Lecture 13
Time Series:
Stationarity, AR(p) & MA(q)

1

Time Series: Introduction

• In the early 1970's, it was discovered that simple time series models performed better than the complicated multivarate, then popular, 1960s macro models (FRB-MIT-Penn). See, Nelson (1972).

• The tools? Simple univariate (ARIMA) models, popularized by the textbook of Box & Jenkins (1970).

Q: What is a time series? A time series y_t is a process observed in sequence over time, t = 1,...., *T* $\Rightarrow Y_t = \{y_1, y_2, y_3, ..., y_T\}$

• Main Issue with Time Series: Dependence.

Given the sequential nature of Y_t, we expect y_t & y_{t-1} to be dependent. Depending on assumptions, classical results (based on LLN & CLT) may not be valid.

Time Series: Introduction

- Usually, time series models are separated into two categories:
  - Univariate ($y_t \in R$, it is scalar)
    - $\Rightarrow$ primary model: Autoregressions (ARs).
  - Multivariate ($y_t \in R^m$, it vector-valued).
    - $\Rightarrow$ primary model: Vectot autoregressions (VARs).

- In time series, $\{..., y_1, y_2, y_3, ..., y_T\}$ are jointly RV. We want to model the conditional expectation:

$$E[y_t | F_{t-1}]$$

where $F_{t-1} = \{y_{t-1}, y_{t-2}, y_{t-3}, ...\}$ is the past history of the series.

Time Series: Introduction

- Two popular models for $E[y_t | F_{t-1}]$:

– An autoregressive (AR) process models $E[y_t | F_{t-1}]$ with lagged dependent variables.

- A moving average (MA) process models $E[y_t | F_{t-1}]$ with lagged errors.

- Usually, $E[y_t | F_{t-1}]$ has been modeled as a linear process. But, recently, non-linearities have become more common.

- In general, we assume the error term, ε_t, is uncorrelated with anything, with mean 0 and constant variance, σ^2. We call a process like this a *white noise* (WN) *process*. We denote it as

$$\varepsilon_t \sim WN(0, \sigma^2)$$

Time Series: Introduction

• We want to select an appropriate time series model to forecast y_t. In this class, the choices are AR(p), MA(q) or ARMA(p, q).

• Steps for forecasting:

(1) Identify the appropriate model. That is, determine p, q.

(2) Estimate the model.

(3) Test the model.

(4) Forecast.

• In this lecture, we go over the statistical theory (stationarity, ergodicity and MDS CLT), the main models (AR, MA & ARMA) and tools that will help us describe and identify a proper model

CLM Revisited: Time Series

With autocorrelated data, we get dependent observations. Recall,

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

The independence assumption (**A2'**) is violated. The LLN and the CLT cannot be easily applied, in this context. We need new tools and definitions.

We will introduce the concepts of *stationarity* and *ergodicity*. The ergodic theorem will give us a counterpart to the LLN.

To get asymptotic distributions, we also need a CLT for dependent variables, using the concept of mixing and stationarity. Or we can rely on the *martingale CLT*.

Time Series – Stationarity

• Consider the joint probability distribution of the collection of RVs:

$$F(z_{t_1}, z_{t_2}, \ldots\ldots z_{t_n}) = P(Z_{t_1} \le z_{t_1}, Z_{t_2} \le z_{t_2}, \ldots Z_{t_n} \le z_{t_n})$$

Then, we say that a process is

1st order stationary if $F(z_{t_1}) = F(z_{t_1+k})$ *for any* t_1, k

2nd order stationary if $F(z_{t_1}, z_{t_2}) = F(z_{t_1+k}, z_{t_2+k})$ *for any* t_1, t_2, k

Nth-order stationary if $F(z_{t_1} \ldots . z_{t_n}) = F(z_{t_1+k} \ldots . z_{t_n+k})$ *for any* t_1, t_n, k

• *Definition*. A process is *strongly* (*strictly*) *stationary* if it is a *N*th-order stationary process for any *N*.

Time Series – Moments

• The moments describe a distribution. We calculate the moments as usual:

$$E(Z_t) = \mu_t = \int Z_t f(z_t) dz_t$$

$$Var(Z_t) = \sigma_t^2 = E(Z_t - \mu_t)^2 = \int (Z_t - \mu_t)^2 f(z_t) dz_t$$

$$Cov(Z_{t_1}, Z_{t_2}) = E[(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2})] = \gamma(t_1 - t_2)$$

$$\rho(t_1, t_2) = \frac{\gamma(t_1 - t_2)}{\sqrt{\sigma_{t_1}^2}\sqrt{\sigma_{t_2}^2}}$$

<u>Note</u>: $\gamma(t_1 - t_2)$ is called the auto-covariance function –think of it as a function of $k = t_1 - t_2$. $\gamma(0)$ is the variance.

• Stationarity requires all these moments to be independent of time.

• If the moments are time dependent, we say the series is *non-stationary*.

Time Series – Moments

• For strictly stationary process: $\mu_t = \mu$ and $\sigma_t^2 = \sigma^2$

because $F(z_{t_1}) = F(z_{t_1+k}) \rightarrow \mu_{t_1} = \mu_{t_1+k} = \mu$

provided that $E(|Z_t|) < \infty, \quad E(Z^2{}_t) < \infty$

Then, $F(z_{t_1}, z_{t_2}) = F(z_{t_1+k}, z_{t_2+k}) \Rightarrow$

$\text{cov}(z_{t_1}, z_{t_2}) = \text{cov}(z_{t_1+k}, z_{t_2+k}) \Rightarrow$

$\rho(t_1, t_2) = \rho(t_1 + k, t_2 + k)$

let $t_1 = t - k$ and $t_2 = t$, then

$\rho(t_1, t_2) = \rho(t - k, t) = \rho(t, t + k) = \rho_k$

The correlation between any two RVs depends on the time difference.

Time Series – Weak Stationarity

• A process is said to be *N*-order *weakly stationary* if all its joint moments up to order *N* exist and are time invariant.

• A *Covariance stationary* process (or 2nd order weakly stationary) has:
- constant mean
- constant variance
- covariance function depends on time difference between R.V.

