Applied Time Series Econometrics

Johannes W. Fedderke
ERSA and University of Cape Town

Second Semester 2006

1. An Introduction to Some Time Series Data Properties and Concepts

Central assumption of classical normal linear regression model:

- observations are independently sampled.

Assumption generally violated in the time series context:

- observations are connected in all kinds of ways: eg. autocorrelation.

2. Stochastic processes

- If a sample is independently sampled it means that the inclusion of one observation should have no bearing on whether or not another observation is included.
- The information we obtain from one observation should tell us nothing about the information that we will obtain from the next observation.

- In a time series context this is generally not the case:
  - Firstly, if we include year 1960 in the sample we will generally include 1961, 1962 and so on.
  - Secondly, we frequently find that the value that our variable takes on in 1960 will provide information about the value it will take on in 1961, 1962 and so on.

- A discrete-time stochastic process is a collection of random variables $\{X_t(\varpi)\}$ where:
  - $t \in T$ denotes time, and
  - $\varpi \in \Omega$ denotes an elementary event or outcome
  - when Ω is the relevant sample space.
- The stochastic process is defined relative to a probability space, (Ω, F, P), where
  - F is the event space and
  - $P(\cdot)$ is an appropriate probability measure such that $P(A)$ is defined for all events $A \in F$.

- For time t_0, $X_{t_0}(\varpi)$ is a random variable for $\varpi \in \Omega$.
- For a fixed value ϖ_0, $X_t(\varpi_0)$ is a *realization* of the stochastic process for $t \in T$, such that different values for ϖ induce different realizations.

- EXAMPLE:
  - Event: rolling a fair, six-sided die every morning before breakfast next week;
  - generates 7 *random variables*: each morning's score generates 1 number;
  - 7 numbers together are the stochastic process
- Where:
  - stochastic process has 1 number at each point in time: *time series*

2.1 The white noise process

Simplest form of stochastic process:

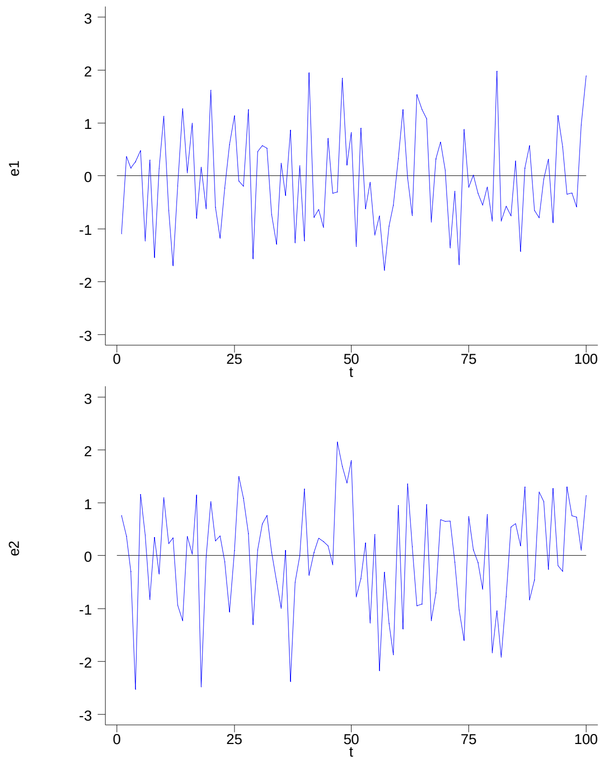
- A mean-zero stochastic process is white noise if its elements are mutually uncorrelated.
- Let $\{u_t\}$ denote a zero-mean, finite-variance stochastic process, so that we have:
  - $E[u_t] = 0$ and
  - $E[u_t^2] < \infty$, then:

* $\{u_t\}$ is weak white noise if $E\left[u_t u_s\right] = 0 \, \forall t \neq s$.
* $\{u_t\}$ is strong white noise if $E\left[u_t u_s\right] = 0 \, \forall t \neq s$, and $E\left[u_t^2\right] = \sigma^2$.
* Any process that is $I.I.D.\left[0, \sigma^2\right]$ would be strong white noise.

- Strong white noise process is given by:

$$\begin{aligned} y_t &= \epsilon_t \\ \epsilon_t &\sim N.I.I.D.(0, \sigma^2) \end{aligned} \tag{1}$$

- For example:



2.2 The Random Walk

Evolution determined by:

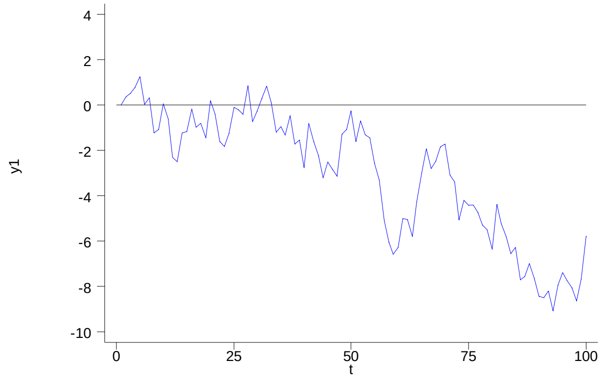
$$y_t = y_{t-1} + \epsilon_t \tag{2}$$

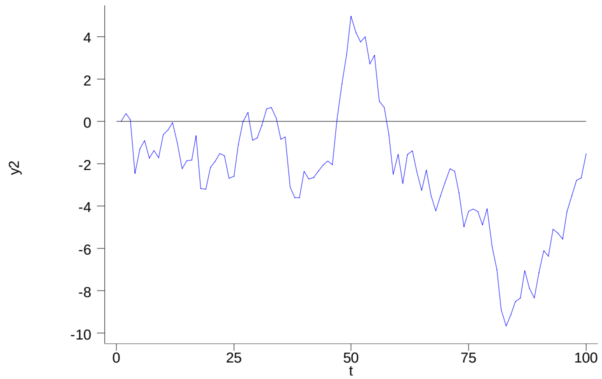
where ϵ_t is a white noise process.

Depends on:

- values of the stochastic error terms,
- "starting value" of the series.

