock" Model



$$
\begin{aligned}
Trend_t &= \mu + \beta_1 \cdot t + (\beta_2 - \beta_1) DT_t^* + e_t \\
DT_t^* &= \begin{cases} t - T_B & \text{if } t > T_B \\ 0 & \end{cases}
\end{aligned}
$$

Model C: The “Combo” Model

Model C: The “Combo” Model



$$
\begin{aligned}
Trend_t &= \mu_1 + \beta_1 \cdot t + (\mu_2 - \mu_1) DU_t + (\beta_2 - \beta_1) DT_t^* + e_t \\
DU_t &= \begin{cases} 1 & \text{if } t > T_B \\ 0 \end{cases} \\
DT_t^* &= \begin{cases} t - T_B & \text{if } t > T_B \\ 0 \end{cases}
\end{aligned}
$$

Test statistics

- Maybe there is more structural breaks but only one big.
- Perron: do unit root test as always, OLS:

$$y_t = \rho y_{t-1} + Trend_t(\lambda) + \sum_{k=1}^{p-1} \phi_k^* \Delta y_{t-k} + \varepsilon_t$$

  - $\lambda = \frac{T_B}{T}$ denotes location of break date.
  - Lagged difference to capture serial correlation.
- The t-statistics

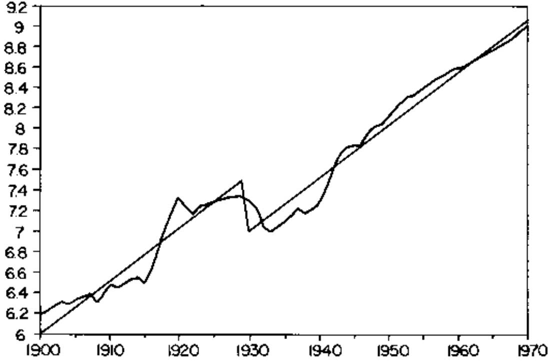
$$\hat{t}_{\rho=1}(\lambda) = \frac{\hat{\rho}(\lambda) - 1}{\widehat{SE}(\hat{\rho}(\lambda))}$$

  - We have more nuisance parameters that affect distribution

$$\hat{t}_{\rho=1}(\lambda) \overset{A}{\sim} \frac{\int_0^1 \underline{W}_\lambda(r) d\underline{W}_\lambda(r)}{\left(\int_0^1 \underline{W}_\lambda(r)^2 dr\right)^{1/2}},$$

  - $\underline{W}_\lambda$ is demeaned, detrended, dedummied Brownian motion.
  - The distribution is shifted further left than ADF.

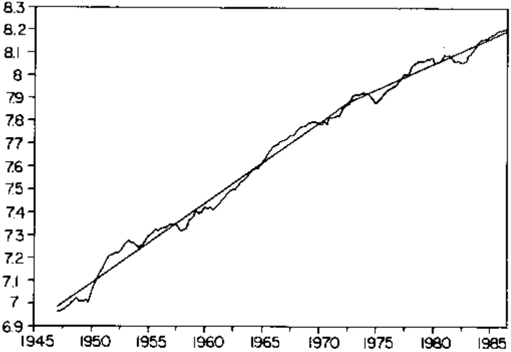
Cochrane (1989)



Note: The broken straight line is a fitted trend (by OLS) of the form $\tilde{y}_t = \tilde{\mu} + \tilde{\gamma} DU_t + \tilde{\beta} t$ where $DU_t = 0$ if $t \leqslant 1929$ and $DU_t = 1$ if $t > 1929$.

FIGURE 1.—Logarithm of "Nominal Wages."

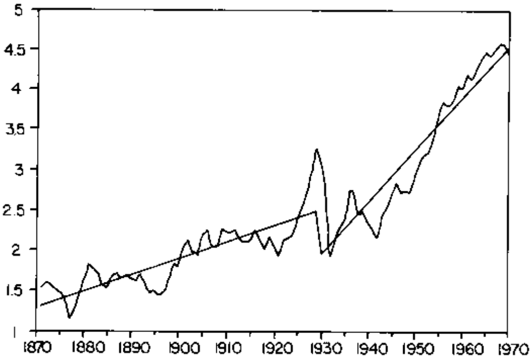
Cochrane (1989)



Note: The broken straight line is a fitted trend (by OLS) of the form: $\tilde{y}_t = \tilde{\mu} + \tilde{\beta}t + \tilde{\gamma}\, DT_t^*$ where $DT_t^* = 0$ if $t \leqslant 1973{:}\mathrm{I}$ and $DT_t^* = t - T_B$ if $t > 1973{:}\mathrm{I} = T_B$.

