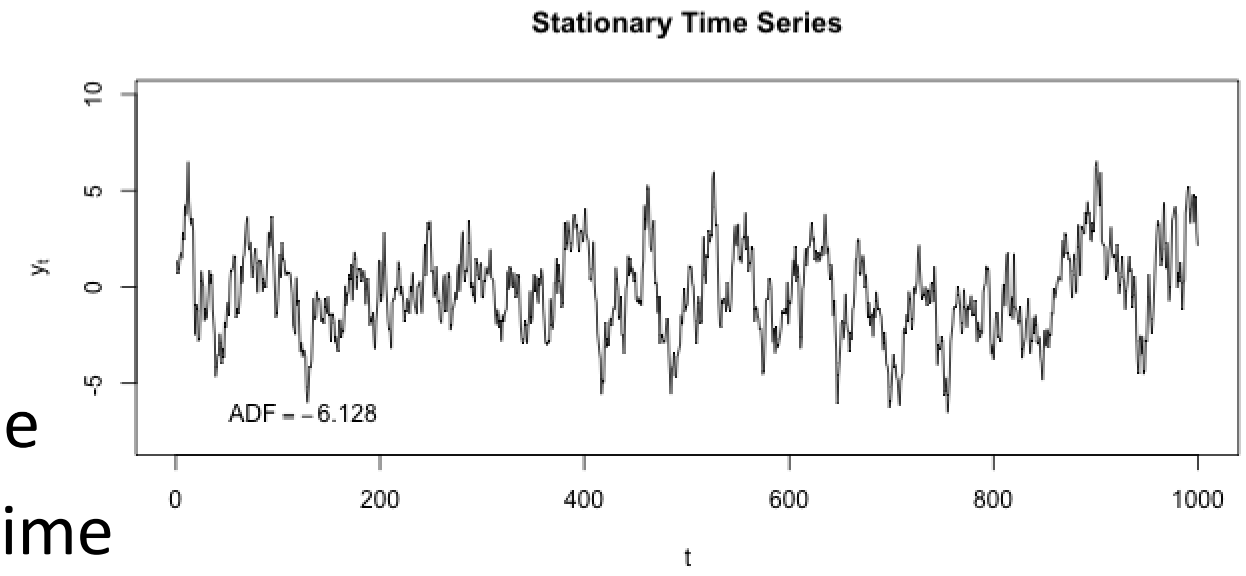
Time Series for Dummies

Pyry Lehtonen

7.3.2017

(Weakly) Stationarity

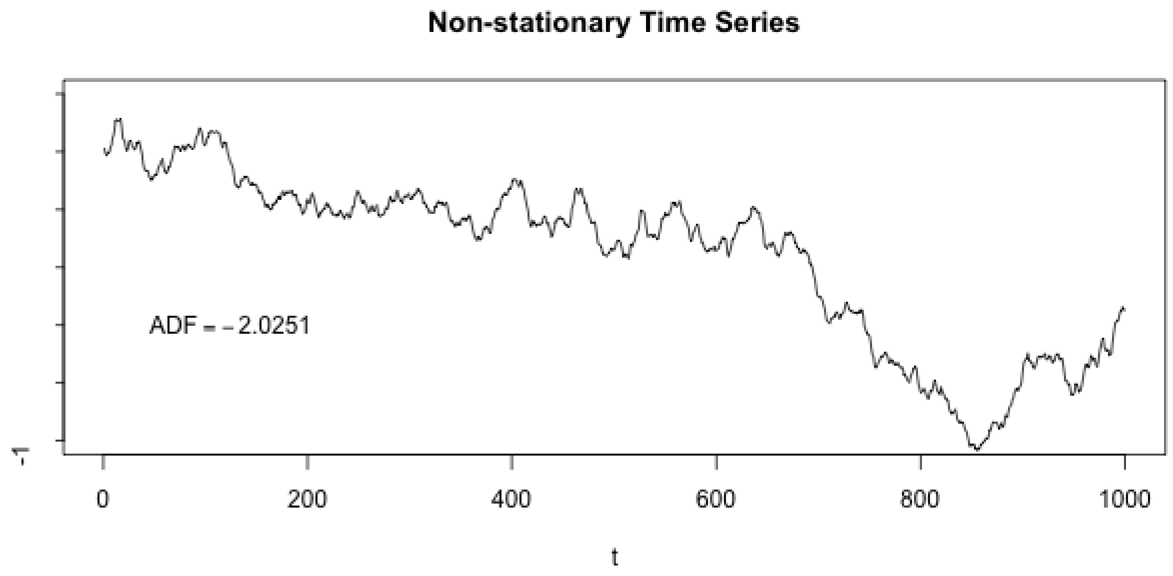
- **Mean** do not change over time
- **Variance** do not change over time
- **Covariance** do not change over time



$$E(y_t) = \mu$$

$$\mathrm{Var}(y_t) = E\left[(y_t - \mu)^2\right] = \gamma_0$$

$$\mathrm{Cov}(y_t, y_{\boxed{t-k}}) = E\left[(y_t - \mu)(y_{t-k} - \mu)\right] = \gamma_k$$



