Class 3: Unit Root

Macroeconometrics - Spring 2011

Jacek Suda, BdF and PSE

November 16, 2011

Outline

Outline:

1. Prediction Error Decomposition
2. State-space Form
3. Kalman Filter
4. Unit Root
    - Dickey-Fuller Test
    - Phillips-Perron
    - Stationarity Tests
    - Variance Ratio Test
5. Structural Breaks

Kalman Filter

- Forecasts based on Wold form assume infinite number of observations.
- *We don't have them in reality.*
- Kalman filter calculates linear projections for finite number of observations,
  - exact finite sample forecast,
  - allow for exact MLE of ARMA models based on Prediction Error Decomposition.

Normal Distribution

Joint Normality:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} = \tilde{y}_{T_{T\times 1}} \sim N(\mu_{T\times 1}, \Omega_{T\times T}),$$

Since it is covariance stationary process, each Y_t has the same mean and variance, $\omega_{11} = \sigma^2 = \omega_{22} = \omega_{TT}$.

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12} & & \omega_{1T} \\ \omega_{21} & \omega_{22} & & \vdots \\ \vdots & & \ddots & \\ \omega_{T1} & & & \omega_{TT} \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & & \\ \vdots & & \ddots & \\ \gamma_{T-1} & & & \gamma_0 \end{bmatrix}, \text{ as } \begin{matrix} \omega_{jj} = \gamma_0, \\ \omega_{ij} = \gamma_{i-j}, \\ i > j. \end{matrix}$$

The likelihood function:

$$L(\tilde{\theta}|\tilde{y}_T) = (2\pi)^{-\frac{T}{2}} \det(\Omega)^{-\frac{1}{2}} e^{-\frac{1}{2}(\tilde{y}_T - \mu)'\Omega^{-1}(\tilde{y}_T - \mu)}.$$

Factorization

- *For large T Ω might be large and difficult to invert.*
- Since Ω is positive definite symmetric matrix then there exists a unique, triangular factorization of Ω,

$$\Omega = AfA',$$

where

$$f_{T\times T} = \begin{bmatrix} f_1 & 0 & \cdots & 0 \\ 0 & f_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & f_T \end{bmatrix}, \qquad f_t > 0\ \forall t \quad \textit{diagonal matrix}$$

$$A_{T\times T} = \begin{bmatrix} 1 & & & 0 \\ a_{21} & 1 & & \\ \vdots & & \ddots & \\ a_{T1} & a_{T2} & \dots & 1 \end{bmatrix}.$$

Likelihood

The likelihood function can be rewritten as:

$$L(\tilde{\theta}|\tilde{y}_T) = (2\pi)^{-\frac{T}{2}} \det(AfA')^{-\frac{1}{2}} e^{-\frac{1}{2}(\tilde{y}_T-\mu)'(AfA')^{-1}(\tilde{y}_T-\mu)}$$

Define

$$\eta = A^{-1}(\tilde{y}_T - \mu)(\textit{prediction error}).$$

where

$$A\eta = (\tilde{y}_T - \mu).$$

Since A is lower-triangular matrix with $1s$ along the principal diagonal,

$$\begin{aligned}
\eta_1 &= y_1 - \mu \\
\eta_2 &= y_2 - \mu - a^*_{21}\eta_1 \\
\eta_3 &= y_3 - \mu - a^*_{31}\eta_1 - a^*_{32}\eta_2 \\
&\vdots \\
\eta_T &= y_T - \mu - \sum_{i=1}^{T-1} a^*_{Ti}\eta_{T-1}
\end{aligned}$$

Likelihood

Also, since A is lower triangular with 1s along the principal diagonal, $\det(A) = 1$

$$\det(AfA) = \det(A) \cdot \det(f) \cdot \det(A') = \det(f).$$

Then,

$$\begin{aligned} L(\tilde{\theta}|\tilde{y}_T) &= (2\pi)^{-\frac{T}{2}} \det(f^{-1})^{-\frac{1}{2}} e^{-\frac{1}{2}\eta'(f^{-1})^{-1}\eta} \\ &= \prod_{t=1}^{T} \left(\frac{1}{\sqrt{2\pi f_t}} e^{-\frac{1}{2}\frac{\eta_t^2}{f_t}} \right), \end{aligned}$$

where η_t is t^{th} element of $\eta_{T\times 1}$ = prediction error $y_t - \hat{y}_{t|t-1}$,

$$\hat{y}_{t|t-1} = \sum_{i=1}^{t-1} a^*_{t,i} y_i, \qquad i = 2, 3, \ldots, T,$$

where $a^*_{t,i}$ is $(t, i)^{th}$ element of A^{-1}.

