Time Series Analysis

ARIMA

Michael Colaresi
colaresi@msu.edu

September 18, 2014

Outline

STATIONARITY

.

Stationarity (Intuitive Definition) Data series is stationary if there is no systematic change in the mean (e.g. no trend), no systematic stochastic variation, and if strict periodic variations (seasonal) are stable. Time plays no role in the sample moments (mean, variance and generally distribution of variable, does not change over time.

Stationary II

- *How do you know a data series are stationary or not?*
  1. Trend is observable
  2. Check seasonality (is Christmas bigger ever year)
  3. Look for outliers/Pattern in variance
  4. Observe Discontinuities (jumps up or down in the series)

  Stationarity means something, non-stationarity is simply absence of stationarity (could be anything)

Strict Stationarity

Strict Stationarity The mean, variance and the joint distribution of a time series do not change given different slices in time. So, for example, the joint density of X_t and X_{t+1} should be no different than that for X_{t+1} and X_{t+2}.

- **Implication:** Shifting the origin (time) by the amount t has no effect on the joint distribution, so we can truncate or analyze $(t_1, t_2, \ldots, t_n)$ with no change in the joint distribution over time.

Weak Stationarity

Weak Stationarity A property that is more tractable empirically. A series is weakly stationarity if

1. $E\left[X_t\right] = \mu$ "finite mean"
2. $Cov\left[X_t, X_{t+i}\right] = g\left(i\right)\ \forall\ i$ "covariance stationarity" (covariance depends only on the lag/shift value (i)).

- *Strict and weak stationarity are the same under a normal distribution, since* $\left(\mu, \sigma^2\right)$ *fully characterize a normal distribution.*

Why is Stationarity Important?

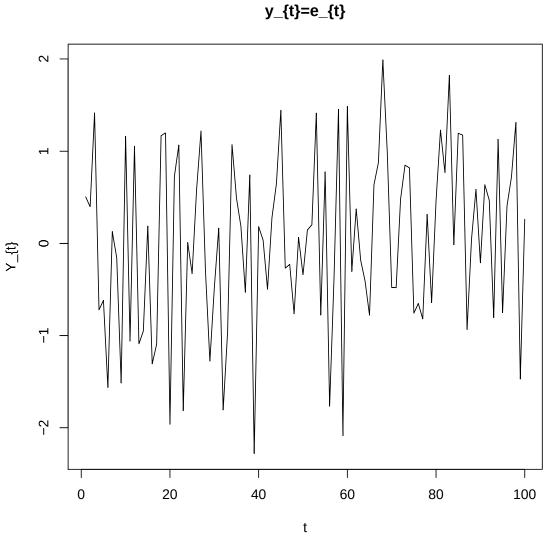
- INFERENCE
- Assumptions
  - errors
  - linearity