- For instance:





2.3 The Random Walk With Drift

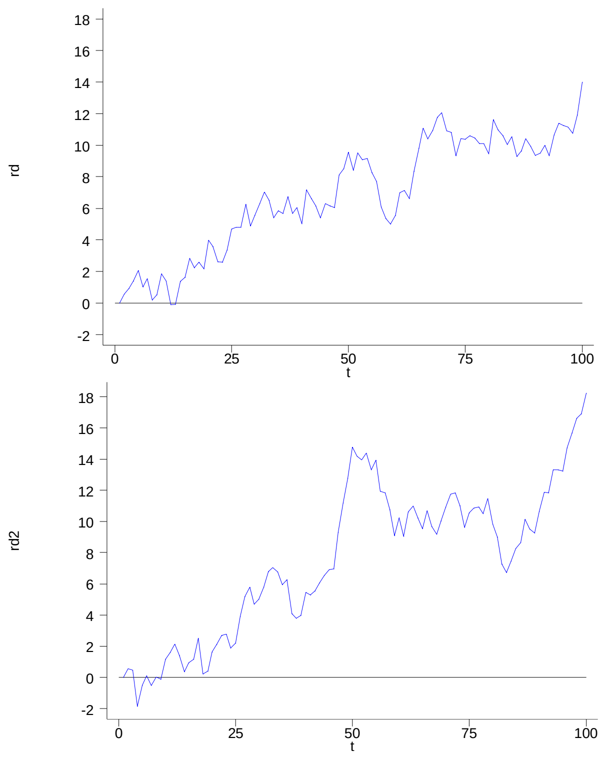
Random walk with drift is given by:

$$y_t = \mu + y_{t-1} + \epsilon_t \tag{3}$$

where ϵ_t is a white noise process and μ is a fixed parameter.

Dominated by drift term in long run.

• For instance:



3. Typology of Stochastic Processes

3.1 Stationary Processes

A stationary process is defined by the fact that the distribution of the random variable y_t must be the same as the distribution of the random variable y_{t+s} for any value of s.

If a series is stationary, it makes sense to talk about the *mean* of the series, its *variance* and the *autocovariance* between terms, i.e.:

- $E(y_t)$ is a constant $= \mu$
- $var(y_t)$ is a constant $= \sigma^2$
- and $cov(y_t, y_{t+k})$ is a constant $= \gamma_k$ $(for\ k = 0, \pm 1, \pm 2.....)$

The values of the autocovariances γ_k can be thought of as a function of k.

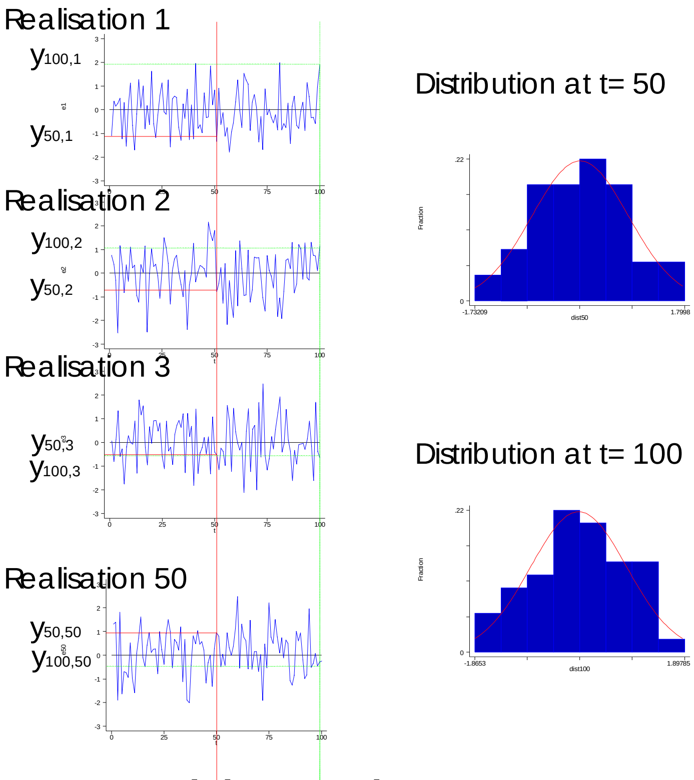
This is referred to as the *autocovariance function*.

We can also define the *autocorrelation function* ρ_k which gives the autocorrelation coefficient ρ_k between y_t and y_{t+k}.

It is easily shown that:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{4}$$

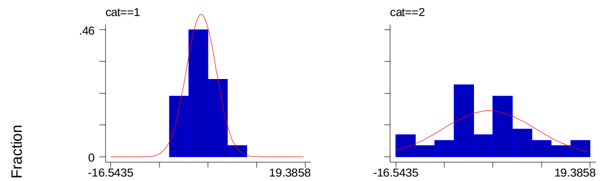
Consider:



White noise process

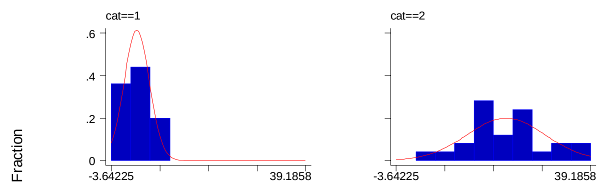
Versus:

Distribution at t= 10 Distribution at t= 100



Random walk

Distribution at t= 10 Distribution at t= 100



Random walk with drift

3.2 Autoregressive (AR) processes

The simplest autoregressive process is the $AR(1)$ process given by:

$$\begin{aligned} y_t &= \delta + \phi y_{t-1} + \epsilon_t \\ \epsilon_t &\sim N.I.D.(0, v^2) \end{aligned} \tag{5}$$

where:

- ϵ_t is a white noise process,
  - with mean zero and
  - a common variance v^2.
- y_t is not a white noise process:
  - the condition that $E\left[y_t y_s\right] = 0 \ \forall t \neq s$ is immediately violated.

- Stationary provided that $| \phi |< 1$:
  - Then $\mu = \frac{\delta}{1-\phi}$ and $\sigma^2 = \frac{v^2}{1-\phi^2}$.
  - Furthermore $\gamma_1 = \phi\sigma^2, \gamma_2 = \phi^2\sigma^2, ..., \gamma_k = \phi^k\sigma^2, ...$
  - $\rho_k = \phi^k$.

- The $AR(p)$ process can be defined by:

$$y_t = \delta + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t \tag{6}$$

$$\epsilon_t \sim N.I.D.(0, v^2)$$

- More complicated to characterise whether or not the series is stationary.
- Do so by means of lag operator.