Kalman Filter

Note: Given $y_t \sim N(\mu, \Omega)$,

$$\eta_t | \Omega_{t-1} \sim N(0, f_t),$$

where f_t is an (t, t) diagonal element of f matrix,

$$\ln L = -\frac{1}{2} \sum_{t=1}^{T} \ln(2\pi f_t) - \frac{1}{2} \sum_{t=1}^{T} \frac{\eta_t^2}{f_t},$$

since $\eta_t \sim N$ and independent of each other.

- The Kalman filter recursively calculates linear projection of y_t on past information Ω_{t-1} for any model that can be cast in state-space form.
- *Kalman filter: for any structure it solves for linear prediction.*

Measurement (Observation) Equation

General form that encompasses a wide variety od models.

1. Measurement (Observation) Equation
   - *Represent the static relationship between observed variables (data) and unobserved state variables.*

$$y_t = H_t\beta_t + Az_t + e_t,$$

where y_t denotes observed data, β_t is a state vector that captures the dynamics, z_t is exogenous, observed variables *for example, lagged values of y_t but also other data*, and e_t is an error term,

$$e_t \sim N(0, R).$$

The existence of the state vector makes this representation not a simple linear model.

Transition (State) Equation

2. Transition (State) Equation
   - *Captures the dynamics in the system, causes the system to go on and on.*

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

where $\tilde{\mu}$ is a vector of constants, F is the transition matrix, and v_t is an error vector,

$$v_t \sim N(0, Q).$$

   - *Like AR(1) but in vector/matrix form.*

Transition (State) Equation

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

- The state vector has and AR(1) kind of representation.
- Describes evolution of state vector.
- These state vectors can be unobservable.
- Transition equation can be used to get information about the unobservable, conditioning on data which is observable (Bayesian).

Error terms

Error terms:

$$e_t \sim N(0, R), \quad v_t \sim N(0, Q),$$

where R, Q are var-cov matrices and

$$E[e_t v_\tau'] = 0, \quad \forall t, \tau$$

- Restrictive assumption
- The model can be represented in a way that is not very restrictive.
- Even with $E[e_t v_\tau'] \neq 0$ we can estimate the model with (modified) Kalman Filter but it becomes more complicated.
- The normality assumption might not be good as always but

  ... It allows to use MLE.

Examples: AR(p)

- *It applies to a very wide variety of time-series models.*

Consider an AR(p) process

$$\begin{aligned} y_t - \mu &= \phi_1(y_{t-1} - \mu) + \ldots + \phi_p(y_{t-p} - \mu) + \varepsilon_t \\ E(\varepsilon_\tau \varepsilon_t) &= \sigma^2 \qquad \text{for } t = \tau \end{aligned}$$

State equation

$$\begin{bmatrix} y_t - \mu \\ y_{t-1} - \mu \\ \vdots \\ y_{t-p+1} - \mu \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} - \mu \\ y_{t-2} - \mu \\ \vdots \\ y_{t-p} - \mu \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Observation equation

$$y_t = \mu + \begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix} \begin{bmatrix} y_t - \mu \\ y_{t-1} - \mu \\ \vdots \\ y_{t-p+1} - \mu \end{bmatrix}.$$

Examples: ARMA(1,1)

ARMA(1,1):
Set $\mu = 0$,

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}, \qquad \varepsilon_t \sim N(0, \sigma^2).$$

- *There might be more than one way to represent a model in a state-space form.*
- *There might be differences in efficiency between different ways.*

Examples: ARMA(1,1)

State equation:

- The general form

$$\beta_t = F\beta_{t-1} + v_t.$$

- Put

$$\beta_t = \begin{bmatrix} y_t \\ \varepsilon_t \end{bmatrix} \Rightarrow \beta_{t-1} = \begin{bmatrix} y_{t-1} \\ \varepsilon_{t-1} \end{bmatrix}$$