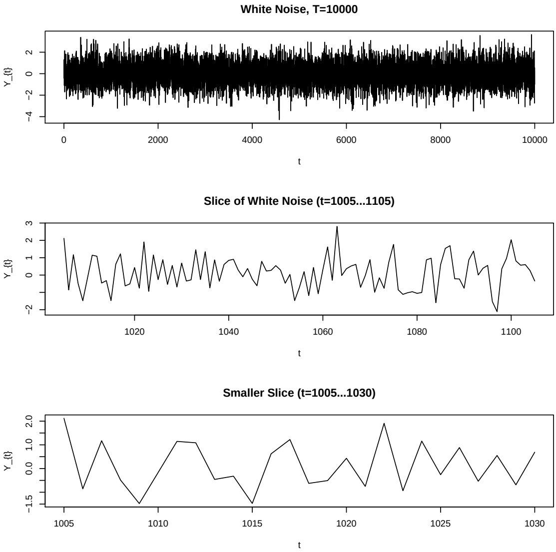
The DNA of Time Series

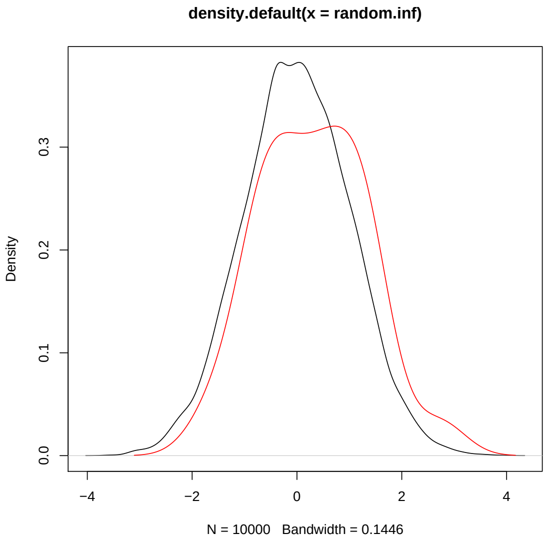
- Consider the *three* basic stochastic properties that are widely discussed: a purely random process or "white noise", a deterministic trend, and a random walk.
  - **A purely random or white noise process:** Y_t is a sequence of random variables that are mutually independent and identically distributed. It follows that

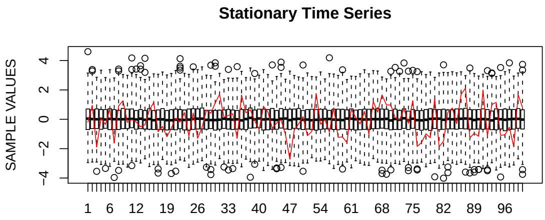
$$\begin{aligned} a &: \quad E[Y] = \bar{Y} \ \text{"constant mean"} \\ b &: \quad E(Y - E[Y]) = \sigma^2 \ \text{"constant variance"} \\ c &: \quad \therefore Cov(Y_t, Y_{t-i}) = 0 \ \forall \ i \end{aligned}$$

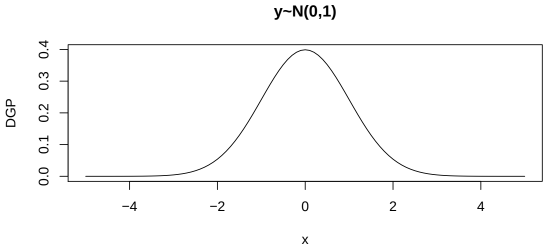
    This is referred to as a "white noise" process. It is weakly stationary, and does not depend on time.


y_{t}=e_{t}
Y_{t}
t
2
1
0
−1
−2
0
20
40
60
80
100


White Noise, T=10000
Y_{t}
t
Slice of White Noise (t=1005...1105)
Y_{t}
t
Smaller Slice (t=1005...1030)
Y_{t}
t

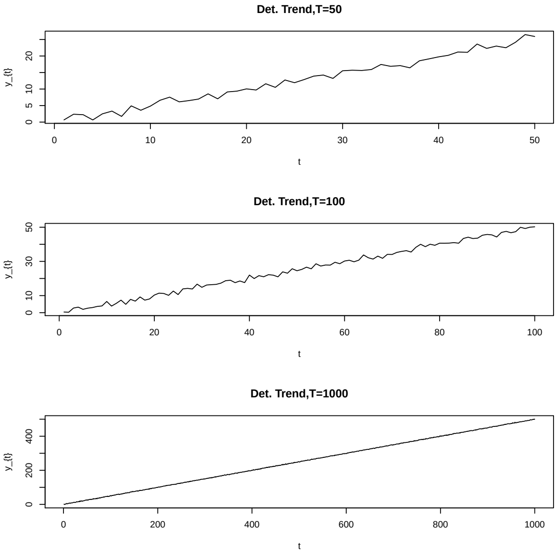

density.default(x = random.inf)
Density
0.0
0.1
0.2
0.3
−4
−2
0
2
4
N = 10000 Bandwidth = 0.1446