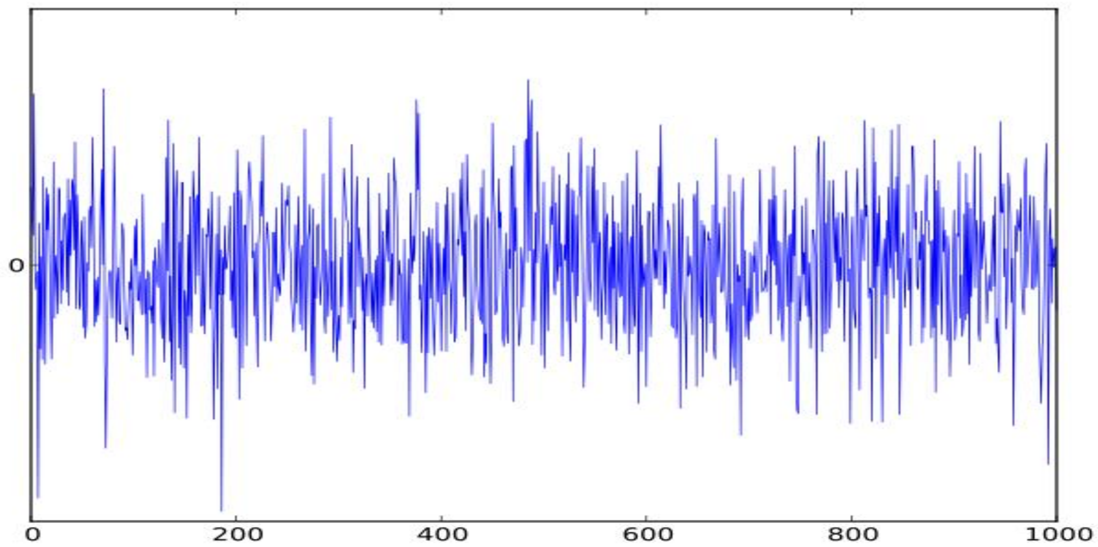
White Noise

- Error term ε_t is white noise (=random signal) if:
- Expected value is 0
- Variance is constant = homoskedasticity
- No autocorrelation

$$E(\varepsilon_t) = 0$$

$$E(\varepsilon_t^2) = \sigma_\varepsilon^2$$

$$E(\varepsilon_t \varepsilon_s) = 0,\ t \neq s$$



Moving Average MA(q) Model

$$y_t = \varepsilon_t + \alpha_1\varepsilon_{t\text{-}1} + \alpha_2\varepsilon_{t\text{-}2} + \cdots + \alpha_q\varepsilon_{t\text{-}q}.$$

- q tells the order of MA(q) model -> how many lags is there
- Exogenous shock lasts only current period and q periods in the future
- y_t depends on error terms -> past values of y do not matter
- $\mathcal{E}_t$ is white noise -> MA(q) is **always stationary**

Autoregressive AR(p) Model

$$y_t = \theta_1 y_{t\text{-}1} + \theta_2 y_{t\text{-}2} + \cdots + \theta_p y_{t\text{-}p} + \varepsilon_t$$

- y_t depends linearly on its previous values
- If AR(p) is not stationary, exogenous shock lasts forever
- If AR(p) is stationary, exogenous shock converges to 0
- Not always stationary, may have a unit root
- Stationarity depends on roots of the lag polynomial

Lag Polynomial

AR(p) model

$$\theta(L)y_t = \varepsilon_t.$$

$$\theta(L) = 1 - \theta_1 L.$$

- Solve roots of L
- If *all roots* are greater than one in absolute value -> AR(p) is stationary

MA(q) model

$$y_t = \alpha(L)\varepsilon_t.$$

$$\alpha(L) = 1 + \alpha_1 L + \alpha_2 L^2 + \ldots + \alpha_q L^q.$$

- MA(q) is always stationary regardless of the roots

Impulse Response Function (IRF)

- IRF is a measure of persistence
  - The more persistence the time series is, the longer will exogenous shock last
  - At the lag where IRF gets value of 0 -> shock do not have effect anymore

Cheat Sheet for Solving IRF

Example is for AR(p) model but same goes for MA(q) model, only lag polynomial is different

- Construct the lag polynomial $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_p L^p$
- Insert polynomial: $\psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots$
- To equation $\psi(L)\theta(L) = 1$
- So we have: $(\psi_0 + \psi_1 L + \psi_2 L^2 + \cdots)(1 - \theta_1 L$ …)=1
- Open the brackets (up to lags that are under interest)
- Set $\psi_0 = 1$, so all the other parameters will be equal to zero
- Solve ψ_1 : take the sum of all polynomials where L^1 exists -> set equal to 0 and solve ψ_1
- Solve ψ_2 : take the all polynomials where L^2 exists -> set equal to 0 and solve ψ_2
- And so on and so forth…