- *Lag operator L*: defined by the operation of taking a one-period lag, i.e.

$$Ly_t = y_{t-1} \tag{7}$$

Naturally enough $L^2 y_t = LLy_t = Ly_{t-1} = y_{t-2}$ and so $L^p y_t = y_{t-p}$.

- We can rewrite the $AR(p)$ process with lag operator notation as:

$$\left(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p\right) y_t = \delta + \epsilon_t \tag{8}$$

- We can think of $\left(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p\right)$ as being a polynomial in the lag operator.
- The series will be stationary, provided that every solution L_i to the equation

$$\left(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p\right) = 0 \qquad (9)$$

  is such that $\mid L_i \mid > 1$.
- If the equation 9 has a solution where $L_i = 1$, we say that the series has a *unit root* and then the series is nonstationary. Where the equation 9 has a solution where $L_i < 1$, the series will be explosive.
- Necessary, not sufficient.

3.3 Moving Average (MA) processes

The simplest $MA(1)$ moving average process is given by:

$$\begin{aligned} y_t &= \delta + \epsilon_t + \theta_1 \epsilon_{t-1} \\ \epsilon_t &\sim N.I.D.(0, v^2) \end{aligned} \tag{10}$$

where:

- $\mu = \delta$ and $\sigma^2 = \left(1 + \theta_1^2\right) v^2$.
- $\gamma_1 = \theta_1 v^2$ and $\gamma_2 = \gamma_3 = ... = 0$.

What is the difference with AR?:

- With AR processes correlations persist (potentially with ever decreasing size)
- In the case of the $MA(1)$ process autocorrelations die out after one period.

- The general MA(q) process is given by:

$$\begin{aligned} y_t &= \delta + \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} \quad (11) \\ \epsilon_t &\sim N.I.D.(0, v^2) \end{aligned}$$

In this case also:

- $\mu = \delta$ and $\sigma^2 = \left(1 + \theta_1^2 + ... + \theta_q^2\right) v^2$.
- This process will be stationary.
- Autocorrelations in this case will die out after q periods.

3.4 Autoregressive Moving Average (ARMA) processes

Can combine autoregressive and moving average components to get the so-called $ARMA$ process.

The general $ARMA(p, q)$ process is:

$$y_t = \alpha + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \qquad (12)$$

This can be written with the lag operator notation as

$$\left(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p\right) y_t \qquad (13)$$
$$= \alpha + \left(1 - \theta_1 L - \theta_2 L^2 - ... - \theta_p L^q\right) \epsilon_t$$

This series is stationary only if the autoregressive component of it is stationary, i.e. only if the polynomial on the left-hand side of the equation does not have unit roots.

3.5 Example

- AR process.
- MA process.

4. Distinguishing Variance Matrixes

4.1 Conditional, Unconditional, Long Run Variance Matrixes

Define:

- Conditional variance: the short-run or contemporaneous variance - takes variables in the function as fixed
- Unconditional matrix: allowing for the dependence of the variance of a random variable on the past
- Long-run matrix: variance of random variable in steady state

Consider:

- Two random variables: $\varepsilon_t \sim \left(0, \sigma_\varepsilon^2\right)$, $\omega_t \sim \left(0, \sigma_\omega^2\right)$
- This gives:
  - $u_t = (\varepsilon_t, \omega_t)'$, a 2×1 vector,
  - and the $E\left\{u_t, u_t'\right\}$ variance matrix:
    - diagonals: variances: $\sigma_\varepsilon^2, \sigma_\omega^2$
    - off-diagonal: covariances: $E\left\{\varepsilon_t, \omega_{t+k}\right\}, E\left\{\omega_t, \varepsilon_{t+k}\right\}$
    - non-symmetric in general (leads, lags)

Example:

- Suppose:

$$\begin{aligned} Y_t &= \varphi_2 X_t + \varepsilon_t \\ X_t &= X_{t-1} + \omega_t \\ \varepsilon_t &\sim \left(0, \sigma_\varepsilon^2\right) \\ \omega_t &\sim \left(0, \sigma_\omega^2\right) \\ cov(\varepsilon_t, \omega_t) &= 0 \end{aligned}$$

- Then for $u_t = (\varepsilon_t, \omega_t)'$:

$$E\left\{u_t, u_t'\right\} = \Phi = \begin{bmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_\omega^2 \end{bmatrix}$$

4.2 Autoregressive Processes

Consider the case of ω_t as AR(1):

$$\begin{aligned} \omega_t &= \gamma\omega_{t-1} + \nu_t \\ |\gamma| &< 1 \\ \nu_t &\sim \left(0, \sigma_\nu^2\right) \\ cov(\varepsilon_t, \nu_t) &= 0 \end{aligned}$$

Now the variance of ω_t *conditional* on ω_{t-1}:

$$Var\left(\omega_t|\omega_{t-1}\right) = \sigma_\nu^2$$

Conditional variance: the short-run or contemporaneous variance - takes variables in the function as fixed

The *unconditional* variance of ω_t is:

$$\begin{aligned} Var(\omega_t) &= Var(\gamma\omega_{t-1} + \nu_t) \\ &= \gamma^2 Var(\omega_{t-1}) + Var(\nu_t) \\ \Longrightarrow \sigma_\omega^2 &= \gamma^2\sigma_\omega^2 + \sigma_\nu^2 \\ \Longrightarrow \sigma_\omega^2 &= \frac{\sigma_\nu^2}{1-\gamma^2} \end{aligned}$$

Unconditional variance: allows for the variance of ω_{t-1}

Note that:

$$\omega_t = \frac{\nu_t}{1 - \gamma L}$$

In steady state:

$$Var\left(\omega_t | L = 1\right)$$

Hence the *long-run* variance of ω_t is:

$$\begin{aligned} Var\left(\omega_t | L = 1\right) &= \left(\frac{\sigma_\nu^2}{\left(1 - \gamma L\right)^2} | L = 1\right) \\ &= \frac{\sigma_\nu^2}{\left(1 - \gamma\right)^2} \end{aligned}$$

Long-run variance: allows for dependence on all past $\omega_{t-i}, i \geq 1$

Note, for $0 < \gamma < 1$:

$$\sigma_\nu^2 < \frac{\sigma_\nu^2}{1-\gamma^2} < \frac{\sigma_\nu^2}{(1-\gamma)^2}$$

or:

- conditional $<$ unconditional $<$ long-run variance
- $\gamma \to 1, \Longrightarrow \frac{\sigma_\nu^2}{1-\gamma^2} \to \infty, \frac{\sigma_\nu^2}{(1-\gamma)^2} \to \infty$
- Even in steady state, in the case of $\gamma \to 1$: feasible space is infinitely large.