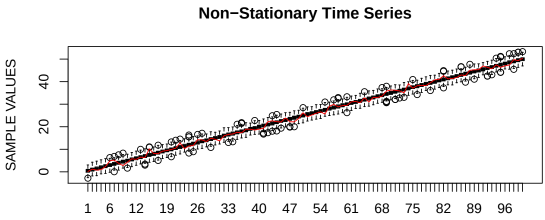

Stationary Time Series
SAMPLE VALUES
4
2
0
−2
−4
1
6
12
19
26
33
40
47
54
61
68
75
82
89
96

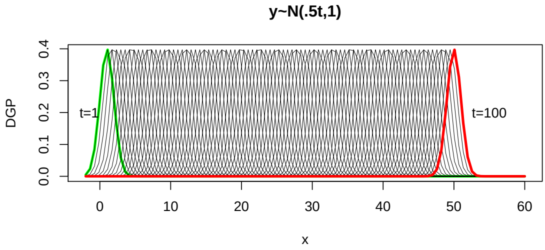

y~N(0,1)
DGP
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
x

DNA of Time Series II

- **Deterministic trend:** If ϵ_t is a purely random process with mean μ and variance σ^2_ϵ, $y_t = \beta t + \epsilon_t$ is a deterministic trend
- at $t = 0 : y_t = \epsilon_0$
- at $t = T : E[y_T] = \beta T$
- sample variance of y_t also moves with t


Det. Trend,T=50
y_{t}
0
5
10
20
0
10
20
30
40
50
t
Det. Trend,T=100
y_{t}
0
10
30
50
0
20
40
60
80
100
t
Det. Trend,T=1000
y_{t}
0
200
400
0
200
400
600
800
1000
t


Non−Stationary Time Series
SAMPLE VALUES
0
20
40
1
6
12
19
26
33
40
47
54
61
68
75
82
89
96