Representations in Other Models

- AR(p) has a MA(∞) representation if AR(p) is stationary
- MA(q) has an AR(∞) representation if its lag polynomial roots are greater than one in absolute value
  - This is called *invertibility*
- **When lag polynomial roots $|L| > 1$ representation can be done**

Autoregressive-Moving-Average ARMA(p,q)

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \cdots + \alpha_q \varepsilon_{t-q}$$

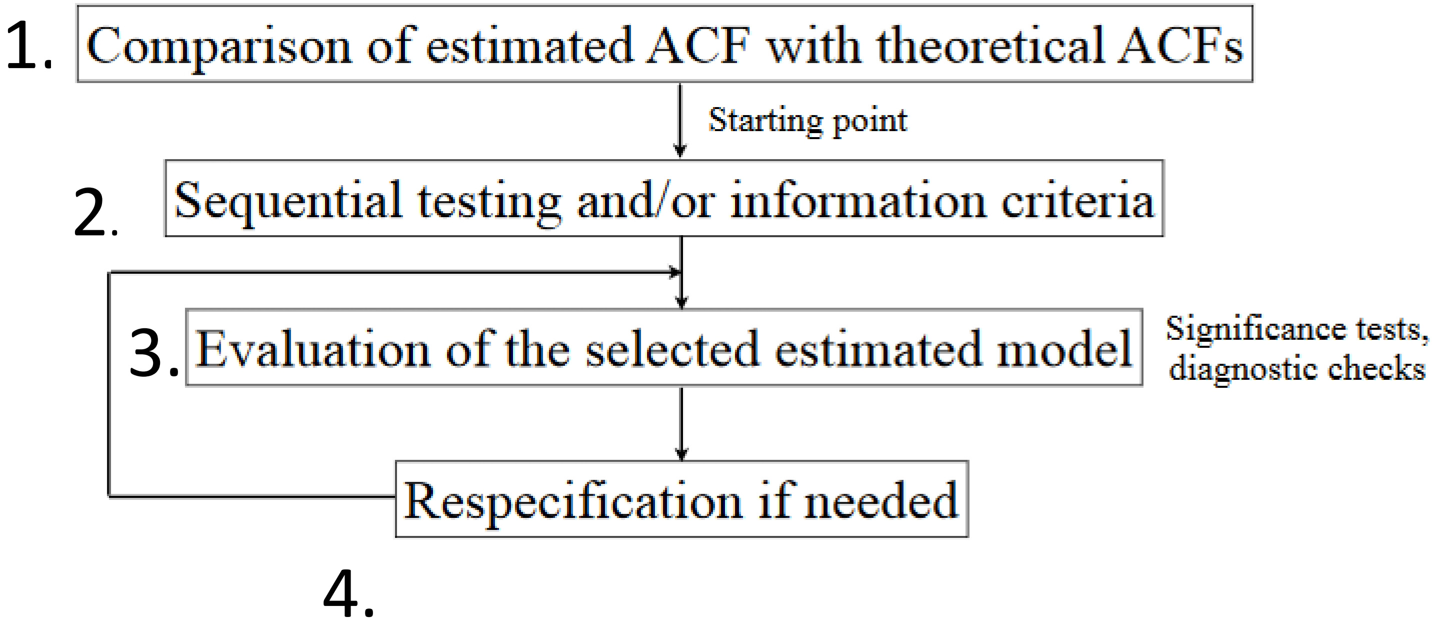
- ARMA combines AR and MA models
- ARMA model is stationary if AR part is stationary (roots greater than one in absolute value)
  - MA part is always stationary
- If ARMA is stationary, it has a MA(∞) representation:

$$\theta(L) y_t = \alpha(L)\varepsilon_t$$

$$\Leftrightarrow y_t = \underbrace{\theta(L)^{-1}\alpha(L)}_{=\psi(L)}\varepsilon_t = \psi(L)\varepsilon_t$$

Model Selection

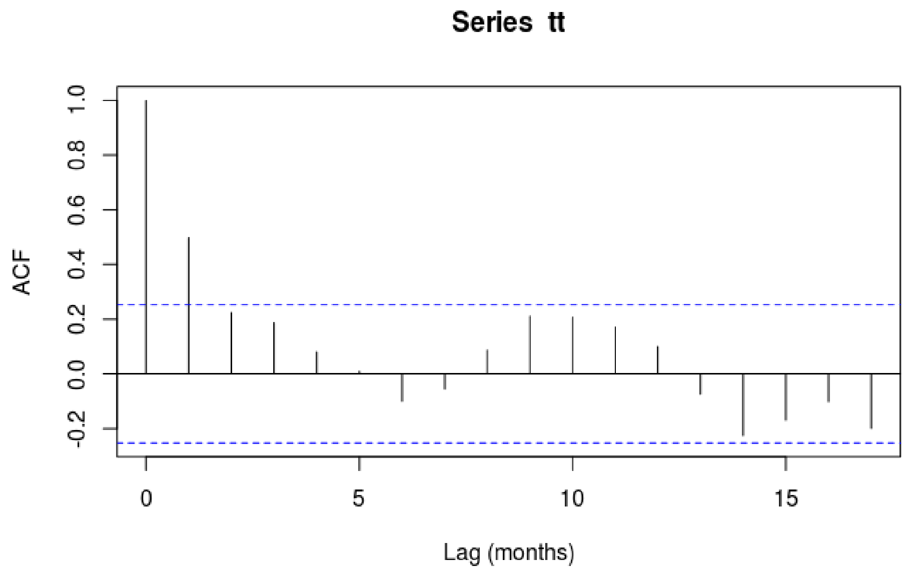
- Goal: Find simplest ARMA model that sufficiently captures dynamics of a time series