FIGURE 2.—Logarithm of "Postwar Quarterly Real GNP."

Cochrane (1989)



Note: The broken straight line is a fitted trend (by OLS) of the form $\bar{y}_t = \bar{\mu} + \bar{\gamma}_1 DU_t + \bar{\beta} t + \bar{\gamma}_2 DT_t$ where $DU_t = DT_t = 0$ if $t \leqslant 1929$ and $DU_t = 1$, $DT_t = t$ if $t > 1929$.

FIGURE 3.—Logarithm of "Common Stock Prices."

Criticism

- Data mining: *How Perron know the structural break was in 1929? He looked into data.*
- λ must be chosen independently of the data for the correct size of the test (or else there is bias against unit root, Zivot and Andrews, 1992 JBES)
  - size: if H_0 true how often you reject it
  - power: if H_1 is true (H_0 false) how often do you reject
  - *Small size and large power is optimal. We normally fix size (e.g.5% size and make test as powerful as possible.*
- $t_{5\%} = -3.8$ critical value
  λ chosen after looking at data: choosing λ so that it generates the largest t-statistics–test distributions is ever ore shifted so actual size might be bigger. Actual size might be 30% even though was set to 5%.

Criticism: Zivot and Andrews, 1992

- Need a model of break date selection procedure (Zivot and Andrews, 1992)
- $\hat{\lambda}_{INF}$ = break dates that produces the largest value of $|\hat{t}_{\rho=1}(\lambda)|$ over all λs in the sample.

$$\hat{t}_{\rho=1}(\hat{\lambda}_{INF}) = \inf_{\lambda \in \Lambda} \{\hat{t}_{\rho=1}(\lambda)\}$$

- Zivot and Andrews find that for 8 of 14 Nelson-Plosser series (including GNP) $\hat{\lambda}_{INF} = 1929$ are stationary and the t-statistics is distributed as

$$\hat{t}_{\rho=1}(\hat{\lambda}_{INF}) \overset{A}{\sim} \inf_{\lambda \in \Lambda} \left\{ \frac{\int_0^1 \underline{W}_\lambda(r) d\, \underline{W}_\lambda(r)}{\left(\int_0^1 \underline{W}_\lambda(r)^2 dr\right)^{1/2}} \right\}.$$

- It shifts distribution further left of Perron's.

Detrending

- Need stationary series:

$$Y_t = X_t\beta + \varepsilon_t$$

  - Granger and Newbold (1974, JoE, "Spurious Regressions in Econometrics")
  - If y_t and X_t are independent random walk ($\beta = 0$), $\hat{\beta}_{OLS} \rightarrow$ non-zero random variable, and $\hat{t}_{\beta=0}$ is large: spurious regression phenomenon.
- Taking difference instead of levels (so we get stationary series) will bring larger standard errors => cannot reject hypothesis.
- Detrending still allows to analyze levels.
- Sometimes we are interested in trend alone.

Trend/Cycle

- Observable series y_t

$$y_t = \tau_t + c_t$$

  - τ_t is trend, and
  - c_t is transitory component, $(I(0))$.
- If trend contains stochastic component, random walk, then if we apply HP we get spurious cycle.

$$\tau_t = \mu + \tau_{t-1} + \eta_t$$

- We have two unobserved components and if we can model the cycle we can try to use unobserved component estimation.

Unobserved Components Approach

- Watson (1986, JME), Clark (1987, QJE), Morley, Nelson, Zivot (2003, ReStat)
- Approach: parametric model for c_t
- Model *("Structural)*

$$
\begin{aligned}
y_t &= \tau_t + c_t \\
\tau_t &= \mu + \tau_{t-1} + \eta_t, \qquad \eta_t \sim iidN(0, \sigma_\eta^2) \\
\phi(L)c_t &= \varepsilon_t, \qquad \varepsilon_t \sim iidN(0, \sigma_\varepsilon^2), \\
& cov(\eta_t, \varepsilon_t) = \sigma_{\varepsilon\eta}
\end{aligned}
$$

Problem: Identification

- We have 1 observable series and 2 unobservable components.
- To get 2 unobservable components, we need some identification assumptions.

<u>Identification</u>:

- If $c_t = \varepsilon_t$ or $c_t = \phi c_{t-1} + \varepsilon_t$, then $\sigma_{\varepsilon\eta}$ is not identified *from the data*.
- There can be infinitely many values of $\sigma_{\varepsilon\eta}$ that would produce the same autocovariance generating function for the first series.
- However, that does not mean that all values of $\sigma_{\varepsilon\eta}$ are equal.
- If it is set to zero, it imposes restriction on autocovariance generating function of 1^{st} differences.