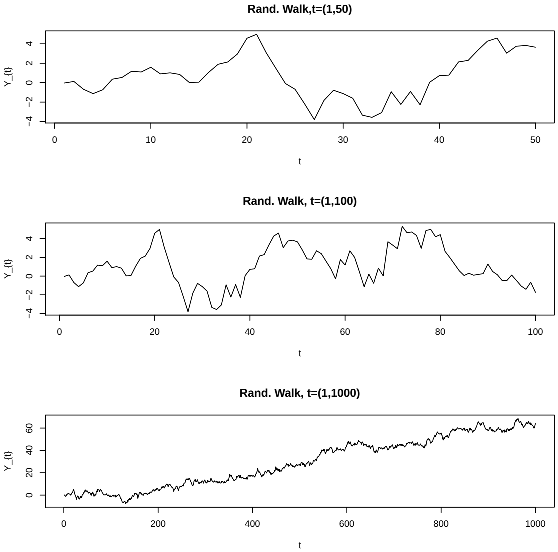

y~N(.5t,1)
DGP
0.0
0.1
0.2
0.3
0.4
t=1
t=100
0
10
20
30
40
50
60
x

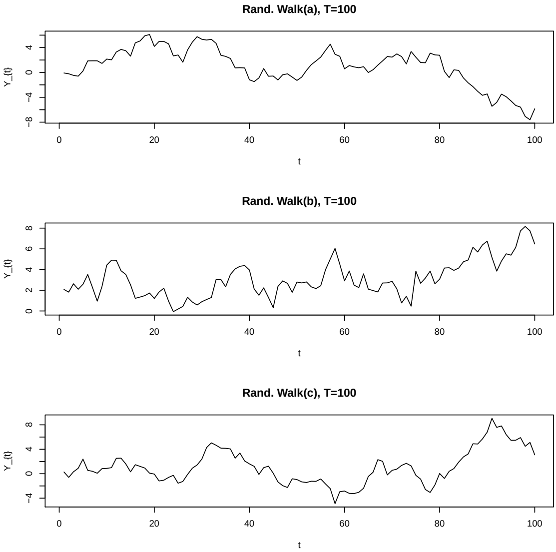
DNA III

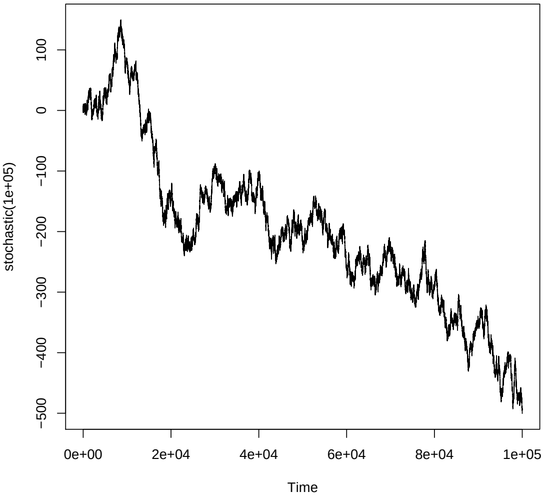
- **Random walk:** Suppose $\epsilon*_t$ is a purely random process with mean μ and variance σ^2_ϵ. A process Y_t is a random walk if $Y_t = \mu + Y_{t-1} + \epsilon_t$
- At $t = 0$ with $\mu = 0$: $Y_t = \epsilon_0$. Then, at time T, $Y_t = \sum_{t=1}^{T} \epsilon_t$. This is called a martingale process. Generally, it follows that

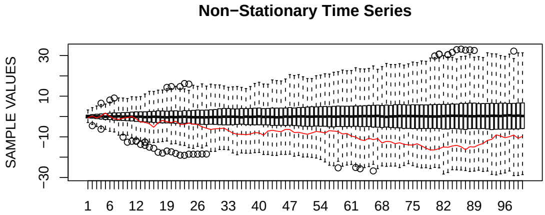
$$
\begin{aligned}
a &: \quad E[Y] = t\mu \\
b &: \quad E(Y - E[Y]) = Var[Y_t] = t\sigma^2_\epsilon
\end{aligned}
$$

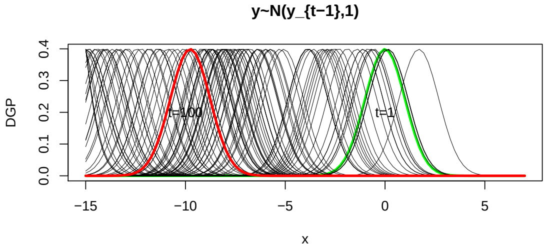
- which imply that the mean and variance change with time, which is a form of non-stationarity.


Rand. Walk,t=(1,50)
Y_{t}
t
Rand. Walk, t=(1,100)
Y_{t}
t
Rand. Walk, t=(1,1000)
Y_{t}
t


Rand. Walk(a), T=100
Y_{t}
t
Rand. Walk(b), T=100
Y_{t}
t
Rand. Walk(c), T=100
Y_{t}
t


stochastic(1e+05)
100
0
−100
−200
−300
−400
−500
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
Time


Non−Stationary Time Series
SAMPLE VALUES
30
10
−10
−30
1 6 12 19 26 33 40 47 54 61 68 75 82 89 96


y~N(y_{t−1},1)
DGP
0.4
0.3
0.2
0.1
0.0
t=100
t=1
−15
−10
−5
0
5
x

Tranformations to Stationarity

Why do we transform non-stationary series?