- Put $y_t = \phi y_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t$ in a matrix notation:

$$\underbrace{\begin{bmatrix} y_t \\ \varepsilon_t \end{bmatrix}}_{\beta_t} = \underbrace{\begin{bmatrix} \phi & \theta \\ 0 & 0 \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} y_{t-1} \\ \varepsilon_{t-1} \end{bmatrix}}_{\beta_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_t \\ \varepsilon_t \end{bmatrix}}_{v_t},$$

and $v_t \sim N(0, Q)$, $Q = \begin{bmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 \end{bmatrix}$.

y_t – *observable*, ε_t–*unobservable, forecast error*

Examples: ARMA(1,1)

Observation equations:

$$\underset{y_t}{y_t} = \underset{H}{\begin{bmatrix} 1 & 0 \end{bmatrix}} \underset{\beta_t}{\begin{bmatrix} y_t \\ \varepsilon_t \end{bmatrix}}$$

no exogenous variables: $A = 0$, also $R = 0$.

$y_t = H\beta_t$ for this case (ARMA(1,1)).

The parameters ϕ, θ, σ^2 are captured in F, Q matrices. The Kalman Filter will estimate them.

- *For KF what goes in β_t doesn't matter.*
- *Only parameters F, Q, H, R will matter.*
- *The state vector is now defined by F, Q, H, and the observations.*

ARMA(1,1): Alternative Representation

A more “elegant” (i.e. easier for computation) representation.
Lag notation (alternative representation for ARMA(1,1))

$$\begin{aligned} (1-\phi L)y_t &= (1+\theta L)\varepsilon_t \\ y_t &= (1-\phi L)^{-1}(1+\theta L)\varepsilon_t \\ y_t &= (1+\theta L)(1-\phi L)^{-1}\varepsilon_t. \end{aligned}$$

Define $x_t = (1-\phi L)^{-1}\varepsilon_t$

$$\begin{aligned} (1-\phi L)x_t &= \varepsilon_t, \qquad (x_t \text{ is AR}(1),\ \textit{not observed}) \\ x_t - \phi x_{t-1} &= \varepsilon_t \end{aligned}$$

Then,

$$\begin{aligned} y_t &= (1+\theta L)x_t \\ y_t &= x_t + \theta x_{t-1}. \end{aligned}$$

So y_t is a linear combination of 2 unobservable AR(1) processes, x_t and x_{t-1}.

ARMA(1,1): State-Space

Observation equation *(all randomness in the state equation)*

$$y_t = H\beta_t,$$

where

$$y_t = \begin{bmatrix} 1 & \theta \end{bmatrix} \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix}$$

- *Inside H there are parameters to be estimated.*
- $A = 0$*, no exogenous,* $R = 0$ *as the observable equation is just the identity (no randomness of* e_t*).*

ARMA(1,1): State-Space

State equation

$$\begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix},$$

$$\beta_t \qquad F \qquad \beta_{t-1} \qquad v_t$$

so

$$v_t \sim N(0, Q), \qquad Q = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

- *So ϕ is in F, θ in H, and σ^2 in Q.*

Given F, Q, H, A, R and data (y_t's), use Kalman Filter to find prediction error decomposition of joint likelihood for $\tilde{y}_T = (y_1, \ldots y_T)$, given by $L(\theta, \phi, \sigma^2 | \tilde{y}_T)$. *(exact likelihood)*

Kalman Filter

Kalman filter:

- *purpose: to make inference about unobservable given the observable,*
- *application: signal extraction in engineering,*
- *economics: don't know the parameters F, Q, H and want to estimate them.*

State-space form

ME: Measurement (Observation) equation:

$$y_t = H\beta_t + e_t, \qquad e_t \sim N(0, R)$$

SE: Transition (State) equation:

$$\beta_t = \tilde{\mu} + F\beta_{t-1} + v_t, \qquad v_t \sim N(0, Q),$$

$$E[e_t v_\tau'] = 0.$$

Mean of β

1. β_t is a random variable
   - it might be unobservable and no data for it,
   - it is normal random variable as it is sum of normal variables, $v_t \sim N$.