That is, Z_t is covariance stationary if:

$E(Z_t) = \text{constant}$

$Var(Z_t) = \text{constant}$

$Cov(Z_{t_1}, Z_{t_2}) = E[(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2})] = \gamma(t_1 - t_2) = f(t_1 - t_2)$

Time Series – Weak Stationarity

Examples: For all assume $\varepsilon_t \sim WN(0,\sigma^2)$

1) $y_t = \Phi\, y_{t-1} + \varepsilon_t$

$E[y_t] = 0$ (assuming $\Phi \neq 1$)

$Var[y_t] = \sigma^2/(1-\Phi^2)$ (assuming $|\Phi| < 11$)

$E[y_t\, y_{t-1}] = \Phi\, E[y_{t-1}^2]$

$\Rightarrow$ stationary, not time dependent

2) $y_t = \mu + y_{t-1} + \varepsilon_t$ $\Rightarrow y_t = \mu\, t + \Sigma_{j=0 \text{ to } t-1}\, \varepsilon_{t-j} + y_0$

$E[y_t] = \mu\, t + y_0$

$Var[y_t] = \Sigma_{j=0 \text{ to } t-1}\, \sigma^2 = \sigma^2\, t$

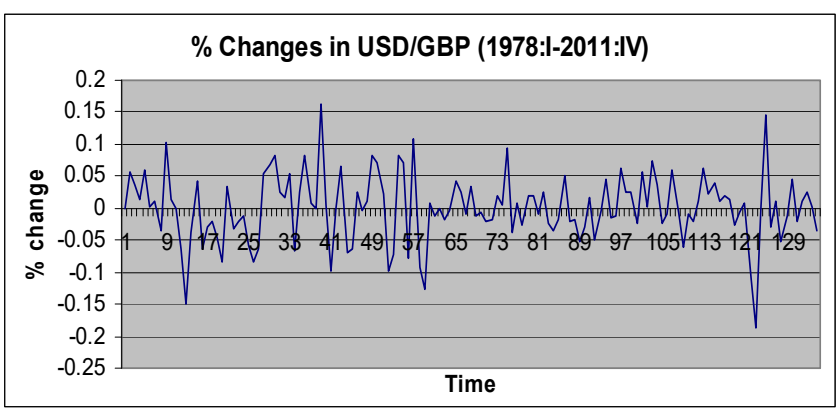
$\Rightarrow$ non-stationary, time dependent

Stationary Series

Examples:

$$y_t = .08 + \varepsilon_t + .4\, \varepsilon_{t-1} \qquad \varepsilon_t \sim WN$$

$$y_t = .13\, y_{t-1} + \varepsilon_t$$

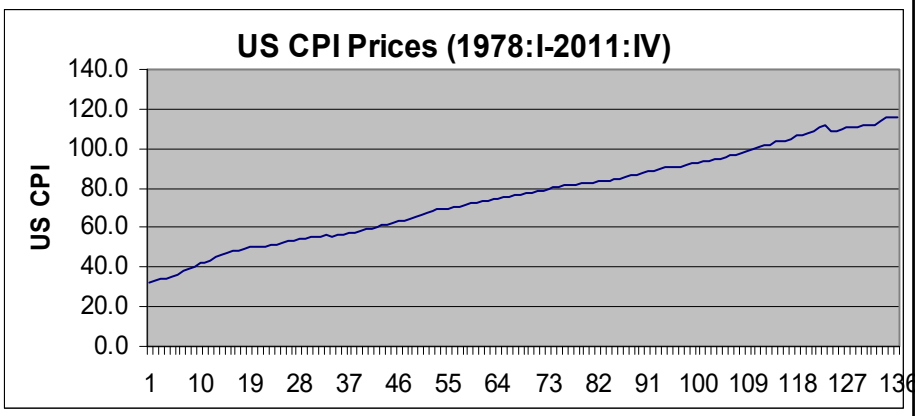


Non-Stationary Series

Examples:

$$y_t = \mu\, t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \qquad \varepsilon_t \sim WN$$

$$y_t = \mu + y_{t-1} + \varepsilon_t \qquad \text{RW with drift}$$



Time Series – Ergodicity

- We want to allow as much dependence as the LLN allows us to do it.
- But, stationarity is not enough, as the following example shows:

- **Example:** Let $\{U_t\}$ be a sequence of *i.i.d.* RVs uniformly distributed on [0, 1] and let Z be N(0,1) independent of $\{U_t\}$.

Define $Y_t = Z + U_t$. Then Y_t is stationary (why?), but

$$\bar{Y}_n = \frac{1}{n}\sum_{t=1}^{n} Y_t \xrightarrow{no} E(Y_t) = \frac{1}{2}$$

$$\bar{Y}_n - Z \xrightarrow{p} \frac{1}{2}$$

The problem is that there is too *much dependence* in the sequence $\{Y_t\}$ (because of Z). In fact the correlation between Y_1 and Y_t is always positive for any value of *t*.

Time Series – Ergodicity of the Mean

• We want to estimate the mean of the process $\{Z_t\}$, $\mu(Z_t)$. But, we need to distinguishing between *ensemble average* and *time average*:

- Ensemble Average $\bar{\bar{z}} = \dfrac{\sum_{i=1}^{m} z_i}{m}$

- Time Series Average $\bar{z} = \dfrac{\sum_{t=1}^{n} z_t}{n}$

Q: Which estimator is the most appropriate?
A: Ensemble Average. But, it is impossible to calculate. We only observe one Z_t.

• Q: Under which circumstances we can use the time average (only one realization of $\{Z_t\}$)? Is the time average an unbiased and consistent estimator of the mean? The *Ergodic Theorem* gives us the answer.

Time Series – Ergodicity of the Mean

• Recall the sufficient conditions for consistency of an estimator: the estimator is asymptotically unbiased and its variance asymptotically collapses to zero.