1. Stabilize the variance
2. Make seasonal effects or shocks additive, and not increasing with the mean.

Transformation methods

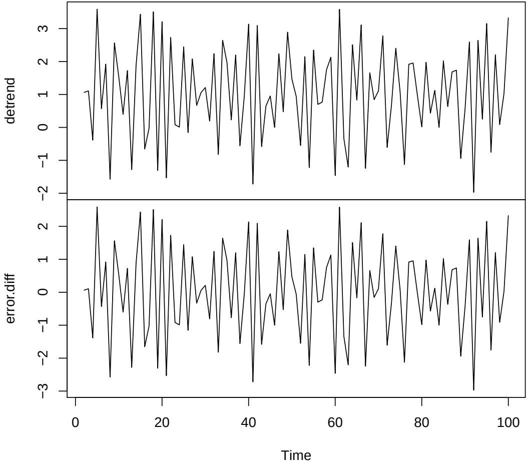
1. Logs (same units, stabilize variance, model percentage change)
2. Differencing: $(Y_t - Y_{t-1})$
3. Linear transformation when cointegration exists $(Y_t - X_t)$
4. Both (log first, then difference)

Differencing Non-stationary Variables

- Deterministic Trend ($y_t = \beta t + \epsilon_t$)
- Stochastic Trend ($y_t = y_{t-1} + \epsilon_t$)

What do you get?

Delta Det. Trend, difference of Y_{t} (up) and errors (down)



Covariance

The covariance between two random variables X and Y is found by the following formula

$$Cov\,(X,Y) \;=\; E\left[(X - E\,[X])\,(Y - E\,[Y])\right]$$

Sample Covariance

- If X and Y are "independent," then their covariance is zero.
- If X and Y are not independent, then the covariance may be positive or negative depending on whether "high" values of X tend to go with "high" values of Y.
- With N observations, the estimated (sample) covariance is $Cov\left(X, Y\right) = \left(N-1\right)^{-1} \sum_{i=1}^{N} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)$

Autocovariance

- If X, Y are random variables from the *same* stochastic process at different times, then the covariance parameter is called an *autocovariance* parameter.

Correlation

- The covariance is difficult to interpret since it depends on the "units" in which X and Y are measured.
- In addition, the "range" is difficult to assess. One way around this is to calculate the correlation coefficient:

$$
\begin{aligned}
Corr\left(X,Y\right) &= \frac{Cov\left(X,Y\right)}{\sqrt{E\left(X-E\left[X\right]\right)^2 E\left(Y-E\left[Y\right]\right)^2}} \\
&= \frac{E\left(X-E\left[X\right]\right) E\left(Y-E\left[Y\right]\right)}{\sqrt{E\left(X-E\left[X\right]\right)^2 E\left(Y-E\left[Y\right]\right)^2}}
\end{aligned}
$$

- The correlation coefficient ranges between ± 1.

Autocorrelation

- If X, Y are random variables from the *same* stochastic process at different times, then the correlation coefficient is called an *autocorrelation* coefficient.
- This may sound familiar.
- Lots of different processes can cause autocorrelation/autocovariance.

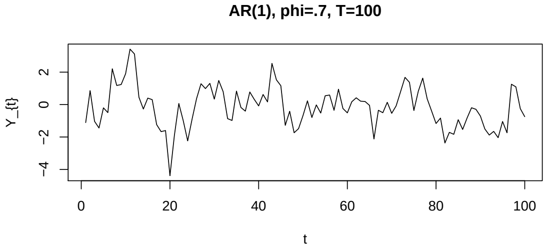
AutoCorrelation Function

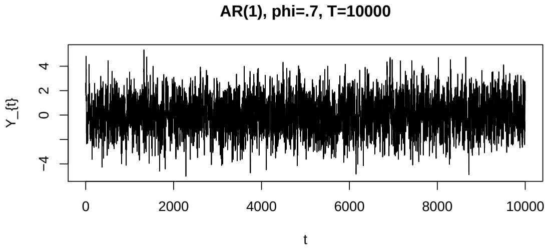
- The ACF is a function that produces correlations for the same series for different lags, such as for the pairs $(Y_1, Y_2), (Y_1, Y_3)$, etc.
- Same formula as above.
- Calculate $Corr(y_t, y_{t-i})$ recursively for all integers $i \in (1, p)$, and then plot.
- Example: ACF for $y_t = .4y_{t-1} + \epsilon_t$, where $\epsilon_t \; N(0, 1)$
  - $i = 1, Corr(y_t, y_{t-1}) = .4$
  - $i = 2, Corr(y_t, y_{t-2}) = .16$
  - $i = 3, Corr(y_t, y_{t-3}) = .064$
  - $\ldots$