Example: AR(1)

Example: AR(1)

$$\begin{aligned} y_t &= \tau_t + c_t \\ \tau_t &= \mu + \tau_{t-1} + \eta_t \\ c_t &= \phi c_{t-1} + \varepsilon_t \end{aligned}$$

- Structural model: 5 parameters: $\mu, \sigma_\eta^2, \sigma_\varepsilon^2, \phi, \sigma_{\varepsilon\eta}$.
- How many parameters can be identified from data?

Reduced-Form

- First-difference equation

$$\begin{aligned} y_t &= \tau_t + c_t \\ (1-L)y_t &= (1-L)\tau_t + (1-L)c_t \\ \Delta y_t &= \mu + \eta_t + (1-L)(1-\phi L)^{-1}\varepsilon_t \end{aligned}$$

Example: AR(1)

- Multiply both sides by $(1-\phi L)$:

$$\begin{aligned}(1-\phi L)\Delta y_t &= (1-\phi L)\mu + (1-\phi L)\eta_t + (1-L)\varepsilon_t \\ &= c + \eta_t + \phi\eta_{t-1} + \varepsilon_t - \varepsilon_{t-1}, \qquad c=(1-\phi)\mu.\end{aligned}$$

- They are unobserved but we have a sum of two iid series

$$\eta_t + \varepsilon_t + (-\phi)\eta_{t-1} + (-1)\varepsilon_{t-1}$$

- The sum of two white noise processes = white noise: same moments as MA(1).
- So this model is observationally equivalent to

$$\Delta y_t = c + \phi\Delta y_{t-1} + e_t + \theta e_{t-1}$$

- ARMA(1,1) $\Longrightarrow$ 4 parameters: $c, \phi, \theta, \sigma_e^2$, *that's how many we can estimate.*
- We have 5 parameters but only 4 observed. So far estimates assumes one of parameters fixed.

Estimation

- Assume $\sigma_{\varepsilon\eta} = 0$ (Watson, Harvey, Clark).

  => shocks that drive transitory movements are not correlated with those that drive long-run behavior.
- With this assumption the model can be estimated:
  1. Find match (functional) of observed/estimated parameters with the ones from structural model, or
  2. Cast the model in a state space form and estimate via Kalman Filter:

State-Space Form

Observation equation

$$y_t = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \tau_t \\ c_t \end{bmatrix} y_t = H\beta_t$$

State equation

$$\begin{aligned} \begin{bmatrix} \tau_t \\ c_t \end{bmatrix} &= \begin{bmatrix} \mu \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & \phi \end{bmatrix} \begin{bmatrix} \tau_{t-1} \\ c_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix}, \\ \beta_t &= \hat{\mu} + F\beta_{t-1} + e_t, \qquad e_t \sim N(0, Q), \\ & Q = \begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix} \end{aligned}$$

Kalman Filter: Results

- Kalman Filter does not care about how we came up with state form.
- KF: $\tau_{t|t}$ and $c_{t|t}$, $\tau_{t|T}$, $c_{t|T}$.
- We say τ_t and c_t are uncorrelated with each other, by assumption.
- $corr(\eta_{t|t}, \varepsilon_{t|t}) = -1$ even though we assume $corr(\eta_t, \varepsilon_t) = 0$.
- In classical approach $corr(x_t, \hat{\varepsilon}_t) = 0$ by construction, even though true relationship is $corr(x_t, \varepsilon_t) \neq 0$.
- Estimates of correlation rather than sample correlation of estimates.
- Identification: If we estimate the model without assuming $\sigma_{\varepsilon\eta}$ Gauss will not converge as there is ∞ many numbers of $\sigma_{\varepsilon\eta}$ for which likelihood doesn't decrease.

Morley, Nelson and Zivot (2003)

- RW + AR(2) makes model identified.

Why?

- AR(1) cycles is not observationally different from RW.
- AR(2) has this feature that cannot be proxied by RW.

Morley, Nelson and Zivot (2003):

- $\sigma_{\varepsilon\eta}$ identified for $c_t \sim ARMA(p,q)$, with $p \geq q+2$.

Example: AR(2)

Model:

$$\begin{aligned} y_t &= \tau_t + c_t \\ \tau_t &= \mu + \tau_{t-1} + \eta_t \\ c_t &= \phi_1 c_{t-1} + \phi_2 c_{t-2} + \varepsilon_t \end{aligned}$$

- 6 parameters: $\mu, \phi_1, \phi_2, \sigma_\eta^2, \sigma_\varepsilon^2, \sigma_{\varepsilon\eta}$.