1. Q: Is the time average is asymptotically unbiased? Yes.

$$E(\bar{z}) = \frac{1}{n}\sum_t E(Z_t) = \frac{1}{n}\sum_t \mu = \mu$$

2. Q: Is the variance going to zero as T grows? It depends.

$$\text{var}(\bar{z}) = \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\text{cov}(Z_t, Z_s) = \frac{\gamma_0}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\rho_{t-s} = \frac{\gamma_0}{n^2}\sum_{t=1}^{n}(\rho_{t-1}+\rho_{t-2}\cdots+\rho_{t-n}) =$$

$$= \frac{\gamma_0}{n^2}[(\rho_0+\rho_1+\cdots+\rho_{n-1})+(\rho_{-1}+\rho_0+\rho_1+\cdots+\rho_{n-2})+$$

$$+\cdots+(\rho_{-(n-1)}+\rho_{-(n-2)}+\cdots+\rho_0)] =$$

Time Series – Ergodicity of the Mean

$$\text{var}(\bar{z}) = \frac{\gamma_0}{n^2} \sum_{k=-(n-1)}^{n-1} (n - |k|)\rho_k = \frac{\gamma_0}{n} \sum_k (1 - \frac{|k|}{n})\rho_k$$

$$\lim_{n\to\infty} \text{var}(\bar{z}) = \lim_{n\to\infty} \frac{\gamma_0}{n} \sum_k (1 - \frac{|k|}{n})\rho_k \xrightarrow{?} 0$$

• If the Z_t were uncorrelated, the variance of the time average would be $O(n^{-1})$. Since independent random variables are necessarily uncorrelated (but not vice versa), we have just recovered a form of the LLN for independent data.

Q: How can we make the remaining part, the sum over the upper triangle of the covariance matrix, go to zero as well?
A: We need to impose conditions on ρ_k. Conditions weaker than "they are all zero;" but, strong enough to exclude the sequence of identical copies.

Time Series – Ergodicity of the Mean

• We use two inequalities to put upper bounds on the variance of the time average:

$$\sum_{t=1}^{n-1} \sum_{k=1}^{n-t} \rho_k \le \sum_{t=1}^{n-1} \sum_{k=1}^{n-t} |\rho_k| \le \sum_{t=1}^{n-1} \sum_{k=1}^{\infty} |\rho_k|$$

Covariances can be negative, so we upper-bound the sum of the actual covariances by the sum of their magnitudes. Then, we extend the inner sum so it covers *all* lags. This might of course be infinite (sequence-of-identical-copies).

• *Definition*: A covariance-stationary process is *ergodic* for the mean if

$$p \lim \bar{z} = E(Z_t) = \mu$$

Ergodicity Theorem: Then, a sufficient condition for ergodicity for the mean is

$$\rho_k \to 0 \quad \text{as } k \to \infty$$

Time Series – Ergodicity of 2nd Moments

• A sufficient condition to ensure ergodicity for second moments is:

$$\sum_k |\rho_k| < \infty$$

A process which is ergodic in the first and second moments is usually referred as *ergodic in the wide sense* .

• *Ergodicity under Gaussian Distribution*
If $\{Z_t\}$is a stationary Gaussian process, $\sum_k |\rho_k| < \infty$

is sufficient to ensure ergodicity for all moments.

Note: Recall that only the first two moments are needed to describe the normal distribution.

Time Series – Ergodicity – Theorems

• We state two essential theorems to the analysis of stationary time series. Difficult to prove in general.

Theorem I
If y_t is strictly stationary and ergodic and $x_t = f(y_t, y_{t-1}, y_{t-2}, ...)$ is a RV, then x_t is strictly stationary and ergodic.

Theorem II (Ergodic Theorem)
If y_t is strictly stationary and ergodic and $E[y_t] < \infty$; then as $T \to \infty$;

$$\frac{1}{\sqrt{T}} \sum_t y_t \xrightarrow{p} E[y_t]$$

• These results allow us to consistently estimate parameters using time-series moments.

Time Series - MDS

• Definition: ε_t is a martingale difference sequence (MDS) if

$$E[\varepsilon_t \mid F_{t-1}]=0.$$

• Regression errors are naturally a MDS. Some time-series processes may be a MDS as a consequence of optimizing behaviour. For example, most asset pricing models imply that asset returns should be the sum of a constant plus a MDS.

• Useful property: ε_t is uncorrelated with any function of the lagged information F_{t-1}. Then, for $k > 0$ $\Rightarrow E[y_{t-k}\, \varepsilon_t] = 0$.

Time Series – MDS CLT

Theorem (MDS CLT)

If u_t is a strictly stationary and ergodic MDS and $E(u_t\, u_t') = \mathbf{\Omega} < \infty$; then as $T \to \infty$;

$$\frac{1}{\sqrt{T}} \sum_t u_t \xrightarrow{d} N(0, \Omega)$$

• Application: Let $x_t = \{y_{t-1}, y_{t-2}, \ldots\}$, a vector of lagged y_t's.

Then $(x_t\, \varepsilon_t)$ is a MDS. We can apply the MDS CLT Theorem. Then,

$$\frac{1}{\sqrt{T}} \sum_t x_t \varepsilon_t \xrightarrow{d} N(0, \Omega), \qquad \Omega = E[x_t x_t' \varepsilon_t^2]$$

• Like in the derivation of asymptotic distribution of OLS, the above result is the key to establish the asymptotic distribution in a time series context.

Autoregressive (AR) Process

• We want to model the conditional expectation of y_t:

$$E[y_t | F_{t-1}]$$

where $F_{t-1} = \{y_{t-1}, y_{t-2}, y_{t-3}, ...\}$ is the past history of the series. We assume the error term, $\varepsilon_t = y_t - E[y_t | F_{t-1}]$, follows a $WN(0,\sigma^2)$.

• An AR process models $E[y_t | F_{t-1}]$ with lagged dependent variables.

• The most common models are AR models. An AR(1) model involves a single lag, while an AR(*p*) model involves *p* lags.

<u>Example</u>: A linear AR(*p*) model (the most popular in practice):

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ...\phi_p y_{t-p} + \varepsilon_t$$

with $E[\varepsilon_t | F_{t-1}]=0$.

AR Process – Lag Operator

• Define the operator *L* as

$$L^k z_t = z_{t-k}$$

• It is usually called *Lag operator*. But it can produces lagged or forward variables (for negative values of *k*). For example:

$$L^{-3} z_t = z_{t+3}$$

• Also note that if *c* is a constant $\Rightarrow L\,c = c$.