Autoregressive Process

- Autoregressive process of order p: Suppose ϵ_t is a purely random process with mean $\mu = 0$ and variance σ^2_ϵ. A process Y_t is said to be an *autoregressive process of order p*, $(AR(p))$, if
  $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t$.
- An $AR(\infty) = \phi B = 1 - \sum_{j=1}^{\infty} \phi_j B^j$

The ARIMA process: ARMA(p,d,q)





AR Properties

- AR process does not depend on time
  1. $\epsilon_t s$ are independent
  2. Mean is constant
  3. Roots are stable $\therefore$ weakly stationary process
  4. If $\epsilon_t \sim N\left(\mu, \sigma^2\right) \Rightarrow$ strictly stationary process.
- ACF "dampens"

Moving Average

- Moving Average Process of order q: Suppose ϵ_t is a purely random process with mean $\mu = 0$ and variance σ^2_ϵ. A process Y_t is said to be a *moving average process of order q* $(MA(q))$, if

$$\begin{aligned} Y_t &= \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q} \\ &= \left(1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \ldots - \theta_q B^q\right) \epsilon_t \end{aligned}$$

- An $MA(\infty) = \theta(B) = 1 + \sum_{j=1}^{\infty} -\theta_j B^j$.

MA Properties

- The MA process does not depend on time.
  1. $\epsilon_t s$ are conditionally independent.
  2. The mean is constant $\therefore$weakly stationary process.
  3. If $\epsilon_t \sim N\left(\mu, \sigma^2\right) \Rightarrow$ strictly stationary process.
  4. Roots are stable and process is “invertible”.
- ACF “cuts off” or “ends” abruptly.

ARIMA(p,d,q)

It is sometimes the case that some data series exhibit mixed AR/MA behavior. As a generic example

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \ldots - \theta_q \epsilon_{t-q} \qquad (1)$$

- Using the lag operator (B), we can rewrite (1) as

$$\phi(B) Y_t = \theta(B) \epsilon_t \qquad (2)$$

- where $\phi(B), \theta(B)$ are polynomials of order p and q respectively. Now components in (2) can be written as:

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \ldots - \phi_p B^p \qquad (3) \\ \theta(B) &= 1 - \theta_1 B - \ldots - \theta_q B^q \qquad (4) \end{aligned}$$

AR to MA and Back Again

One of the most important things to remember about ARIMA models is that AR and MA process are mathematically equivalent to each other at the limit.

- You can turn an AR into an MA
- You can turn an MA into an MA
- *Without loss of information (just paper)*

To Infinity and Beyond

Both AR and MA processes can be represented as infinite order representations of the other.

The implication:

- AR process helps "identify" an MA process (more on this below).
- The MA process tells us whether an AR process is stationary.
- Using both allows us to better understand how to use ACF and PACFs.

MA process to infinite-order AR process

- Consider the following $MA(1)$ process:

$$Y_t = \epsilon_t - \theta\epsilon_{t-1} \tag{5}$$

- Shift (5) back one period:

$$\begin{aligned} Y_{t-1} &= \epsilon_{t-1} - \theta\epsilon_{t-2} && (6) \\ \epsilon_{t-1} &= Y_{t-1} + \theta\epsilon_{t-2} && (7) \end{aligned}$$

- Substitute (7) into (5) to get a new expression for ϵ_t :

$$\begin{aligned} Y_t &= \epsilon_t - \theta\left(Y_{t-1} + \theta\epsilon_{t-2}\right) && (8) \\ Y_t &= \epsilon_t - \theta Y_{t-1} - \theta^2\epsilon_{t-2} \end{aligned}$$