Model Selection

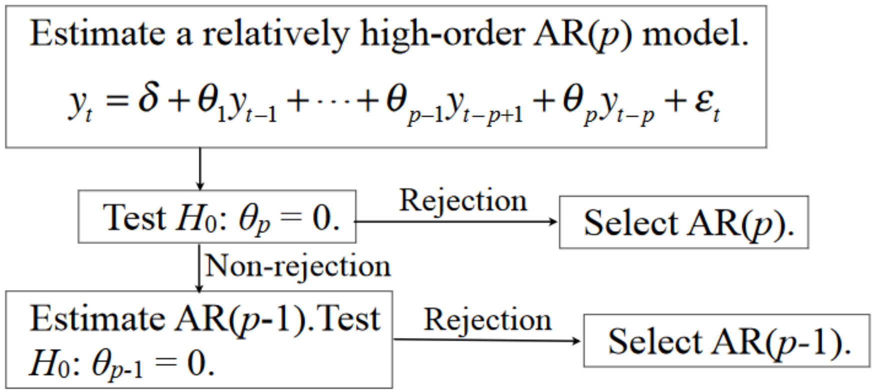
Comparison of autocorrelation functions

- AR(p) -> Autocorrelation last to infinity
- MA(q) -> Autocorrelation lasts q periods, then go to 0
- ARMA(p,q) -> Autocorrelation lasts something between these two
- Look at autocorrelation function and how long autocorrelation last
  - Predict which model is adequate



Model Selection

Sequential testing and information criteria



- Sequential testing
  - Test if last lag p of time series has influence to model
    - $H_0\colon\ \theta_p = 0$
  - If yes (H_0 is rejected) take p lags into the model
  - If no (H_0 not rejected) try influence of lag p-1 and continue until first rejection)
- Information criteria
  - Trade-off between complexity and goodness of fit -> helps to find a balance
  - Bayesian information criterion (BIC) and Akaike information criterion (AIC)
  - Choose a model with smallest value of criterion
  - BIC imposes higher penalty on additional parameters -> imposes more parsimonious model

Model Selection

Diagnostic checking

- If model is adequate*, error term should be white noise*
- There shouldn’t be autocorrelation nor heteroskedasticity
- **Ljung-Box** test for **autocorrelation** in residuals
  - H_0: No autocorrelation, H_1: autocorrelation
  - If p-value < critical value -> model is not adequate -> respesification
  - If p-value > critical value -> all is good
- **Mc Leod-Li** test for conditional **heteroscedasticity** in residuals
  - H_0: No heteroskedasticity (homoskedasticity), H_1: heteroskedasticity
  - If p-value < critical value -> model is not adequate -> respesification
  - If p-value > critical value -> all is good

Forecasting MA(q)

- Forecasting can be done only up to q lags
  - If forecast horizont h > q, forecast yields only constant (if no constant -> 0)
- Error terms of the future can not be forecasted & past values are known

$$y_t = 0.8 + \varepsilon_t + 0.4\varepsilon_{t-1} - 0.1\varepsilon_{t-2}$$

One-step ahead forecast:

$$y_{T+1|T} = \mathrm{E}_T[y_{T+1}] = 0.8 + \mathrm{E}_T[\varepsilon_{T+1}] + 0.4\mathrm{E}_T[\varepsilon_T] - 0.1\mathrm{E}_T[\varepsilon_{T-1}] = 0.8 + 0.4\varepsilon_T - 0.1\varepsilon_{T-1}$$

Two-step ahead forecast:

$$y_{T+2|T} = \mathrm{E}_T[y_{T+2}] = 0.8 + \mathrm{E}_T[\varepsilon_{T+2}] + 0.4\mathrm{E}_T[\varepsilon_{T+1}] - 0.1\mathrm{E}_T[\varepsilon_T] = 0.8 - 0.1\varepsilon_T$$

Forecasting AR(p)