• Sometimes the notation for *L* when working as a lag operator is *B (backshift operator),* and when working as a forward operator is *F*.

• Important application: Differencing

$$\Delta z_t = (1-L) z_t = z_t - z_{t-1}$$
$$\Delta^d z_t = (1-L)^d z_t$$

Autoregressive (AR) Process

• Let's work with the linear AR(p) model is:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t$$

$$y_t = \mu + \sum_{i=1}^{p} \phi_i L^i y_t + \varepsilon_t \qquad \text{L : Lag operator}$$

• We can write this process as:

$$\phi(L) y_t = \mu + \varepsilon_t \qquad \text{where } \phi(L) = 1 - \phi_1 L^1 - L^2 \phi_2 - .... - \phi_p L^p$$

ϕ(L) is called the autoregressive polynomial of y_t. Note that

$$y_t = \phi(L)^{-1}(\mu + \varepsilon_t)$$

delivers an infinite sum on the ε_{t-j}'s $\Rightarrow$ an MA(∞) process!

• Q: Can we do this inversion?

AR Process - Stationarity

• Let's compute moments of y_t using the infinite sum (assume μ=0):

$$E[y_t] = \phi(L)^{-1} E[\varepsilon_t] = 0 \qquad \Rightarrow \phi(L) \neq 0$$

$$Var[y_t] = \phi(L)^{-2} Var[\varepsilon_t] \qquad \Rightarrow \phi(L)^{-2} > 0$$

$$E[y_t y_{t-j}] = \gamma(t-j) = E[(\phi_1 y_{t-1} y_{t-j} + \phi_2 y_{t-2} y_{t-j} .... + \phi_p y_{t-p} y_{t-j} + \varepsilon_t y_{t-j})]$$

$$= \phi_1 \gamma(j-1) + \phi_2 \gamma(j-2) + ... + \phi_p \gamma(j-p)$$

where, abusing notation,

$$\phi(L)^{-2} = 1/(1 - \phi_1^2 L^1 - L^2 \phi_2^2 L^2 - .... - \phi_p^2 L^p)$$

Using the *fundamental theorem of algebra*, $\phi(z)$ can be factored as

$$\phi(z) = (1 - r_1^{-1} z)(1 - r_2^{-1} z)...(1 - r_p^{-1} z)$$

where the r_1, ... $r_k \in C$ are the roots of $\phi(z)$. If the ϕ_j's coefficients are all real, the roots are either real or come in complex conjugate pairs.

AR Process – Stationarity

Theorem: The linear AR(p) process is strictly stationary and ergodic if and only if $|r_j|>1$ for all j, where $|r_j|$ is the modulus of the complex number r_j.

• We usually say "*all roots lie outside the unit circle*."

Note: If one of the r_j's equals 1, $\phi(L)$ (& y_t) has a unit root –i.e., $\Phi(1)=0$. This is a special case of *non-stationarity*.

• Recall $\phi(L)^{-1}$ produces an infinite sum on the ε_{t-j}'s. If this sum does not explode, we say the process is *stable*.

• If the process is stable, we can calculate $\delta y_t/\delta\varepsilon_{t-j}$: How much y_t is affected today by an innovation (a shock) $t-j$ periods ago. We call this the *impulse response function (IRF)*.

AR Process – Example: AR(1)

Example: AR(1) process

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

$$E[y_t] = \frac{E[\mu+\varepsilon_t]}{1-\phi} = \frac{\mu}{1-\phi} = \mu^* \qquad \Rightarrow \phi \neq 1 \quad (r_1 \neq 0)$$

$$Var[y_t] = \frac{Var[\varepsilon_t]}{(1-\phi^2)} = \frac{\sigma^2}{(1-\phi^2)}; \qquad \text{since } \sigma^2 > 0 \Rightarrow |\phi| < 1 \ (r_1 > 1)$$

Note: $1/(1-\phi^i) = \sum_{j=0}^{\infty} \phi^{ij} \qquad i = 1,2$

These infinite sums will not explode (*stable* process) if

$|\phi| < 1 \qquad \Rightarrow$ stationarity condition.

Under this condition, we can calculate the impulse response function:

$$\delta y_t/\delta\varepsilon_{t-j} = \phi^j$$

AR Process – Example: AR(1)

• The autocovariance function is:

$$\gamma_k = Cov\left(Y_t, Y_{t-k}\right) = E\left[\left(Y_t - \mu\right)\left(Y_{t-k} - \mu\right)\right]$$
$$\gamma_k = E\left[\left\{\phi\left(Y_{t-1} - \mu\right) + \varepsilon_t\right\}\left(Y_{t-k} - \mu\right)\right]$$
$$\gamma_k = \phi E\left[\left(Y_{t-1} - \mu\right)\left(Y_{t-k} - \mu\right)\right] + E\left[\varepsilon_t\left(Y_{t-k} - \mu\right)\right]$$
$$\gamma_k = \phi\gamma_{k-1} + E\left[\varepsilon_t\left(Y_{t-k} - \mu\right)\right] = \phi\gamma_{k-1}$$

• There is a recursive formula for γ_k:

$$\gamma_1 = \phi\gamma_0$$
$$\gamma_2 = \phi\left(\phi\gamma_0\right) = \phi^2\gamma_0$$
$$\gamma_k = \phi^k\gamma_0$$

• Again, when $|\phi|<1$, the autocovariance do not explode as k increases. There is an exponential decay towards zero.

AR Process – Example: AR(1)

• Note: $\gamma_k = \phi^k\gamma_0$

– when $0 < \phi < 1$ $\Rightarrow$ All autocovariances are positive.

– when $-1 < \phi < 0$ $\Rightarrow$ The sign of γ_k shows an alternating pattern beginning a negative value.

• The AR(1) process has the Markov property: The distribution of Y_t given $\{Y_{t-1}, Y_{t-2}, \ldots\}$ is the same as the distribution of Y_t given $\{Y_{t-1}\}$.