MA to AR II

- Repeat this process for two lags:

$$\epsilon_{t-2} = Y_{t-2} - \theta \epsilon_{t-3} \tag{9}$$

- Substitute (9) into (8):

$$\begin{aligned} Y_t &= \epsilon_t - \theta Y_{t-1} - \theta^2 \left(Y_{t-2} - \theta \epsilon_{t-3} \right) \\ &= \epsilon_t - \theta Y_{t-1} - \theta^2 Y_{t-2} + \theta^3 \epsilon_{t-3} \end{aligned}$$

- and so on. Note, eventually, ϵ_{t-i} will approach zero, which leaves:

$$Y_t = \epsilon_t - \sum_{i=1}^{\infty} \theta^i Y_{t-i} \tag{10}$$

- which is an infinite-order AR process.

AR process to infinite-order MA process

- Alternatively, we can express an AR process as a infinite order MA process. Consider the following $AR(1)$ process:

$$Y_t = \phi Y_{t-1} + \epsilon_t \tag{11}$$

- Again, shift back (11) one period:

$$Y_{t-1} = \phi Y_{t-2} + \epsilon_{t-1} \tag{12}$$

- Plugging (12) into equation (11):

$$\begin{aligned} Y_t &= \phi(\phi Y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \phi^2 Y_{t-2} + \phi\epsilon_{t-1} + \epsilon_t \end{aligned} \tag{13}$$

AR to MA II

- Repeat the process for two lags:

$$Y_{t-2} = \phi Y_{t-3} + \epsilon_{t-2} \tag{14}$$

- Substituting equation (14) into equation (13):

$$\begin{aligned} Y_t &= \phi^2 (\phi Y_{t-3} + \epsilon_{t-2}) + \phi \epsilon_{t-1} + \epsilon_t \\ &= \phi^3 Y_{t-3} + \phi^2 \epsilon_{t-2} + \phi \epsilon_{t-1} + \epsilon_t \end{aligned}$$

- Like (10), one process disappears (this time it is the AR process), and another assumes an infinite-order MA process:

$$Y_t = \sum_{i=1}^{\infty} \phi^i \epsilon_{t-i} \tag{15}$$

A Higher Order

- There is no reason AR process would just be of order 1.
  - What would ACF of an AR(2) look like?
  - AR(3)?
- Simulate the process: $y_t = .5y_{t-1} + .3y_{t-2} + \epsilon_t$

Time 0: $0 + 1 \leftarrow OneUnitShock$
Time 1: $.5(1) + .3(0) = .5$
Time 2: $.5(.5) + .3(1)) = .55$
Time 3: $.5(.55) + .3(.5) = .425$
Time 4: $.5(.425) + .3(.55) = .3775$

Identification Problem for MA

Problem There are 2^q different $MA(q)$ coefficients that are *observationally equivalent* in the sense that they possess the same $p(k)$-ACF.

- Consider the $MA(1)$ model with the 2^1 outcomes that are *equivalent* and where $\theta = 0.5$. MODEL A:

$$\begin{aligned} Y_t &= \epsilon_t - \theta\epsilon_{t-1} \\ &= (1 - \theta B)\,\epsilon_t \end{aligned} \tag{16}$$

- Also, consider MODEL B

$$\begin{aligned} Y_t &= \epsilon_t - \frac{1}{\theta}\epsilon_{t-1} \\ &= \left(1 - \frac{1}{\theta}B\right)\epsilon_t \end{aligned} \tag{17}$$

- Both Model A and Model B have a first-order spike–$\rho(1)$–equivalent to,
  $\rho(1) = \frac{-\theta}{(1+\theta^2)}$
  - In Model A, $\theta = 0.5$ and $\rho(0) = 1$, so
    $\rho(1) = \frac{-0.5}{(1+0.5^2)} = -0.4$
  - In Model B, $\theta = 2$, $\rho(0) = 1$, so
    $\rho(1) = \frac{-2}{5} = -0.4$

Hence, we do not know which θ identifies the ACF spike.