Conditional mean

$$\beta_t|\Omega_{t-1} \sim N\left(E[\beta_t|\Omega_{t-1}], var(\beta_t|\Omega_{t-1})\right)$$
$$E[\beta_t|\Omega_{t-1}] = \beta_{t|t-1}, \qquad \textit{conditional expectations.}$$

- *We may not know what β's are.*
- *If we have information about its distribution, we can calculate mean, variance, etc.*
- *β_{t-1} may be not observable: take expectations of it*

$$E[\beta_t|\Omega_{t-1}] \equiv \beta_{t|t-1} = \tilde{\mu} + FE[\beta_{t-1}|\Omega_{t-1}] + 0$$
$$\beta_{t|t-1} = \tilde{\mu} + F\beta_{t-1|t-1},$$

- *In AR(1): $E[y_t] = \mu + \phi E[y_{t-1}]$, last term is observable.*

Variance of β

Conditional variance

$$Var(\beta_t|\Omega_{t-1}) \equiv P_{t|t-1} = E[(\beta_t - \beta_{t|t-1})(\beta_t - \beta_{t|t-1})'].$$

Recall

$$var(ax) = a^2 var(x), \qquad a - scalar, x - random\ vector.$$

Two sources of randomness (variation) for β_t:

1. v_t is a random variable,
2. β_{t-1} is also random so there might be difference between β_{t-1} and $\beta_{t|t-1}$, *there may not be equal to each other.*

$$P_{t|t-1} = F\, P_{t-1|t-1}\, F' + Q,$$

where $P_{t|t-1}$, uncertainty about β_t equals sum of uncertainty about β_{t-1}, $P_{t-1|t-1}$, and uncertainty about v_t.
Note: $cov(\beta_{t-1}, v_t) = 0$.

Kalman Filter

2. y_t is a random variable.
    - Now, we have data on y_t.
    - We have some joint density of y_t, β_t and some prior.
    - Using data we get posterior of β_t.

    - *We want to make inference for β_t which we don't observe.*
    - *We see y_t which is related to β_t.*
    - *We make inferences on β_t by observing joint density (distribution) of ys and βs (Bayesian view).*

Distribution of y_t

Distribution of y_t given state-space

$$y_t|\Omega_{t-1} \sim N(E[y_t|\Omega_{t-1}], var(y_t|\Omega_{t-1}),$$

- Conditional mean

$$E[y_t|\Omega_{t-1}] \equiv y_{t|t-1} = H\beta_{t|t-1} + 0$$

- Conditional variance

$$var(y_t|\Omega_{t-1}) \equiv f_{t|t-1} = HP_{t|t-1}H' + Q,$$

since we don't know β_t.

- Note: $cov(H\beta_t, e_t) = 0$ because $E[v_t e_t] = 0$.

If $E[v_t e_t] \neq 0$ *we will add another term in the* $var(y_t|\Omega_{t-1})$ *capturing that.*

Joint Distribution

- Covariance between β_t and y_t:

$$cov(y_t, \beta_t|\Omega_{t-1}) = P_{t|t-1}H',$$

as $cov(H\beta_t + e_t, \beta_t) = cov(H\beta_t, \beta_t) + cov(e_t, \beta_t) = cov(\beta_t, \beta_t)H' + 0.$

Then, the joint distribution for y_t and β_t is joint normal:

$$\begin{array}{c|c} \beta_t & \\ y_t & \end{array} \Omega_{t-1} \sim N\left(\begin{bmatrix} \beta_{t|t-1} \\ H\beta_{t|t-1} \end{bmatrix}, \begin{bmatrix} P_{t|t-1} & P_{t|t-1}H' \\ P_{t|t-1}H' & f_{t|t-1} \end{bmatrix} \right).$$

Kalman Filter

Two steps of Kalman Filter :

(a) Prediction,

(b) Given y_t updating inference on β_t.

Definition

Given $\beta_{0|0}$, $P_{0|0}$, Kalman Filter solves the following six equations for $i = 1, \ldots, T$

Prediction of y_t, β_t

$$(1) \quad \beta_{t|t-1} = \tilde{\mu} + F\beta_{t-1|t-1},$$

$$(2) \quad P_{t|t-1} = F\,P_{t-1|t-1}\,F' + Q,$$

Forecast error:

$$(3) \quad \eta_{t|t-1} \equiv y_t - y_{t|t-1} = y_t - H\beta_{t|t-1},$$

Variance of forecast error:

$$(4) \quad f_{t|t-1} = H\,P_{t|t-1}\,H' + R$$

Updating of y_t, β_t

$$(5) \quad \beta_{t|t} = \beta_{t|t-1} + \kappa_t \eta_{t|t-1},$$

$$(6) \quad P_{t|t} = P_{t|t-1} - \kappa_t H P_{t|t-1},$$

$$\kappa_t \equiv P_{t|t-1} H' f_{t|t-1}^{-1} \quad \text{“Kalman gain”}.$$

Kalman Filter

- $\beta_{0|0}$, $P_{0|0}$, are equal to unconditional mean and variance, and reflect prior beliefs.
- Equation (5) is a linear combination of previous guess and forecast error.