AR Process – Example: AR(2)

Example: AR(2) process

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \quad \Rightarrow (1 - \phi_1 L - \phi_2 L^2) y_t = \mu + \varepsilon_t$$

We can invert $(1 - \phi_1 L - \phi_2 L^2)$ to get the MA(∞) process.

• Stationarity Check

– $E[y_t] = \mu/(1 - \phi_1 - \phi_2) = \mu^*$ $\quad \Rightarrow \phi_1 + \phi_2 \neq 1.$

– $Var[y_t] = \sigma^2/(1 - \phi_1^2 - \phi_2^2)$ $\quad \Rightarrow \phi_1^2 + \phi_2^2 < 1$

Stationarity condition: $|\phi_1 + \phi_2| < 1$

• The analysis can be simplified by rewriting the AR(2) in matrix form as an AR(1):

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \mu \\ 0 \end{bmatrix} + \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix} \Rightarrow \tilde{y}_t = \tilde{\mu} + A\tilde{y}_{t-1} + \tilde{\varepsilon}_t$$

Note: Now, we check $[I - \mathbf{A}^i]$ (i=1,2) for stationarity conditions

AR Process – Stationarity

$$\tilde{y}_t = \tilde{\mu} + A\tilde{y}_{t-1} + \tilde{\varepsilon}_t \qquad \Rightarrow \tilde{y}_t = [I - AL]^{-1} \tilde{\varepsilon}_t$$

Note: Recall $(I - F)^{-1} = \sum_{j=0}^{\infty} F^j = I + F + F^2 + \ldots$

Checking that $[I - \mathbf{A}L]$ is not singular, same as checking that $\mathbf{A}^j$ does not explode. The stability of the system can be determined by the eigenvalues of $\mathbf{A}$. That is, get the λ_i's and check if $|\lambda_i| < 1$ for all i.

$$A = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \Rightarrow |A - \lambda I| = \det \begin{bmatrix} \phi_1 - \lambda & \phi_2 \\ 1 & -\lambda \end{bmatrix} = -(\phi_1 - \lambda)\lambda - \phi_2$$

$$\lambda_i = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

• If $|\lambda_i| < 1$ for all i=1,2, y_t is stable (it does not explode) and stationary. Then: $\lambda_1 \lambda_2 = \phi_2 \Rightarrow |\lambda_1 \lambda_2| = |\phi_2| < 1$

$$\lambda_1 + \lambda_2 = \phi_1 \Rightarrow |\lambda_1 + \lambda_2| = |\phi_1| < 2$$

AR Process – Stationarity

• The autocovariance function is given by:

$$\begin{aligned}\gamma_k &= E\left[(Y_t-\mu)(Y_{t-k}-\mu)\right]\\ &= E\left[(\phi_1(Y_{t-1}-\mu)+\phi_2(Y_{t-2}-\mu)+\varepsilon_t)(Y_{t-k}-\mu)\right]\\ &= \phi_1\gamma_{k-1}+\phi_2\gamma_{k-2}+E\left[\varepsilon_t(Y_{t-k}-\mu)\right]\end{aligned}$$

• Again a recursive formula. Let's get the first autocovariances:

$$\begin{aligned}\gamma_0 &= \phi_1\gamma_{-1}+\phi_2\gamma_{-2}+E\left[\varepsilon_t(Y_t-\mu)\right]\\ &= \phi_1\gamma_1+\phi_2\gamma_2+\sigma^2\\ \gamma_1 &= \phi_1\gamma_0+\phi_2\gamma_1+E\left[\varepsilon_t(Y_{t-1}-\mu)\right]\\ &= \phi_1\gamma_0+\phi_2\gamma_1 \qquad \Rightarrow \gamma_1=\frac{\phi_1\gamma_0}{1-\phi_2}\\ \gamma_2 &= \phi_1\gamma_1+\phi_2\gamma_2+E\left[\varepsilon_t(Y_{t-2}-\mu)\right]\\ &= \phi_1\gamma_1+\phi_2\gamma_2 \qquad \Rightarrow \gamma_2=\frac{\phi_1^2+\phi_2-\phi_2^2}{1-\phi_2}\gamma_0\end{aligned}$$

AR Process – Stationarity

• The AR(2) in matrix AR(1) form is called *Vector AR(1)* or VAR(1).

<u>Nice property</u>: The VAR(1) is Markov -i.e., forecasts depend only on today's data.

• It is straightforward to apply the VAR formulation to any AR(*p*) processes. We can also use the same eigenvalue conditions to check the stationarity of AR(*p*) processes.

AR Process – Causality

• The AR(*p*) model:

$$\phi(L)y_t = \mu + \varepsilon_t \qquad \text{where } \phi(L) = 1 - \phi_1 L^1 - L^2\phi_2 - .... - \phi_p L^p$$

Then, $y_t = \phi(L)^{-1}(\mu+\varepsilon_t)$, $\Rightarrow$ an MA(∞) process!

• But, we need to make sure that we can invert the polynomial ϕ(L).

• When ϕ(L) $\neq$0, we say the process y_t is *causal* (strictly speaking, a *causal function of* $\{\varepsilon_t\}$).

<u>Definition</u>: A linear process $\{y_t\}$ is *causal* if there is a

$$\psi(L) = 1 + \psi_1 L + \psi_2 L^2 + ...$$

$$\text{with} \quad \sum_{j=0}^{\infty} |\psi_j(L)| < \infty$$

$$\text{with} \quad y_t = \psi(L)\varepsilon_t .$$

AR Process – Causality

<u>Example</u>: AR(1) process:

$$\phi(L)y_t = \mu + \varepsilon_t, \qquad \text{where } \phi(L) = 1 - \phi_1 L$$

Then, y_t is causal if and only if:

$$|\phi_1| < 1$$

or

the root r_1 of the polynomial $\phi(z) = 1 - \phi_1 z$ satisfies $|r_1| > 1$.