Long-run variance matrix, Ω, for $u_t = (\varepsilon_t, \omega_t)'$:

- Assuming Φ diagonal:

$$\Omega = \begin{bmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \frac{\sigma_\nu^2}{(1-\gamma)^2} \end{bmatrix}$$

$$given$$

$$\Phi = \begin{bmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_\nu^2 \end{bmatrix}$$

- Assuming Φ non-diagonal, $cov(\varepsilon_t, \nu_t) \neq 0$:

$$\Omega = \begin{bmatrix} \sigma_\varepsilon^2 & \frac{\sigma_{\varepsilon \upsilon}}{(1-\gamma)} \\ \frac{\sigma_{\varepsilon \upsilon}}{(1-\gamma)} & \frac{\sigma_\nu^2}{(1-\gamma)^2} \end{bmatrix}$$

$$given$$

$$\Phi = \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon \upsilon} \\ \sigma_{\varepsilon \upsilon} & \sigma_\omega^2 \end{bmatrix}$$

- Allowing for both ε_t, ω_t as AR(1):

$$\begin{aligned}
\varepsilon_t &= \gamma_\varepsilon \varepsilon_{t-1} + \zeta_t,\ |\gamma_\varepsilon| < 1 \\
\omega_t &= \gamma_\omega \omega_{t-1} + \nu_t,\ |\gamma_\omega| < 1 \\
\sigma_{\zeta\nu} &= 0
\end{aligned}$$

such that:

$$\begin{pmatrix} \varepsilon_t \\ \omega_t \end{pmatrix} = \begin{bmatrix} \gamma_\varepsilon & 0 \\ 0 & \gamma_\omega \end{bmatrix} \begin{pmatrix} \varepsilon_{t-1} \\ \omega_{t-1} \end{pmatrix} + \begin{pmatrix} \zeta_t \\ \nu_t \end{pmatrix}$$

or:

$$\begin{aligned}
u_t &= A u_{t-1} + \eta_t \\
u_t &= (\varepsilon_t, \omega_t)' \\
\eta_t &= (\zeta_t, \nu_t)'
\end{aligned}$$

- Conditional variance:

$$\Phi = \begin{bmatrix} \sigma_\zeta^2 & 0 \\ 0 & \sigma_\nu^2 \end{bmatrix}$$

- Unconditional variance:

$$\begin{aligned}
\Psi &= E\left\{u_t, u_t'\right\} \\
&= E\left\{(Au_{t-1} + \eta_t)(Au_{t-1} + \eta_t)'\right\} \\
&= E\left\{Au_{t-1}A'u_{t-1}'\right\} + E\left\{\eta_t \eta_t\right\}', \ given\ E\left\{u_{t-1}\eta_t'\right\} = 0 \\
&= AE\left\{u_{t-1}u_{t-1}'\right\} A' + \Phi \\
\Psi &= A\Psi A' + \Phi
\end{aligned}$$

- Long-run variance; from:

$$\begin{aligned} u_t &= Au_{t-1} + \eta_t \\ (I - AL)\, u_t &= \eta_t \\ \Longrightarrow u_t &= \frac{\eta_t}{(I - AL)} \end{aligned}$$

- Hence:

$$\begin{aligned}
\Omega &= E\left\{u_t, u_t' | L = 1\right\} \\
&= E\left\{(I - AL)^{-1}\,\eta_t\eta_t'\left(I - A'L\right)^{-1} | L = 1\right\} \\
&= \left[(I - AL)^{-1}\,E\left\{\eta_t\eta_t'\right\}\left(I - A'L\right)^{-1} | L = 1\right] \\
&= \left[(I - AL)^{-1}\,\Phi\left(I - A'L\right)^{-1} | L = 1\right] \\
&= (I - A)^{-1}\,\Phi\left(I - A'\right)^{-1}
\end{aligned}$$

- Interequation serial correlation now follows trivially from:

$$\begin{aligned}\varepsilon_t &= \gamma_\varepsilon \varepsilon_{t-1} + \gamma_{\varepsilon\omega}\omega_{t-1} + \zeta_t,\ |\gamma_\varepsilon| < 1 \\ \omega_t &= \gamma_\omega \omega_{t-1} + \gamma_{\omega\varepsilon}\varepsilon_{t-1} + \nu_t,\ |\gamma_\omega| < 1 \\ \sigma_{\zeta\nu} &= 0\end{aligned}$$

such that:

$$\begin{pmatrix} \varepsilon_t \\ \omega_t \end{pmatrix} = \begin{bmatrix} \gamma_\varepsilon & \gamma_{\varepsilon\omega} \\ \gamma_{\omega\varepsilon} & \gamma_\omega \end{bmatrix} \begin{pmatrix} \varepsilon_{t-1} \\ \omega_{t-1} \end{pmatrix} + \begin{pmatrix} \zeta_t \\ \nu_t \end{pmatrix}$$

The rest follows mutatis mutandis.

- Finally, suppose $\sigma_{\zeta\nu} \neq 0$.
- Then:

$$\Phi = \begin{bmatrix} \sigma_\zeta^2 & \sigma_{\zeta\nu} \\ \sigma_{\zeta\nu} & \sigma_\nu^2 \end{bmatrix}$$

- EXAMPLE:

$$
\begin{aligned}
\varepsilon_t &= 0.2\varepsilon_{t-1} + 0.1\omega_{t-1} + \zeta_t \\
\omega_t &= 0.5\omega_{t-1} + 0.4\varepsilon_{t-1} + \nu_t \\
\zeta_t &\sim (0,1) \\
\nu_t &\sim (0,1) \\
\sigma_{\zeta\nu} &= 0
\end{aligned}
$$

such that:

$$
\begin{pmatrix} \varepsilon_t \\ \omega_t \end{pmatrix} = \begin{bmatrix} 0.2 & 0.1 \\ 0.5 & 0.4 \end{bmatrix} \begin{pmatrix} \varepsilon_{t-1} \\ \omega_{t-1} \end{pmatrix} + \begin{pmatrix} \zeta_t \\ \nu_t \end{pmatrix}
$$

and:

$$
u_t = \begin{bmatrix} 0.2 & 0.1 \\ 0.5 & 0.4 \end{bmatrix} u_{t-1} + \eta_t
$$