Pre-multiplying both sides with $(1-L)$:

$$\begin{aligned} \Delta y_t &= (1-L)\tau_t + (1-L)c_t \\ &= \mu + \eta_t + (1-L)(1-\phi_1 L - \phi_2 L^2)^{-1}\varepsilon_t \\ (1-\phi_1 L - \phi_2 L^2)\Delta y_t &= (1-\phi_1-\phi_2)\mu + \eta_t - \phi_1\eta_{t-1} - \phi_2\eta_{t-2} + \varepsilon_t - \varepsilon_{t-1} \end{aligned}$$

- The model is observationally equivalent to ARMA(2,2) model: $\Delta y_t \sim ARMA(2,2)$ with 6 parameters: $c, \phi_1, \phi_2, \theta_1, \theta_2, \sigma_e^2$.

Results

- We can map parameters of ARMA(2,2) to our structural model or estimate KF with.

$$Q = \begin{bmatrix} \sigma_\eta^2 & \sigma_{\varepsilon\eta} \\ \sigma_{\varepsilon\eta} & \sigma_\varepsilon^2 \end{bmatrix}$$

- For US real GDP, setting $\sigma_{\varepsilon\eta} = 0$ can be rejected: $\rho_{\varepsilon\eta} = -0.9$.
- τ_t is volatile
- Structural model with ARMA(3) has 7 structural parameters but is observationally equivalent to reduced-form version ARMA(3,3) with 8 parameters: overidentification.
- Not such a big problem; $\rho_{\varepsilon\eta} < 0$ still holds.

Homework 1

1. Written Exercise
   For an AR(1) with $\phi = 0.7$, calculate and plot the true ACF and PACF. and the IRF for $j = 0, 1, \ldots, 5$.
   Do the same for an AR(2) with $\phi_1 = 1.1$ and $\phi_2 = -0.6$.
2. Eviews question

Homework 1

1. Autocorrelation Function, ACF:

$$\rho_j \equiv \frac{\gamma_j}{\gamma_0} \equiv corr(Y_t, Y_{t-j})$$

2. Partial Autocorrelation Function, PACF:
Regression coefficient (for the population) ϕ_{kk} in k-th order autoregression:

$$Y_t = c + \phi_1^{(k)} Y_{t-1} + \phi_2^{(k)} Y_{t-2} + \ldots + \phi_k^{(k)} Y_{t-k} + \varepsilon_t$$

3. Impulse Response Function, IRF :

$$\frac{\partial X_{t+1}}{\partial \varepsilon_t} = \psi_j$$

Homework 1

Question 1:

- For an AR(1) with $\phi = 0.7$, calculate and plot the true ACF and PACF and the IRF for $j = 0, 1, \ldots, 5$.

For AR(1), $Y_t = \phi Y_{t-1} + \varepsilon_t$:

1. ACF, $\rho_j = \phi \rho_{j-1} \implies \rho_j = \phi^j$.
2. PACF, $\phi_{11} = \phi$ and $\phi_{kk} = 0$ for $k > 1$.
3. IRF, $\psi_j = \phi^j$.

AR(1), $\phi = 0.7$

j	ACF(j)	PACF(j)	IRF(j)
0	1	.	1
1	0.7	0.7	0.7
2	0.49	0	0.49
3	0.343	0	0.343
4	0.240	0	0.240
5	0.168	0	0.168

Homework 1

Question 1:

- For an AR(2) with $\phi_1 = 1.1$ and $\phi_2 = -0.6$, calculate and plot the true ACF and PACF and the IRF for $j = 0, 1, \ldots, 5$.

1. ACF: Using definition, $\rho_1 = \rho_{-1}$, and Yule-Walker equation we get: $\rho_0 = 1$, $\rho_1 = \frac{\phi_1}{1-\phi_2}$, and $\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2}$ for $j > 1$.
2. PACF: the vector $\phi^{(k)}$ can be calculated from

$$\begin{bmatrix} \phi_1^{(k)} \\ \phi_2^{(k)} \\ \vdots \\ \phi_k^{(k)} \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{bmatrix}$$

Using $\rho_j = \frac{\gamma_j}{\gamma_0}$ and

$$\begin{bmatrix} \phi_1^{(2)} \\ \phi_2^{(2)} \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 \\ \gamma_1 & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} \frac{\rho_1 - \rho_1\rho_2}{1-\rho_1^2} \\ \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} \end{bmatrix},$$

$\phi_1^{(1)} = \frac{\gamma_1}{\gamma_0}$, $\phi_2^{(2)} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}$ and $\phi_k^{(k)} = 0$ for $k > 2$.