$$
\begin{aligned}
(5) \qquad \beta_{t|t} &= \beta_{t|t-1} + \kappa_t \eta_{t|t-1}, \\
(6) \qquad P_{t|t} &= P_{t|t-1} - \kappa_t H P_{t|t-1}, \\
\kappa_t &\equiv P_{t|t-1} H' f_{t|t-1}^{-1} \quad \text{“\underline{Kalman gain}”.}
\end{aligned}
$$

- *The Kalman gain depends on the relationship between y_t and β_t since $P_{t|t-1}H' = cov(\beta_t, y_t)$ and $f_{t|t-1}^{-1}$ is the precision of the forecast error.*
- *The bigger the variance of forecast error the smaller the Kalman gain and less weight put to updating.*
- *Equation (6) measures conditional variance.*
- *Since we observe y_t the uncertainty declines.*

Kalman Gain

- The stronger the covariance between y_t and β_t, the more we will update when we see high forecast error.
- If the relationship is weaker, we don't put much weight as probably it is not driven by β_t.
- The weight depends on the variance of forecast error: if f^{-1} big, put high weight on that observations.
- Once we have $\eta_{t|t-1}, f_{t|t-1}$, we can do MLE after constructing the joint likelihood of prediction error decomposition.

UNIT ROOT

Random Walk - Example

Basic random walk

$$y_t = y_{t-1} + \varepsilon_t; \qquad \varepsilon_t \sim WN$$

Note the property

$$E_t y_{t+1} = y_t.$$

- random walks are popular in finance...
  - models for asset prices, exchange rates
- ...as well as in macroeconomics
  - permanent income hypothesis
  - unit root in macro variable: Nelson and Plosser (1982)

Random Walk: Properties

Properties of random walks:

1. The impulse-response function of a random walk is one at all horizons.

$$y_t = y_{t-1} + \varepsilon_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$$

   - The impulse-response function of stationary processes dies out eventually.

2. The forecast variance of the random walk grows linearly with the forecast horizon

$$var(y_{t+k}|y_t) = var(y_{t+k} - y_t) = k\sigma^2$$

   - forecast error variance of a stationary series approaches a constant, the unconditional variance of that series.
   - the variance of the random walk is infinite.

Random Walk: Properties

Properties of random walks:

3. The autocovariances of a random walk, seen as the limit of an AR(1), $y_t = \phi y_{t-1} + \varepsilon_t$, as $\phi \to 1$.

$$\rho_j = 1 \qquad \text{for all } j$$

   - All estimated autocorrelations are near 1; they die out "too slowly".

4. The variance of a random walk is primarily due to low-frequency components.

Stochastic Trends

Until the late 70s:

- fit a linear trend to log GNP (by OLS), and then
- define the stochastic part of the time series as deviations from this trend.

Problem in late 70s:

- The "trend", "potential" GNP growth rate slowed down.
- Slowdown was not foreseen.
- Hard to think about more complex deterministic trends, (e.g. polynomials)

Approach to deal with it:

- stochastic trends,
- random-walk type processes

They give a convenient representation of such trends since they wander around at low frequencies.

Permanent Shocks

Are there only temporary shocks:

- Business cycles seen as (by assumption) the stationary deviation about the time trend
- But: shocks to GNP might not more closely resemble the permanent shocks of a random walk than the transitory shocks of the old AR(2) about a linear trend.
- Classic: Nelson and Plosser (1982) test macroeconomic time series for unit roots.
- They found they could not reject unit roots in most time series.
- Evidence for the possibility of long run movements in time-series.
- Also: financial economics and the question of whether stock returns are less than perfect random walks...

  ... random walk test considered as convincing evidence about “efficient markets”.

Statistical issues: Distribution of AR(1) estimates

Recall

$$y_t = y_{t-1} + \varepsilon_t; \qquad \varepsilon_t \sim WN$$

Dickey and Fuller:

- Test for a random walks by running $y_t = \phi y_{t-1} + \varepsilon_t$ and testing whether $\phi = 1$ not correct:
  1. OLS estimates are biased down (towards stationarity)
  2. OLS standard errors are tighter than the actual standard errors
- Many series thought to be stationary based on OLS regressions could be in fact generated by random walks.