- Forecasting can be done to infinity
- Example: AR(1) with intercept

$x_t = 0.5 + 0.5x_{t-1} + \varepsilon_t$

One-step ahead forecast:

$$x_{T+1|T} = \mathrm{E}_T[x_{T+1}] = 0.5 + 0.5x_T + \mathrm{E}_T[\varepsilon_{T+1}] = 0.5 + 0.5x_T$$

Two-step ahead forecast:

$$\begin{aligned} x_{T+2|T} =& \mathrm{E}_T[x_{T+2}] = 0.5 + 0.5\mathrm{E}_T[x_{T+1}] + \mathrm{E}_T[\varepsilon_{T+2}] \\ =& 0.5 + 0.5^2 + 0.5^2 x_T \end{aligned}$$

Intercept has effect more than once

Random Walk

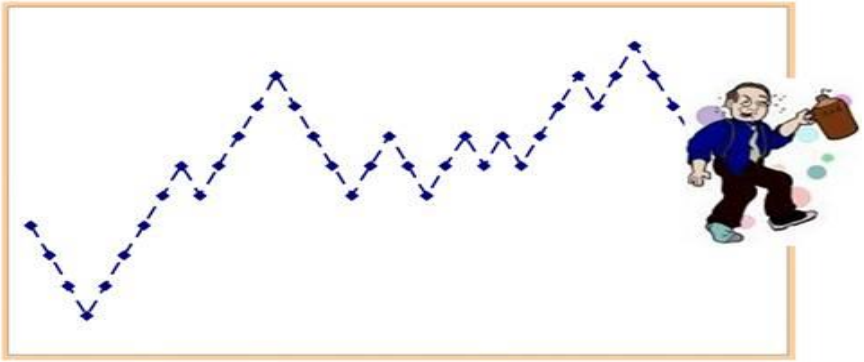
$$y_t = \varepsilon_t + \varepsilon_{t-1} + \cdots \varepsilon_1$$

- The next value of a random walk is obtained by summing a random shock to latest value
- Shock can go to any direction with same probability “randomly”
- $E[y_t] = t\mu$ and $Var[y_t] = t\sigma_\epsilon^2$
- Mean and variance change when t changes
  - Non-stationary, I(1), process
- Random Walk is like drunk bird
  - It can fly wherever with non-trendy direction

Random Walk with Drift

$$y_t = t\delta + \varepsilon_t + \cdots + \varepsilon_1$$

- Random walk exhibits trending behaviour
- Constant term is included -> makes the drift as seen from equation
- Random walk with a drift is like drunk student heading back to home: it has clear direction where to go but every step is random



I(0) and I(1) Processes

- I(0) means stationary process
- I(1) is a process that becomes I(0) after differencing once
- Differencing means changing **data points to changes**
  - Helps to stabilize time series $\Delta X_t = X_t - X_{t-1}$
- Many economic data is not stationary, but differencing ones make them stationary (GDP, inflation, production etc.)
- I(1) processes is like random walk with drift, but I(0) is stationary
- Effect of an error/shock to I(0) is temporary, but in I(1) infinite
- Stationary data is easier to handle -> this is why differencing is done

Testing for an Unit Root

- If time series has an unit root, it is I(1) process
- **Dickey-Fuller** test is for unit root
    - H_0: Unit root exists, H_1: Unit root do not exists
- If test statistics is **smaller than critical value**
    - -> H_0 rejected -> **No unit root** -> **I(0) process**

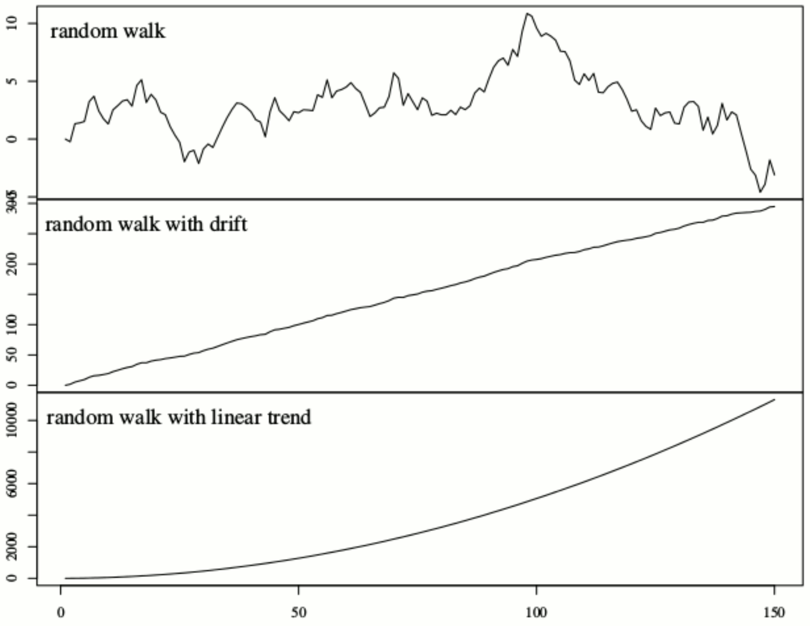
Table 8.1 1% and 5% critical values for Dickey–Fuller tests (Fuller, 1976, p. 373)