- Gives conditional variance:

$$\Phi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- Unconditional variance:

$$
\begin{aligned}
\Psi &= A\Psi A' + \Phi \\
&= \begin{bmatrix} 0.2 & 0.1 \\ 0.5 & 0.4 \end{bmatrix} \Psi \begin{bmatrix} 0.2 & 0.5 \\ 0.1 & 0.4 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
\Longrightarrow \begin{bmatrix} 0.2 & -0.5 \\ -0.1 & 0.4 \end{bmatrix} \Psi &= \Psi \begin{bmatrix} 0.2 & 0.5 \\ 0.1 & 0.4 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
&\Longrightarrow \begin{bmatrix} 13.13 & -0.1 & -16.67 & 0 \\ -0.5 & 12.93 & 0 & -16.67 \\ -3.33 & 0 & 6.47 & -0.1 \\ 0 & -3.33 & -0.5 & -6.27 \end{bmatrix} \begin{bmatrix} \Psi_{11} \\ \Psi_{12} \\ \Psi_{21} \\ \Psi_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}
\end{aligned}
$$

- Long-run variance:

$$\begin{aligned}\Omega &= (I - AL)^{-1}\,\Phi\left(I - A'L\right)^{-1} \\ &= \begin{bmatrix} 0.8 & -0.1 \\ -0.5 & 0.6 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.8 & -0.5 \\ -0.1 & 0.6 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 2.001 & 2.055 \\ 2.005 & 4.813 \end{bmatrix}\end{aligned}$$

4.3 Moving Average Processes

- Suppose:

$$
\begin{aligned}
Y_t &= \varphi_2 X_t + \varepsilon_t \\
X_t &= X_{t-1} + \omega_t \\
\varepsilon_t &= \zeta_t + \theta_{11}\zeta_{t-1} + \theta_{12}\nu_{t-1} \\
\omega_t &= \nu_t + \theta_{21}\zeta_{t-1} + \theta_{22}\nu_{t-1} \\
Var(\varepsilon_t, \omega_t) &= \Phi
\end{aligned}
$$

or:

$$
\begin{pmatrix} \varepsilon_t \\ \omega_t \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \zeta_t \\ \nu_t \end{pmatrix} + \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} \begin{pmatrix} \zeta_{t-1} \\ \nu_{t-1} \end{pmatrix}
$$

$$
\begin{aligned}
u_t &= \eta_t + \theta_1 \eta_{t-1} \\
&= (I + \theta_1 L)\, \eta_t \\
u_t &= (\varepsilon_t, \omega_t)' \\
\eta_t &= (\zeta_t, \nu_t)'
\end{aligned}
$$

- Unconditional variance:

$$\begin{aligned}
\Psi &= E\left\{u_t, u_t'\right\} \\
&= E\left\{\eta_t + \theta_1\eta_{t-1}, \eta_t' + \eta_{t-1}'\theta_1'\right\} \\
&= E\left\{\eta_t\eta_t'\right\} + \theta_1 E\left\{\eta_{t-1}\eta_{t-1}'\right\}\theta_1', \ given\ E\left\{\eta_t\eta_{t-1}\right\} = 0 \\
&= \Phi + \theta_1\Phi\theta_1'
\end{aligned}$$

- Long-run variance:

$$
\begin{aligned}
\Omega &= E\left\{u_t, u_t' | L = 1\right\} \\
&= E\left\{(I + \theta_1 L)\, \eta_t \eta_t' \,(I + \theta_1 L)' \,|L = 1\right\} \\
&= E\left\{(I + \theta_1 L)\, \Phi \,(I + \theta_1 L)' \,|L = 1\right\} \\
&= (I + \theta_1)\, \Phi \,(I + \theta_1)' \\
&= \Psi + \Phi\theta_1' + \theta_1\Phi \\
&= \Psi + \Lambda + \Lambda',\ \Lambda' = \Phi\theta_1'
\end{aligned}
$$

- EXAMPLE:

$$\begin{aligned} \varepsilon_t &= \zeta_t + 0.2\zeta_{t-1} + 0.3\nu_{t-1} \\ \omega_t &= \nu_t + 0.1\zeta_{t-1} + 0.4\nu_{t-1} \\ \zeta_t &\sim (0,1) \\ \nu_t &\sim (0,1) \\ \sigma_{\zeta\nu} &= 0 \end{aligned}$$

such that:

$$\begin{pmatrix} \varepsilon_t \\ \omega_t \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \zeta_t \\ \nu_t \end{pmatrix} + \begin{bmatrix} 0.2 & 0.3 \\ 0.1 & 0.4 \end{bmatrix} \begin{pmatrix} \zeta_{t-1} \\ \nu_{t-1} \end{pmatrix}$$

Given $\Phi = I$, this provides:

$$\begin{aligned} \Psi &= I + \theta_1\theta_1' = \begin{bmatrix} 1.13 & 0.14 \\ 0.14 & 1.17 \end{bmatrix} \\ \Omega &= (I + \theta_1)\left(I + \theta_1'\right) = \begin{bmatrix} 1.53 & 0.54 \\ 0.54 & 1.17 \end{bmatrix} \end{aligned}$$

5. Stationarity

Consider a simple DGP, which is $AR(1)$:

$$\begin{aligned} y_t &= \rho y_{t-1} + u_t \\ u &\sim IID(0, \sigma^2) \end{aligned} \tag{14}$$

Now y will be:

- **weakly (second order, covariance) stationary where** $\mid \rho \mid < 1$
- **nonstationary where** $\mid \rho \mid = 1$
- **explosive where** $\mid \rho \mid > 1$

Strong stationarity has all the requirements of weak stationarity, as well as the requirement that the constant, finite mean, variance and covariance are independent of the time frame over which the sample data is drawn.