Statistical issues: Inappropriate detrending

- Suppose the real model is

$$y_t = c + y_{t-1} + \varepsilon_t; \qquad \varepsilon_t \sim WN$$

- Suppose you detrend by OLS, and then estimate an AR(1), i.e., fit the model

$$y_t = \alpha + \beta \cdot t + \phi y_{t-1} + \varepsilon_t$$

- OLS estimate $\hat{\phi}$ even more biased downward and standard errors more misleading.

Why?

- In a relatively small sample, the random walk is likely to drift up or down;
- Drift could well be (falsely) modeled by a linear (or nonlinear, “breaking” , etc.) trend.

Statistical issues: Spurious regression

Suppose two series are generated by independent random walks,

$$
\begin{aligned}
x_t &= x_{t-1} + \epsilon_t \\
y_t &= y_{t-1} + \nu_t \\
& E(\epsilon_t \nu_t) = 0 \quad \text{for all} \ \ t, s
\end{aligned}
$$

Suppose we run y_t on x_t by OLS,

$$
yt = \alpha + \beta x_t + \upsilon_t
$$

- Assumptions behind the usual distribution theory are violated.
- We find statistically significant β more often than the we should.

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

- If one of the roots is equal to one, it can be factored out

$$\phi(z) = (1 - z)\phi^*(z),$$

$\Rightarrow \phi^*(z) = 0$ has roots outside unit circle

$$\begin{aligned} \Rightarrow \phi^*(L)(1 - L)y_t &= \phi(L)\varepsilon_t \\ \Delta y_t &= \phi^*(L)^{-1}\theta(L)\varepsilon_t \\ \Delta y_t &= \Psi^*(L)\varepsilon_t \\ \Delta y_t &= u_t, \qquad u_t = \Psi^*(L)\varepsilon_t \sim I(0) \end{aligned}$$

- n in $I(n)$ denotes order of integration.
- $I(0)$ denotes covariance stationary process.

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 $H_0: \quad \phi = 0$,
$$t_{\phi=0} = \frac{\hat{\phi} - 0}{\hat{SE}(\hat{\phi})} \overset{A}{\sim} N(0,1)$$
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 $\phi = 1$ these nuisance parameter 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 - it's 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

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

- 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, but 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.

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

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

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

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

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

$$\sigma^2 = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E(u_t^2)$$

$$\lambda^2 = \lim_{T\to\infty} \sum_{t=1}^{T} E(T^{-1} S_T^2) = \text{“long run variance”}$$

$$S_T = \sum_{t=1}^{T} u_t.$$

- 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

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

μ_t is unobserved component.

- it takes both cases of unit root and stationarity
- μ_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:

$$H_0: \quad \sigma_u^2 = 0 \implies y_t \sim I(0)$$
$$H_1: \quad \sigma_u^2 > 0 \implies y_t \sim I(1)$$

- 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:

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

- 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 hypothesis

$$H_0 : \qquad \sigma_u^2 = 0 \qquad \text{no random walk component, just constant} H_1 :$$

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

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

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 $\hat{\eta}_\tau > 0.146$.

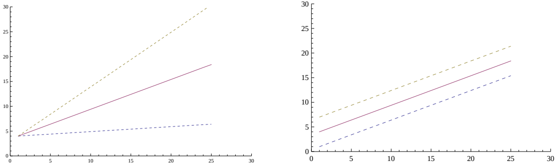
Testing

- 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

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

- Result

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

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

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

- Variance

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

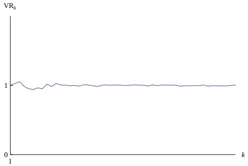
- Variance ratio

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

- 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 $\hat{\Lambda}_{NW}^2, \hat{\gamma}_0^*, \hat{\gamma}_j^*$.