• Q: How do we calculate the Ψ's coefficients for an AR(*p*)?
A: Matching coefficients:

$$(Y_t - \mu) = \frac{1}{(1 - \phi_1 L)}\varepsilon_t \overset{|\phi_1|<1}{=} \sum_{i=0}^{\infty} \phi_1^{\,i} L^i \varepsilon_t$$

$$= \left(1 + \phi_1 L + \phi_1^{\,2} L^2 + \cdots\right)\varepsilon_t \qquad \Rightarrow \Psi_i = \phi_1^{\,i}, i \geq 0$$

AR Process – Calculating the ψ's

• Example: AR(2) - Calculating the ψ's by matching coefficients.

$$\underbrace{(1 - \phi_1 L - \phi_2 L^2)}_{\Phi(L)}(y_t - \mu) = \varepsilon_t \quad \Rightarrow \quad \Phi(L)\Psi(L) = 1$$

$$\begin{aligned}
\Psi_0 &= 1 \\
\Psi_1 &= \phi_1 \\
\Psi_2 &= \phi_1^2 + \phi_2 \\
\Psi_3 &= \phi_1^3 + 2\phi_1\phi_2 \\
\Psi_j &= \phi_1 \Psi_{j-1} + \phi_2 \Psi_{j-2}, \; j \geq 2
\end{aligned}$$

We can solve these *linear difference equations* in several ways:

- Numerically
- Guess the form of a solution and using an inductive proof
- Using the theory of linear difference equations.

AR Process – Estimation and Properties

• Define

$$x_t = (1 \; y_{t-1} \; y_{t-2} \ldots y_{t-p})$$
$$\beta = (\mu \; \phi_1 \; \phi_2 \; \ldots \; \phi_p)$$

Then the model can be written as $y_t = x_t'\beta + \varepsilon_t$

• The OLS estimator is $b = \hat{\beta} = (X'X)^{-1}X'y$

• Recall that $u_t = x_t\varepsilon_t$ is a MDS. It is also strictly stationary and ergodic.

$$\frac{1}{\sqrt{T}}\sum_t x_t\varepsilon_t = \frac{1}{\sqrt{T}}\sum_t u_t \xrightarrow{p} E[u_t] = 0.$$

• The vector x_t is strictly stationary and ergodic, and by Theorem I so is $x_t x_t'$. Then, by the Ergodic Theorem

$$\frac{1}{\sqrt{T}}\sum_t x_t x_t' = \xrightarrow{p} E[x_t x_t'] = Q$$

AR Process – Estimation and Properties

• *Consistency*

Putting together the previous results, the OLS estimator can be rewritten as:

$$b = \hat{\beta} = (X'X)^{-1}X'y = \beta + \underbrace{\left(\frac{1}{T}\sum_t x_t x_t'\right)^{-1}}_{x_t}\left(\frac{1}{T}\sum_t x_t'\varepsilon_t\right)$$

Then,

$$b = \beta + \left(\frac{1}{T}\sum_t x_t x_t'\right)^{-1}\left(\frac{1}{T}\sum_t x_t'\varepsilon_t\right) \xrightarrow{p} \beta + Q^{-1}0 = \beta$$

$\Rightarrow$ the OLS estimator is consistent.

AR Process – Asymptotic Distribution

• *Asymptotic Normality*

We apply the MDS CLT to $x_t\varepsilon_t$. Then, it is straightforward to derive the asymptotic distribution of the estimator (similar to the OLS case):

Theorem If the AR*(p)* process y_t is strictly stationary and ergodic and $E[y_t^4]$, then as $T \to \infty$;

$$\sqrt{T}(b-\beta) \xrightarrow{d} N(0, Q^{-1}\Omega Q^{-1}), \qquad \Omega = E[x_t x_t' \varepsilon^2]$$

• Identical in form to the asymptotic distribution of OLS in cross-section regression $\Rightarrow$ asymptotic inference is the same.

• The asymptotic covariance matrix is estimated just as in the cross-section case: The sandwich estimator.

AR Process – Bootstrap

• So far, we constructed the bootstrap sample by randomly resampling from the data values (y_t,x_t). This created an *i.i.d* bootstrap sample.

• This is inappropriate for time-series, since we have dependence.

• There are two popular methods to bootstrap time series.

(1) *Model-Based (Parametric) Bootstrap*

(2) *Block Resampling Bootstrap*

AR Process – Bootstrap

(1) *Model-Based (Parametric) Bootstrap*

1. Estimate **b** and residuals **e**:
2. Fix an initial condition $\{y_{t-k+1}, y_{t-k+2}, y_{t-k+3}, ..., y_0\}$
3. Simulate *i.i.d.* draws **e*** from the empirical distribution of the residuals $\{e_1, e_2, e_3, ..., e_T\}$.
4. Create the bootstrap series y_t by the recursive formula

$$y_t^* = \hat{\mu} + \hat{\phi}_1 y_{t-1}^* + \hat{\phi}_2 y_{t-2}^* + ... + \hat{\phi}_p y_{t-p}^* + \varepsilon_t^*$$

<u>Pros</u>: Simple. Similar to the usual bootstrap.

<u>Cons</u>: This construction imposes homoskedasticity on the errors **e*** ; which may be different than the properties of the actual **e**. It also imposes the AR*(p)* as the DGP.

AR Process – Bootstrap

(2) *Block Resampling*

1. Divide the sample into T/m blocks of length m.
2. Resample complete blocks. For each simulated sample, draw T/m blocks.
3. Paste the blocks together to create the bootstrap time-series y_t^*.

Pros: It allows for arbitrary stationary serial correlation, heteroskedasticity, and for model misspecification.

Cons: It may be sensitive to the block length, and the way that the data are partitioned into blocks. May not work well in small samples.

Moving Average Process

- An MA process models $E[y_t | F_{t-1}]$ with lagged error terms. An MA(q) model involves q lags.

- We keep the white noise assumption for ε_t.

Example: A linear MA(q) model:

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} = \mu - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t$$

$$y_t = \mu - \sum_{i=1}^{q} \theta_i L^i \varepsilon_t + \varepsilon_t = \mu + \theta(L)\varepsilon_t \qquad \theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q$$

- Q: Is y_t stationary? Check the moments.