	Without trend		With trend	
Sample size	1%	5%	1%	5%
$T = 25$	−3.75	−3.00	−4.38	−3.60
$T = 50$	−3.58	−2.93	−4.15	−3.50
$T = 100$	−3.51	−2.89	−4.04	−3.45
$T = 250$	−3.46	−2.88	−3.99	−3.43
$T = 500$	−3.44	−2.87	−3.98	−3.42
$T = \infty$	−3.43	−2.86	−3.96	−3.41

No growth

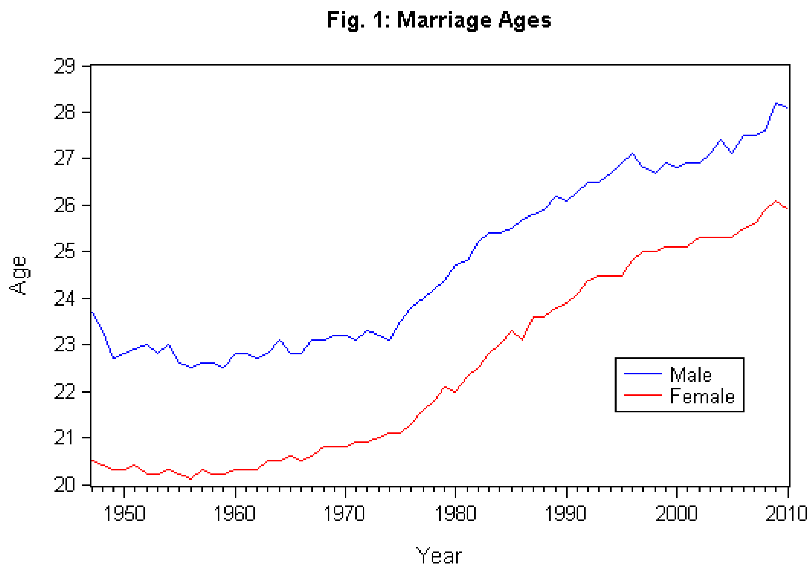
Linear growth

Exponential growth



Cointegration

- Two cointegrated time series follow the same long-run path
- All linear combinations of β are also cointegrating vectors
- Both time series x_t and y_t are I(1)
- If time series are cointegrated there is some cointegration vector β = (β_1, β_2) that makes $\beta_1 x_t + \beta_2 y_t$ I(0) process
- Multiplying β with any non-zero constant is also cointegration vector



Testing for Cointegration

Table 9.2 Asymptotic critical values residual unit root tests for cointegration (with constant term) (Davidson and MacKinnon, 1993)

Number of variables (incl. Y_t)	Significance level 1%	5%	10%
2	−3.90	−3.34	−3.04
3	−4.29	−3.74	−3.45
4	−4.64	−4.10	−3.81
5	−4.96	−4.42	−4.13

- Regress y_t on x_t
- Take residuals from OLS
- Make Augmented Dickey-Fuller test to residuals
  - H_0: No cointegration relation between y_t and x_t
  - H_1: There is cointegration relation between y_t and x_t
- If test statistics is smaller than critical value
  - H_0 rejected and there is cointegration relation
- Notice different critical values
  - This is because uncertainty in estimation with OLS

Vector Autoregression VAR

$$\mathbf{y}_t = \Theta_1 \mathbf{y}_{t-1} + \Theta_2 \mathbf{y}_{t-2} + \cdots \Theta_p \mathbf{y}_{t-p} + \varepsilon_t$$

- Same as AR but multivariate case
- Weakly stationary if roots of the following equation greater than unity in absolute value

$$\left| \mathbf{I}_K - \Theta_1 z - \Theta_1 z^2 - \cdots - \Theta_1 z^p \right| = 0$$

- Diagnostic checks etc. can be done the same way as in univariate AR case

Granger Causality

- If x_t helps to predict y_t, x_t Granger causes y_t
- Granger causality means that information about one time series help to predict the other
- Granger causality do not tell anything about causality relations but it does tell about **correlation** between time series
- Granger causality from x_t to y_t can be tested as H_0: $\theta_{i,12} = 0, i = 1,2, \ldots, p$

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \sum_{i=1}^{p} \begin{pmatrix} \theta_{i,11} & \theta_{i,12} \\ \theta_{i,21} & \theta_{i,22} \end{pmatrix} \begin{pmatrix} y_{t-i} \\ x_{t-i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Vector Error Correction Model VECM

- VECM tells how deviations from the long-run equilibrium are corrected in the short-run

$$\mathbf{y}_t = \Theta_1 \mathbf{y}_{t-1} + \cdots + \Theta_p \mathbf{y}_{t-p} + \varepsilon_t$$

$$\Delta \mathbf{y}_t = \Pi \mathbf{y}_{t-1} + \Gamma_1 \Delta \mathbf{y}_{t-1} + \cdots + \Gamma_{p-1} \Delta \mathbf{y}_{t-p+1} + \varepsilon_t$$