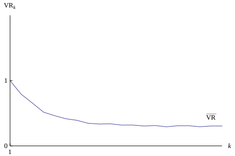
Implications

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



   - random walk, all shocks are permanent

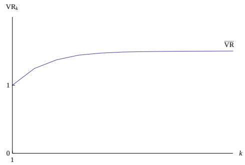
2. If $\widehat{VR}_k \to 0 < \overline{VR} < 1$



   - “mean reverting” I(1) process

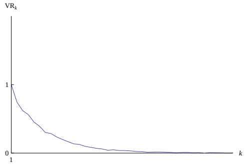
Implications

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

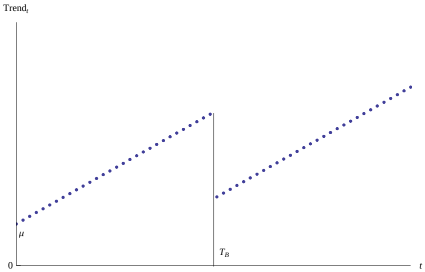
Trend stationary	$\Longrightarrow$	never
TS with one structural break	$\Longrightarrow$	once
TS with n structural breaks	$\Longrightarrow$	n times
TS with T structural breaks	$\Longrightarrow$	Unit root (difference stationary)

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

- Given a level a level breal 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

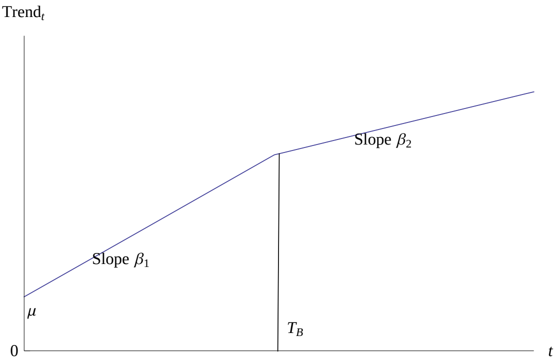


- Model of trend fluctuations of GDP.
- Trend is deterministic, when removed one gets the covariance stationary series.

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

Model B: The "Oil Shock" Model

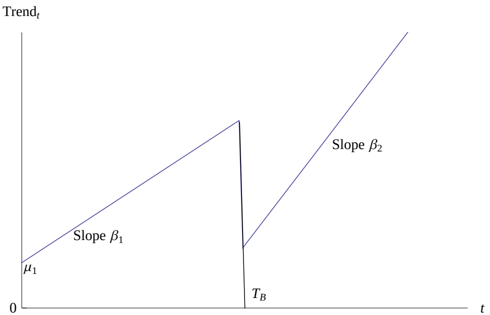
Model B: The "Oil Shock" Model



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

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

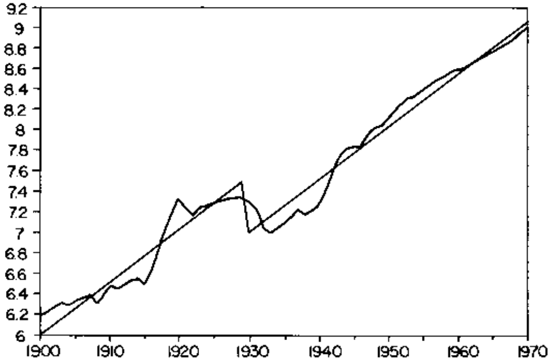
$$\hat{t}_{\rho=1}(\lambda) = \frac{\hat{\rho}(\lambda) - 1}{\widehat{SE}(\hat{\rho}(\lambda))}$$

  - 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.

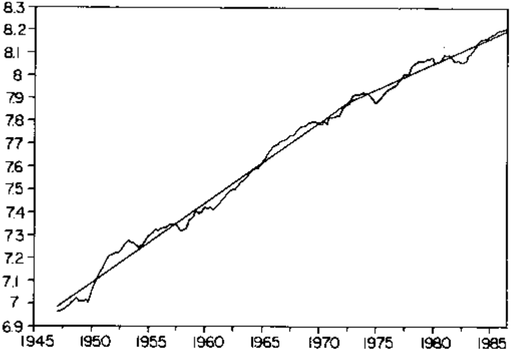
Cochrane (1989)



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."

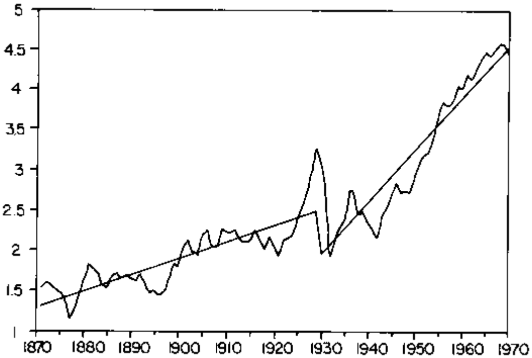
Cochrane (1989)



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:I and $DT_t^* = t - T_B$ if $t >$ 1973: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% lee0 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.