Moving Average Process – Stationarity

• Q: Is y_t stationary? Check the moments. WLOG, assume $\mu=0$.

$$E[y_t] = 0$$
$$Var[y_t] = (1+\theta_1^2+\theta_2^2+...+\theta_q^2)\sigma^2$$
$$\gamma(t-j) = E[y_t y_{t-j}] = E[(-\theta_1\varepsilon_{t-1}y_{t-j} - \theta_2\varepsilon_{t-2}y_{t-j} - .... - \theta_q\varepsilon_{t-q}y_{t-j} + \varepsilon_t y_{t-j})]$$
$$= \sigma^2[\sum_{j=1}^{q}\theta_j\theta_{j+|k|}] \qquad |k| \le q; \quad \text{otherwise} \quad \gamma(t-j) = 0.$$

• It is easy to verify that the sums are finite $\Rightarrow$ MA(q) is stationary.

• Note that an MA(q) process can generate an AR process.

$$y_t = \mu + \theta(L)\varepsilon_t \qquad => \theta(L)^{-1}y_t = \mu^* + \varepsilon_t$$

• We have an infinite sum polynomial on θL. That is, an AR(∞).

$$\sum_{j=0}^{\infty}\pi_j(L)y_t = \mu^* + \varepsilon_t$$

MA Process – Invertibility

• We need to make sure that $\theta(L)^{-1}$ is defined: We require $\theta(L) \neq 0$. When this condition is met, we can write ε_t as a causal function of y_t. We say the MA is *invertible*. For this to hold, we require:

$$\sum_{j=0}^{\infty}|\pi_j(L)| < \infty$$

<u>Definition</u>: A linear process $\{y_t\}$ is *invertible* strictly speaking, an *invertible function of* $\{\varepsilon_t\}$, if there is a

$$\pi(L) = 1 + \pi_1 L + \pi_2 L^2 + ...$$
$$\text{with} \quad \sum_{j=0}^{\infty}|\pi_j(L)| < \infty$$
$$\text{with} \quad \varepsilon_t = \pi(L)y_t.$$

MA Process – Example: MA(1)

• Example: MA(1) process:

$$y_t = \mu + \theta(L)\varepsilon_t \qquad \theta(L) = 1 + \theta_1 L$$

- Moments

$$E(Y_t) = \mu$$
$$\gamma_0 = Var(y_t) = \sigma^2 + \theta_1^2\sigma^2$$
$$\gamma_1 = E[y_t, y_{t-1}] = \theta_1\sigma^2$$
$$\gamma_k = E[y_t, y_{t-k}] = 0, \quad |k| > 1$$

Note: The autocovariance function is zero after lag 1.

– Invertibility: If $|\theta_1| < 1$, we can write $(1+\theta_1 L)^{-1} y_t + \mu^* = \varepsilon_t$

$$(1 - \theta_1 L + \theta_1^2 L^2 + \ldots + \theta_1^j L^j + \ldots) y_t + \mu^* = \mu^* + \sum_{i=1}^{\infty} \pi_i(L) y_t = \varepsilon_t$$

MA Process – Example: MA(2)

• Example: MA(2) process:

$$y_t = \mu + \theta(L)\varepsilon_t \qquad \theta(L) = 1 - \theta_1 L - \theta_2 L^2$$

- Moments

$$E(Y_t) = \mu$$
$$\gamma_k = \begin{cases} \sigma^2(1+\theta_1^2+\theta_2^2), & k = 0 \\ -\theta_1\sigma^2(1-\theta_2), & |k| = 1 \\ -\theta_2\sigma^2, & |k| = 2 \\ 0, & |k| > 2 \end{cases}$$

Note: the autocovariance function is zero after lag 2.

MA Process – Example: MA(2)

- Inveritbility: The roots of $\lambda^2 - \theta_1\lambda - \theta_2 = 0$ all lie inside the unit circle. It can be shown the invertibility condition for MA(2) process is:

$$\theta_1 + \theta_2 < 1$$
$$\theta_2 - \theta_1 < 1$$
$$-1 < \theta_2 < 1$$

MA Process – Estimation

• MA are more complicated to estimate. In particular, there are nonlinearities. Consider an MA(1):

$$y_t = \varepsilon_t + \theta\,\varepsilon_{t-1}$$

The auto-correlation is $\rho_1 = \theta/(1+\theta^2)$. Then, MM estimate of θ satisfies:

$$r_1 = \frac{\hat{\theta}}{(1+\hat{\theta}^2)} \quad \Rightarrow \quad \hat{\theta} = \frac{1 \pm \sqrt{1-4r_1^2}}{2r_1}$$

• A nonlinear solution and difficult to solve.

• Alternatively, if $|\theta| < 1$, we can try $a \in (-1; 1)$,

$$\varepsilon_t(a) = y_t + a\ y_{t-1} + a^2 y_{t-2} + \ldots$$

and look (numerically) for the least-square estimator

$$\hat{\theta} = \arg_a \min\{S_T(a) = \textstyle\sum_{t=1}^{T} \varepsilon_t^{\,2}(a)\}$$

The Wold Decomposition

Theorem - Wold (1938).
Any covariance stationary $\{y_t\}$ has infinite order, moving-average representation:

$$y_t = \sum_{j=0}^{\infty} \psi_j L^j \varepsilon_t + \kappa_t, \qquad \psi_0 = 1,$$

where κ_t : deterministic term (perfectly forecastable). Say, $\kappa_t = \mu$

$$\sum_{j=0}^{\infty} \psi_j^{\ 2} < \infty$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

• y_t is a linear combination of innovations over time.

• A stationary process can be represented as an MA(∞) plus a deterministic "trend."

The Wold Decomposition

<u>Example</u>:
Let $x_t = y_t - \kappa_t$. Then, check moments:

$$E[x_t] = E[y_t - \kappa_t] = \sum_{j=0}^{\infty} \psi_j E[\varepsilon_{t-j}] = 0.$$

$$E[x_t^{\ 2}] = \sum_{j=0}^{\infty} \psi_j^2 E[\varepsilon_{t-j}^2] = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 < \infty.$$

$$E[x_t, x_{t-j}] = E[(\varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \ldots)(\varepsilon_{t-j} + \psi_1 \varepsilon_{t-j-1} + \psi_2 \varepsilon_{t-j-2} + \ldots)$$

$$= \sigma^2 (\psi_j + \psi_1 \psi_{j+1} + \psi_2 \psi_{j+2} + \ldots) = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+2}$$

X_t is a covariance stationary process.