More precisely, if $\rho = 1$, the current value of y will be:

$$y_t = y_{t-n} + \sum_{i=0}^{n-1} u_{t-i} \tag{15}$$

and recall that:

$$Var(y_t|y_{t-1}) = \sigma^2$$

$$Var(y_t) = \frac{\sigma^2}{1-\rho^2} \rightarrow \infty, \ given\ \rho = 1$$

$$Var(y_t|L = 1) = \frac{\sigma^2}{(1-\rho)^2} \rightarrow \infty, \ given\ \rho = 1$$

Moreover, y does not converge on a mean value in the normal sense, since if at any given time $y = c$, it will take infinitely long for y to return to the c value.

- The DGP is collecting stochastic shocks over time, and does so with non-decaying weights.
- In effect a stochastic shock will never die out in the series, and will reappear with a full weight in all subsequent time periods.
- Over a finite time horizon this series of stochastic shocks will be non zero, and it is this which will give the DGP the appearance of trending.
- Moreover, it is this appearance of trend that will in general mislead OLS techniques when the series is regressed on another with similar characteristics.

By contrast, if $\rho < 1$, if we start with an initial value of y_{t-n}, then we have a finite moving average (MA) process of order n, with greatest weight being placed on the most recent elements of the disturbance terms.

For the complete series we have:

$$y_t = \rho^n y_{t-n} + \sum_{i=0}^{n-1} \rho^i u_{t-i} \tag{16}$$

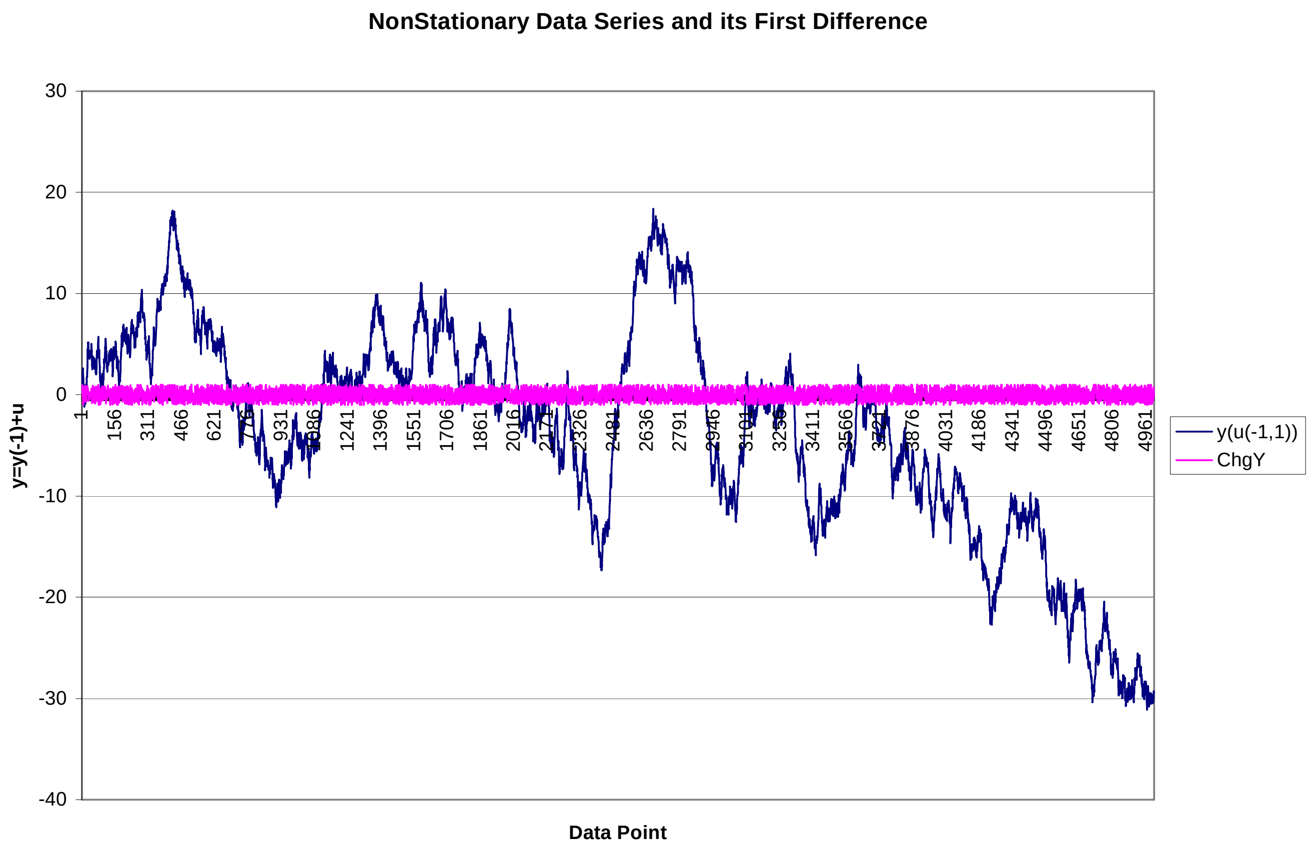
The declining weight attaching to the random disturbances has as a consequence that:

- The mean of the series is constant,
- The variance constant and finite,
- The covariance is constant,
- i.e. all are independent of time.