  - $x_t = x_{t-1} + \Delta x_t$
- Π is coefficient matrix $\Pi = -\left(\mathbf{I}_K - \Theta_1 - \cdots - \Theta_p\right)$
- Π can de represented as Π $=\alpha\beta$

A solution for $\mathbf{\Pi} = \alpha\beta'$ is

$$\begin{bmatrix} -.5 & -1. \\ -.25 & -.5 \end{bmatrix} = \begin{pmatrix} -.5 \\ -.25 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix}' = \begin{pmatrix} -.5 \\ -.25 \end{pmatrix} \begin{pmatrix} 1 & 2 \end{pmatrix}$$

Vector Error Correction Model VECM

- Given VAR(p) of x_i's that are I(1)
- There is error correction representation of the model

$$\begin{pmatrix} \Delta R_t \\ \Delta \pi_t \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}}_{\text{Adjustment vector}} \underbrace{\left(R_{t-1} - \pi_{t-1} - \beta_3 \right)}_{\text{Error correction term}} + \begin{pmatrix} \gamma_{1,11} & \gamma_{1,12} \\ \gamma_{1,21} & \gamma_{1,22} \end{pmatrix} \begin{pmatrix} \Delta R_{t-1} \\ \Delta \pi_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

- VECM uses cointegration vectors (from Johansen's method) to explain short-run adjustments to equilibrium
- VECM gives more information than VAR

Vector Error Correction Model VECM

- VECM enables you to use non-stationary data (but cointegrated) for intepretation. This helps retain the relevant information in the data (which would otherwise get missed if data would be differenced)

Estimation and testing of cointegration

- Determine cointegration rank r with Johansen's trace test

```
TABLE 1
######################
# Johansen-Procedure #
######################
Test type: trace statistic , without linear trend and constant in cointegration

Values of test statistic and critical values of test:
          test 10pct  5pct  1pct
r <= 2 |  3.69  7.52  9.24 12.97
r <= 1 | 18.43 17.85 19.96 24.60
r = 0  | 74.87 32.00 34.91 41.07
```

When test value is greater than critical value -> reject H_0 of no cointegration and move forward. Start from r=0 and go until test value is smaller than critical value

β

```
Eigenvectors, normalised to first column (These are the cointegration relations):
                 r1.l1       r12.l1        pi.l1    constant
r1.l1     1.0000000000  1.000000000  1.000000000  1.00000000
r12.l1   -1.6145290295 -4.035966898 -2.205454650 -2.04453702
pi.l1     0.6131495820  2.883785437  1.266070084  1.12451013
constant  0.0001627575  0.001860306 -0.005412181 -0.01153677
```

α

```
Weights W (This is the loading matrix):
          r1.l1     r12.l1     pi.l1     constant
r1.d -2.2183245 0.3410717 -0.4602082 9.518250e-14
r12.d-1.0972162 0.3353384 -0.6070851 5.677882e-14
pi.d -0.4324214 0.2462456 -0.7829379 9.309052e-14
```

Testing hypothesis on cointegration vector β

- Hypothesis can be expressed as: $$\underset{(K\times r)}{\beta} = \underset{(K\times s)}{\mathbf{H}}\underset{(s\times r)}{\varphi}$$
- Suppose we have

Let $\mathbf{y}_t = (r_{1t}, r_{12t}, \pi_t)'$. Hypothesis (i) implies that the spread $r_{12t} - r_{1t}$ is $I(0)$, while hypothesis (ii) implies that the real interest rate $r_{12t} - \pi_t$ is $I(0)$. Thus the hypotheses (i) and (ii) jointly imply a cointegrating rank of 2 on $\mathbf{y_t}$.

(b) If $\mathbf{y_t}$ is cointegrated, a VAR(p) model for it can be written as

$$\Delta\mathbf{y_t} = \Pi\mathbf{y_{t-1}} + \Gamma_1\Delta\mathbf{y_{t-1}} + \cdots + \Gamma_{p-1}\Delta\mathbf{y_{t-p+1}} + \varepsilon_t,$$

where $\Pi = \alpha\beta'$. The cointegration implications of hypotheses (i) and (ii) can be written as

$$\beta = \mathbf{H}\varphi$$

Assuming that intercept terms are restricted to the cointegration relations, what does $\mathbf{H}$ look like?