ARMA Process

• A combination of AR(p) and MA(q) processes produces an ARMA(p, q) process:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q}$$

$$= \mu + \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{q} \theta_i L^i \varepsilon_t + \varepsilon_t$$

$$=> \phi(L) y_t = \mu + \theta(L) \varepsilon_t$$

• Usually, we insist that $\phi(L) \neq 0$, $\theta(L) \neq 0$ & that the polynomials $\phi(L)$, $\theta(L)$ have no common factors. This implies it is not a lower order ARMA model.

ARMA Process

<u>Example</u>: Common factors.
Suppose we have the following ARMA(2,3) model $\phi(L) y_t = \theta(L) \varepsilon_t$ with

$$\phi(L) = 1 - .6L + .3L^2$$

$$\theta(L) = 1 - 1.4L + .9L^2 - .3L^3 = (1 - .6L + .3L^2)(1 - L)$$

This model simplifies to: $y_t = (1 - L)\varepsilon_t$ $\Rightarrow$ an MA(1) process.

• Pure AR Representation: $\Pi(L)(y_t - \mu) = a_t \Rightarrow \Pi(L) = \dfrac{\phi_p(L)}{\theta_q(L)}$

• Pure MA Representation: $(y_t - \mu) = \Psi(L) a_t \Rightarrow \Psi(L) = \dfrac{\theta_q(L)}{\phi_p(L)}$

• Special ARMA(p,q) cases:
  – p = 0: MA(q)
  – q = 0: AR(p).

ARMA: Stationarity, Causality and Invertibility

Theorem: If ϕ(L) and $\theta(L)$ have no common factors, a (unique) *stationary* solution to $\phi(L)y_t = \theta(L)\varepsilon_t$ exists if and only if

$$|z| = 1 \Rightarrow \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p \neq 0.$$

This ARMA(*p, q)* model is causal if and only if

$$|z| \leq 1 \Rightarrow \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p \neq 0.$$

This ARMA(*p, q)* model is invertible if and only if

$$|z| \leq 1 \Rightarrow \theta(z) = 1 + \theta_1 z - \theta_2 z^2 + \ldots + \theta_p z^p \neq 0.$$

• Note: Real data cannot be *exactly* modeled using a finite number of parameters. We choose *p, q* to create a good approximated model.

ARMA Process – SDE Representation

• Consider the ARMA(*p,q)* model:

$$\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t$$

Let $x_t = y_t - \mu$ and $w_t = \phi(L)\varepsilon_t$.

Then, $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots. + \phi_p x_{t-p} + w_t$

$\Rightarrow x_t$ is a *p*-th-order linear stochastic difference equation (SDE).

Example: 1st-order SDE (AR(1)): $x_t = \phi\, x_{t-1} + \varepsilon_t$

Recursive solution (Wold form): $x_t = \phi^{t+1} x_{-1} + \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} = \phi^{t+1} x_{-1} + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$

where x_{-1} is an initial condition.

ARMA Process – Dynamic Multiplier

• The dynamic multiplier measurers the effect of ε_t on subsequent values of x_t: That is, the first derivative on the Wold representation:

$$\delta x_{t+j}/\delta\varepsilon_t = \delta x_j/\delta\varepsilon_0 = \psi_j.$$

For an AR(1) process:

$$\delta x_{t+j}/\delta\varepsilon_t = \delta x_j/\delta\varepsilon_0 = \phi^j.$$

• That is, the dynamic multiplier for any linear SDE depends only on the length of time j, not on time t.

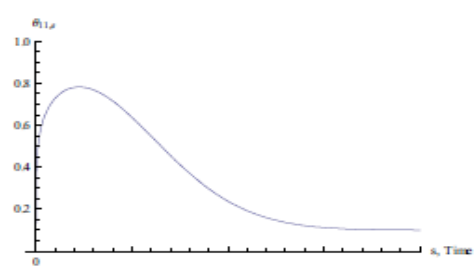
ARMA Process – Impulse Response Function

• The *impulse-response function* (IRF) a sequence of dynamic multipliers as a function of time from the one time change in the innovation, ε_t.

• Usually, IRF are represented with a graph, that measures the effect of the innovation, ε_t, on y_t over time:

$$\delta y_{t+j}/\delta\varepsilon_t + \delta y_{t+j+1}/\delta\varepsilon_t + \delta y_{t+j+21}/\delta\varepsilon_t + \ldots = \psi_j + \psi_{j+1} + \psi_{j+2} + \ldots$$

• Once we estimate the ARMA coefficients, it is easy to draw an IRF.



ARMA Process – Addition

• Q: We add two ARMA process, what order do we get?

• Adding MA processes

$$x_t = A(L)\varepsilon_t$$
$$z_t = C(L)u_t$$
$$y_t = x_t + z_t = A(L)\varepsilon_t + C(L)u_t$$

- Under independence:

$$\gamma_y(j) = E[y_t y_{t-j}] = E[(x_t + z_t)(x_{t-j} + z_{t-j})]$$
$$= E[(x_t x_{t-j} + z_t z_{t-j})] = \gamma_x(j) + \gamma_z(j)$$

- Then, $\gamma(j) = 0$ for $j > \text{Max}(q_x, q_z)$ $\Rightarrow$ y_t is ARMA(0, $max(q_x, q_z)$)

- Implication: MA(2) + MA(1) = MA(2)

ARMA Process – Addition

• Q: We add two ARMA process, what order do we get?

• Adding AR processes

$$(1 - A(L))x_t = \varepsilon_t$$
$$(1 - C(L))z_t = u_t$$
$$y_t = x_t + z_t = ?$$

- Rewrite system as:

$$(1-C(L))(1-A(L))x_t = (1-C(L))\varepsilon_t$$
$$(1-A(L))(1-C(L))z_t = (1-A(L))u_t$$
$$(1-A(L))(1-C(L))y_t = (1-C(L))\varepsilon_t + (1-A(L))u_t = \varepsilon_t + u_t - [C(L)\varepsilon_t + A(L)u_t]$$

- Then, y_t is ARMA(p_x+p_z, max(p_x, p_z))