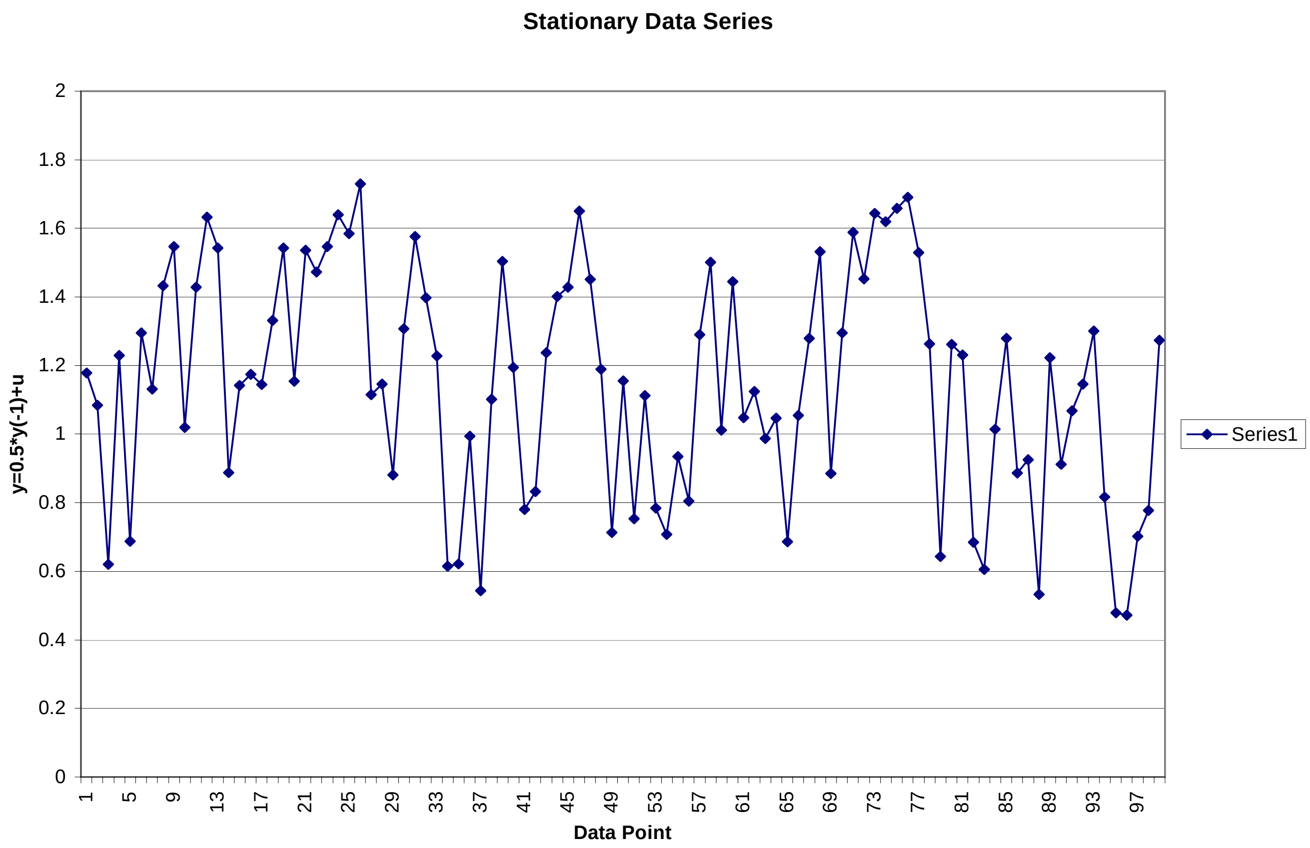
The DGP is collecting stochastic shocks over time, but in contrast to the case where $\rho = 1$ it does so with *decaying* weights.

While stochastic shocks will never die out in the series, they reappear with rapidly declining weights, and soon have a negligible impact on subsequent values of y_t.

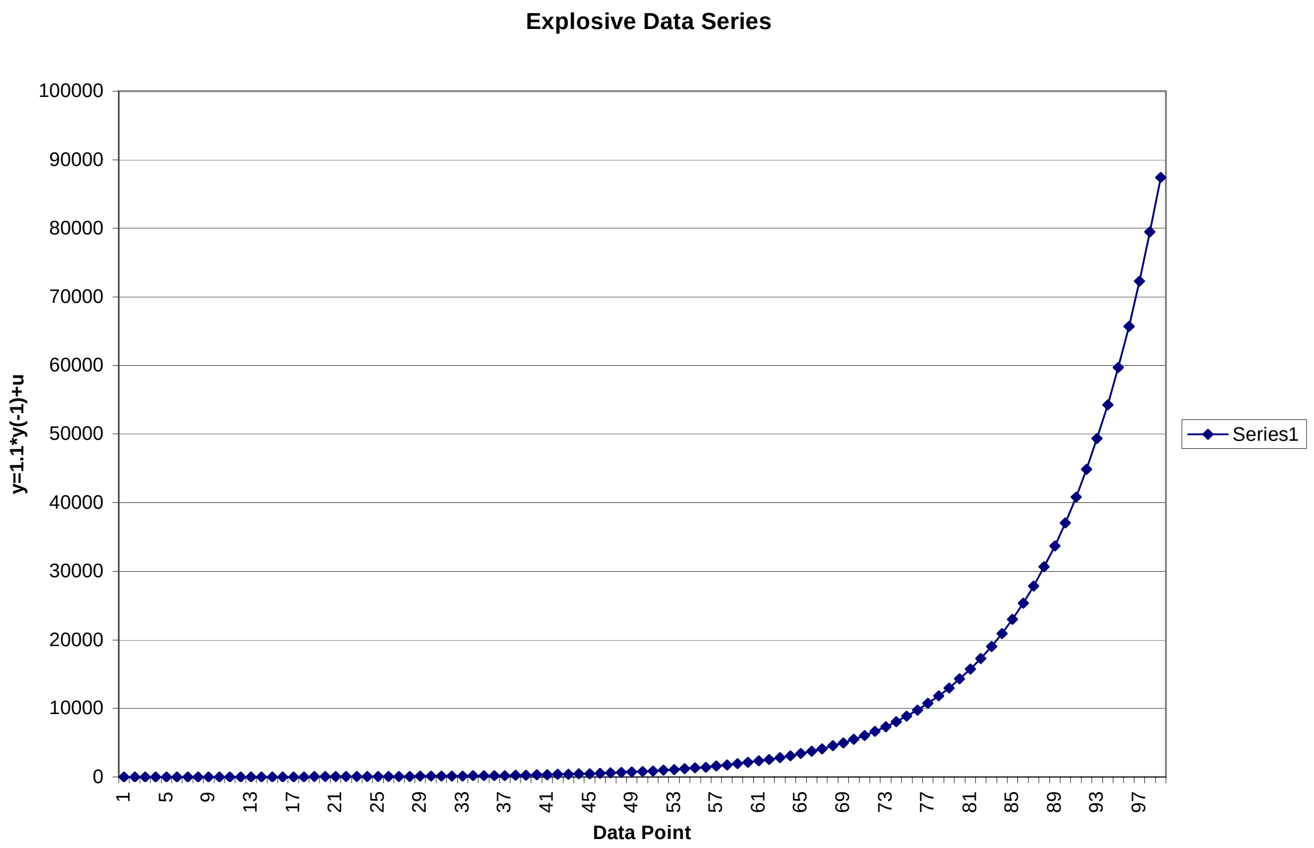
EXAMPLE: Non-stationarity: $|\rho| = 1$:



EXAMPLE: Stationarity: $\mid \rho \mid < 1$:



EXAMPLE: Explosive: $|\rho| > 1$:



- EXAMPLES: Simulations

5.1 Unit Roots

We also note that the stationarity properties of a data series depend on whether the series has a unit root or not.