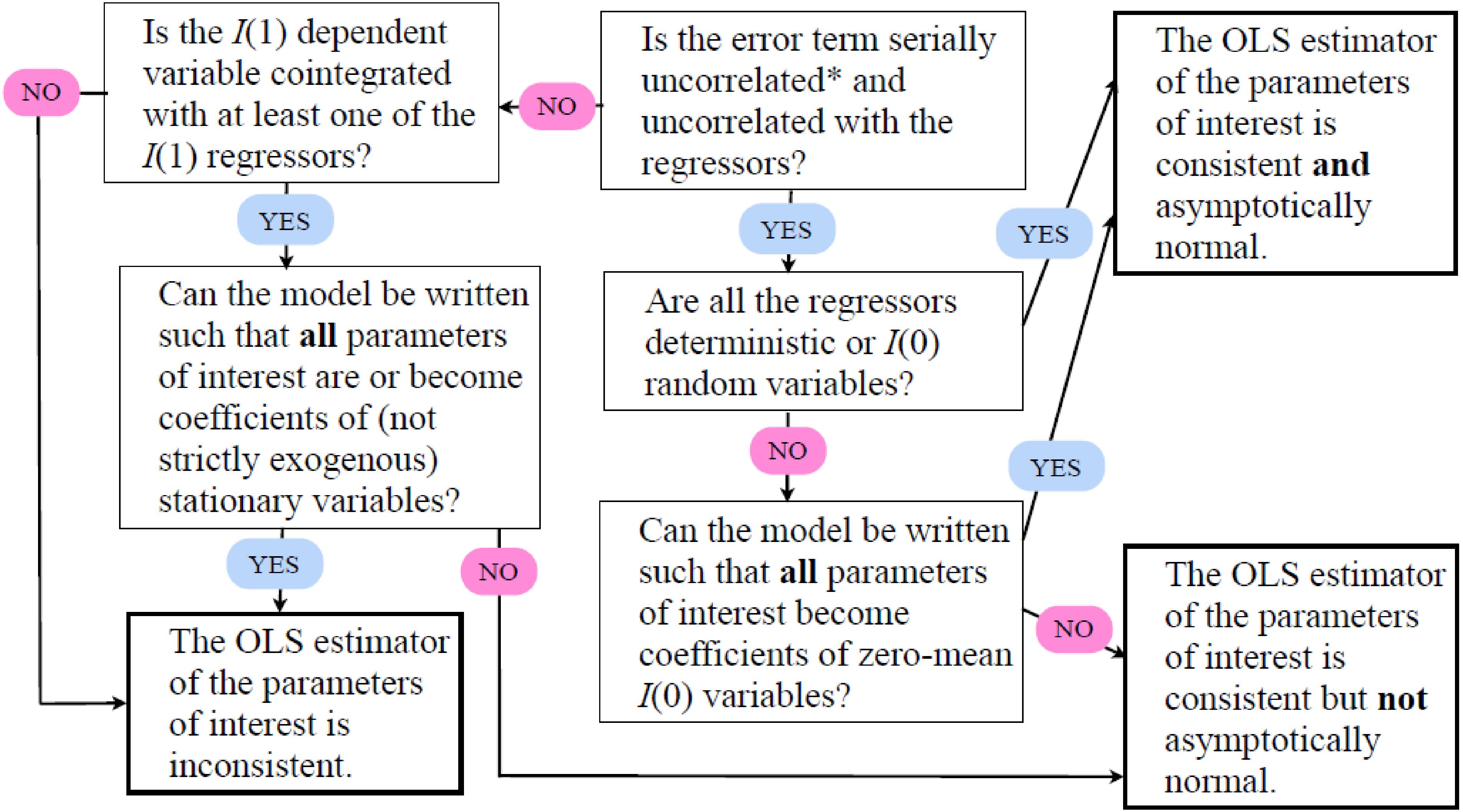
- H can be interpret as:

	$r_{12t} - r_{1t}$	$r_{12t} - \pi_t$	Intercept
r_{1t}	-1	0	0
r_{12t}	1	1	0
π_t	0	-1	0
Intercept	0	0	1

OLS with Time Series

- Consistency and asymptotical normality of β
  - Asymptotical normality is needed for testing
- Newey-West estimator
  - Estimate of the covariance matrix
  - Is used to try to remove autocorrelation and heteroskedasticity
- Breusch-Godfrey test for autocorrelation
  - H_0: no autocorrelation (p-value)
- Breusch-Pagan test for heteroskedasticity
  - H_0: homoskedasticity (p-value)

OLS Regression with $I(1)$ Variables Cheat Sheet



* Not autocorrelated after H-1 lags.

Test	What we are testing?	Null hypothesis	Alternative hypothesis	Critical values
Sequential testing	last lag p of time series has influence to model	No influence	Has influence	p-values
Ljung-Box test	Autocorrelation of residuals	No autocorrelation	Autocorrelation exists	p-values
Mc Leod Li test	Heteroskedasticity of residuals	Homoskedasticity	Heteroskedasticity	p-values
Dickey-Fuller	Unit roots	Unit root exists	No unit root	H0 rejected, if test result < critical value
Augmented Dickey-Fuller	Cointegration of time series	No cointegration	Cointegration exists	H0 rejected, if test result < critical value
Johansen Procedure	Cointegration rank	No cointegration at rank r	Cointegration at rank r	Start from r=0 and stop when test statistic < critical value