MA ID Problem II

Intuition for a Solution Here we make use of the fact that an MA process can be represented as an "inverted" infinite-order AR process to solve this identification problem. We solve the identification problem as follows: Plug in coefficients from the MA process, invert them and set the polynomial equal to zero. The identified model will be the one where the past matters less than the present.

MA ID Solution

- **Solution to the Identification Problem for Model A:** Find the $AR(\infty)$ model from the $MA(1)$ and solve for the AR parameters.

$$Y_t = 0.5Y_{t-1} + 0.5^2 Y_{t-2} + 0.5^3 Y_{t-3} + \dots$$

- **Solution to the Identification Problem for Model B:** As above:

$$Y_t = 2Y_{t-1} + 2^2 Y_{t-2} + 2^3 Y_{t-3} + \dots$$

- Clearly the past in Model A matters less than the present. Model B shows the opposite. The model that is identified is Model A.

PACF

PACF The PACF has the pattern of the AR(k) form:

$$Y_t = \phi_{k_1} Y_{t-1} + \phi_{k_2} Y_{t-2} + \phi_{k_3} Y_{t-3} + \ldots + \phi_{k_k} Y_{t-k} + \epsilon_t$$

An AR(1) provides a single spike such that $Y_t = \phi Y_{t-1}$ and an AR(2) provides $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2}$. The PACF of MA processes will be infinitely lagged and decaying because the inverted MA is an infinitely lagged AR.

Rinse and Repeat

Box-Jenkins Methods

1. Plot the data
2. Difference/Transform to Stationarity
3. Check ACF/PACF and Diagnose
4. Filter and get residuals
5. Test residuals for patterns
6. Repeat until you have white noise.

Independence Tests: Mind Your Q's

- The Q-**statistic** is given by

$$Q = T \sum_{k=1}^{m} r_k^2 \sim \chi^2_{(m-p-q)}$$

where T = total number of time point; m = the number of "events" in the correlogram; k = the index for each autocorrelation parameter; p = the order of the AR component; and q = the order of the MA component. This test is asymptotically valid for the null hypothesis that errors are white noise. Remember that if the null is true, the expected value is $(m - p - q)$.

Box-Ljung

- The **Box-Ljung statistic** is given by

$$T\left(T+2\right)\sum_{k=1}^{m}\frac{r_k^2}{\left(T-k\right)} \sim \chi^2_{(m-p-q)}$$

This statistic represents an attempt to obtain better small sample properties than the Q. Note it is asymptotically valid.

Jarque-Bera Test

The Jarque-Bera test makes use of the following properties of normal distributions:

1. All odd moments greater than 2 equal zero.
2. The 4th central moment (kurtosis) is equal to 3.

Jarque-Bera Strikes Back

- Consider the $r'th$ central moment:
  $\mu_r = T^{-1} \sum_t (X_t - \mu)^r \quad \forall \; r, t$
  - When $r = 3$, the ratio for the coefficient of skewness becomes, $\beta_1 = \frac{\mu_3}{(\mu_2)^{3/2}}$
  - and when $r = 4$, the coefficient of kurtosis becomes, $\beta_2 = \frac{\mu_4}{\mu_2}$.

  If a random variable is normally distributed, skewness equals zero and kurtosis equals 3. In addition, the standard errors of these two sample moments are equal to:

$$\begin{aligned}
\sqrt{Var(\beta_1)} &= \sqrt{\frac{6}{T}} \\
\sqrt{Var(\beta_2)} &= \sqrt{\frac{24}{T}}.
\end{aligned}$$

The Return of Jarque-Bera

The **Jarque-Bera test** is equivalent to:

$$\left(\frac{T}{6}\right)\beta_1 + \left(\frac{T}{24}\right)(\beta_2 - 3)^2 \sim \chi^2_2$$

where under normality, the expected value of the test statistic is two.

- The hypotheses are set up as follows:

$$\begin{aligned} H_0 &: \text{Normality} \\ H_A &: \text{Non-normality} \end{aligned}$$