Homework 1

Question 1:

- For an AR(2) with $\phi_1 = 1.1$ and $\phi_2 = -0.6$, calculate and plot the true ACF and PACF and the IRF for $j = 0, 1, \ldots, 5$.

3. IRF: According to Yule-Walker equation for AR(2)

$$\psi_j = \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2}$$

or recursively solving Y_t

$$\begin{aligned} Y_t &= (\phi_1^5 + 4\phi_1^3\phi_2 + 3\phi_1\phi_2^2)Y_{t-5} + (\phi_1^4\phi_2 + 3\phi_1^2\phi_2^2 + \phi_2^3)Y_{t-6} \\ &\quad + (\phi_1^4 + 3\phi_1^2\phi_2 + \phi_2^2)\varepsilon_{t-4} + (\phi_1^3 + 2\phi_1\phi_2)\varepsilon_{t-3} \\ &\quad + (\phi_1^2 + \phi_2)\varepsilon_{t-2} + \phi_1\varepsilon_{t-1} + \varepsilon_t \end{aligned}$$

Homework 1

Question 1:

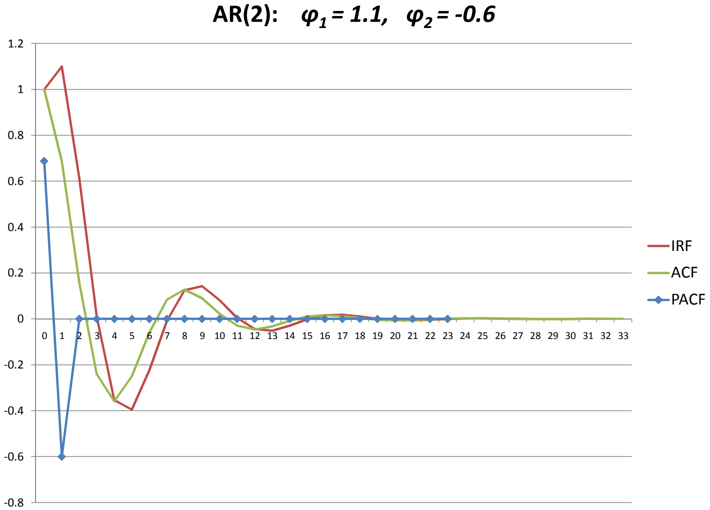
- For an AR(2) with $\phi_1 = 1.1$ and $\phi_2 = -0.6$, calculate and plot the true ACF and PACF and the IRF for $j = 0, 1, \ldots, 5$.

AR(2), $\phi_1 = 1.1$, $\phi_2 = -0.6$

j	ACF	PACF	IRF
0	1	.	1
1	0.688	0.688	1.1
2	0.156	-0.6	0.61
3	-0.241	0	0.011
4	-0.358	0	-0.354
5	-0.250	0	-0.396
6	-0.060	0	-0.223

Homework 1

Question 1:



Homework 2

Question 1:

- Hai, Mark and Wu's (1996) common stochastic trend model:

$$
\begin{aligned}
y_{1,t} &= z_t + x_{1,t}, \\
y_{2,t} &= z_t + x_{2,t}, \\
z_t &= z_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iidN(0, \sigma_\varepsilon^2), \\
x_{1,t} &= \mu_1 + \phi_{11} x_{1,t-1} + \phi_{12} x_{1,t-1} + e_{1,t} + \theta_{11} e_{1,t-1} + \theta_{12} e_{2,t-1}, \\
x_{2,t} &= \mu_2 + \phi_{21} x_{1,t-1} + \phi_{22} x_{1,t-1} + e_{2,t} + \theta_{21} e_{1,t-1} + \theta_{22} e_{2,t-1}, \\
& \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \sim iid\, N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \right),
\end{aligned}
$$

Homework 2: Q1

The states-space representation for this model is:

- Measurement equation

$$y_t = H\beta_t$$

where,

$$y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix}, \quad \beta_t = \begin{bmatrix} z_t \\ x_{1,t} \\ x_{2,t} \\ e_{1,t} \\ e_{2,t} \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix},$$

Homework 2: Q1

- State equation

$$\beta_t = \mu + F\beta_{t-1} + Gu_t,$$

where

$$\mu = \begin{bmatrix} 0 \\ \mu_1 \\ \mu_2 \\ 0 \\ 0 \end{bmatrix}, \quad F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \phi_{11} & \phi_{12} & \theta_{11} & \theta_{12} \\ 0 & \phi_{21} & \phi_{22} & \theta_{21} & \theta_{22} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad u_t = \begin{bmatrix} \varepsilon_t \\ e_{1t} \\ e_{2t} \end{bmatrix}.$$