Rewrite the DGP as:

$$(1 - \rho L)\, y_t = u_t \tag{17}$$

where L denotes the lag operator.

This provides the characteristic equation:

$$(1 - \rho L) = 0$$

If the root of the characteristic equation:

- is unity, the series is nonstationary;
- if the root of the equation is greater than unity, the series is stationary.

Since here there is only one root, $L = 1/\rho$, for the root to exceed unity requires that $\mid \rho \mid < 1$.

It is for this reason that tests on the univariate properties of the data (essentially testing the stationarity of the series) are termed unit root tests.

Alternatively, we might say that we test for a root inside the unit circle, to cover the possibility of both non-stationary and explosive $DGP's$.

5.2 Time Series with Drift, and Deterministic Trends

We can consider the impact and nature of trending in time series in more detail.

For instance, given:

$$y_t = \delta + \rho y_{t-1} + u_t \tag{18}$$

for $\rho = 1$, the series will show a systematic displacement (to the value of δ) in each and every time period (which we term **drift**).

The solution for y_t is:

$$y_t = y_{t-n} + \delta t + \sum_{j=1}^{t} u_j \tag{19}$$

Note:

- there is a trend in the series generated by the drift term over time,
- but the series cannot settle down to the deterministic trend in the DGP, since the accumulated error terms will displace the series away from the trend.
- the direction of the drift will depend on the sign of the δ term, since under $\rho = 1$, $\Delta y_t = \delta + u_t$, such that $E(\Delta y_t) = \delta$.

Under the alternative condition where $\mid \rho \mid < 1$, we again have stochastic shocks which persist, but they do so with decaying weights, such that their influence becomes negligible.

Once again therefore, we have stationarity under the $\mid \rho \mid < 1$ condition.

We can see this from the general expression for the drifting random walk:

$$y_t = \rho^n y_{t-n} + \delta \sum_{i=0}^{n-1} \rho^i + \sum_{i=0}^{n-1} \rho^i u_{t-i} \qquad (20)$$

where $\rho = 1$ stochastic shocks are non-decaying, while $\mid \rho \mid < 1$ allows the shocks to persist, but with declining weight.

By contrast to the preceding DGP, consider:

$$y_t = \alpha + \phi t + u_t \tag{21}$$

where $\alpha+\phi t$ constitutes a deterministic trend with drift, while u_t is the stochastic non-trend component.

It follows that $\Delta y_t = \phi + u_t - u_{t-1}$ such that $E(\Delta y_t) = \phi$, a constant, and $var(y_t) = \sigma^2$. The DGP is said to be *trend stationary*, since although the series will trend and be subject to drift, deviations from the deterministic component of the series will be stationary.

Note therefore:

- Both of the preceding $DGP's$ exhibit a trend, but
- they differ crucially in that while the former does not show stationary disturbances (since the effect of the disturbances is cumulative), the second DGP does manifest stationary disturbances.
- thus we may have either a stochastic (non-stationary) or a deterministic (stationary) trend in the DGP,
- which can make testing for the presence of unit roots complicated, (the power of the relevant tests by means of which we attempt to identify the univariate time series characteristics of $DGP's$ is low).

- EXAMPLES: Simulations

5.3 Difference stationarity

Any non-stationary series can be rendered stationary by differencing.

Consider:

$$\begin{aligned} y_t &= y_{t-1} + u_t \\ u &\sim IID(0, \sigma^2) \\ \implies y_t - y_{t-1} &= u_t \end{aligned}$$

$$\begin{aligned} y_t &= \delta + y_{t-1} + u_t \\ u &\sim IID(0, \sigma^2) \\ \implies y_t - y_{t-1} &= \delta + u_t \end{aligned}$$

Note therefore:

- We characterize the stationarity properties of a time series by referring to its order of integration.
- Where a series does not need to be differenced in order to be rendered stationary (i.e. where it is already stationary in levels), we say that it is integrated of order zero, which we denote as $I(0)$.
- Where a series needs to be differenced once in order to be rendered stationary (i.e. its first difference is stationary), we say that it is integrated of order one, which we denote as $I(1)$.

- Where a series becomes stationary after $n'th$ order differencing (i.e. its $n'th$ order difference is stationary), we say that it is integrated of order n, which we denote as $I(n)$.
- EXAMPLES: Simulations

5.4 Wold's decomposition theorem

- Any covariance stationary process, x_t, can be represented as a deterministic component, $\overline{x}_t$, and an indeterministic component:

$$x_t = \overline{x}_t + \sum_{i=0}^{\infty} \theta_i \xi_{t-i}$$

- Note:
  - $\overline{x}_t$ need not be constant, but it must be stationary (eg. circular function). In practice often treated as a constant.
  - Can then "remove" $\overline{x}_t$, and assuming only that $\sum_{i=0}^{\infty} \theta_i \xi_{t-i}$ is square-summable, and that its coefficients are (very) small beyond a certain lag,
  - then we can approximate the indeterministic component by a finite-order MA-process, or ARMA-process.

- Consider:

$$
\begin{aligned}
x_t &= \xi_t \left(1 + L + 0.5L^2 + 0.25L^3 + ...\right) \\
\Longrightarrow \sum_{i=0}^{\infty} \theta_i^2 &= \left[1 + 1 + 0.5^2 + 0.25^2 + ...\right] \\
&= \frac{1}{1 - 0.5^2} + 1 = 2\frac{1}{3} < \infty
\end{aligned}
$$

- And: ARMA(1,1)

$$\begin{aligned} x_t &= 0.5x_{t-1} + \xi_t + 0.5\xi_{t-1} \\ \Longrightarrow \; x_t\,(1-0.5L) &= \xi_t\,(1+0.5L) \\ \Longrightarrow \; x_t &= \xi_t\,(1+0.5L)\,(1-0.5L)^{-1} \\ &= \xi_t\left(1 + L + 0.5L^2 + 0.25L^3 + ...\right) \end{aligned}$$

- Hence Wold.