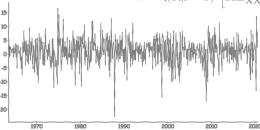


Univariate Time Series Analysis

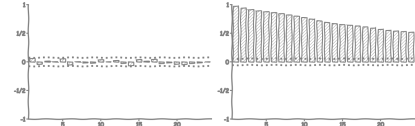
Kevin Sheppard

<https://kevinsheppard.com/teaching/mfe/>

$$\begin{bmatrix} \Delta y_t \\ \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \pi_{p1} & \alpha_{p1} \\ \pi_{p2} & \alpha_{p2} \\ \vdots & \vdots \\ \pi_{p,p-1} & \alpha_{p,p-1} \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta y_{t-2} \\ \vdots \\ \Delta y_{t-p} \end{bmatrix} + \begin{bmatrix} \pi_{p,p} \\ \alpha_{p,p} \\ \vdots \\ \alpha_{p,p} \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta y_{t-2} \\ \vdots \\ \Delta y_{t-p} \end{bmatrix} + \begin{bmatrix} \eta_{t,1} \\ \eta_{t,2} \\ \vdots \\ \eta_{t,p} \end{bmatrix}$$



$$\rho_z = \frac{\gamma_z}{\gamma_\sigma} = \frac{E[(y_t - E[y_t])(y_{t-z} - E[y_{t-z}])]}{V[y_t]} \Rightarrow -2X'(y - X\beta) = -2X'\hat{e} = 0$$



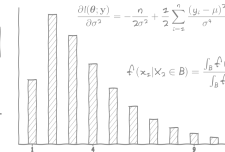
$$\text{Var}_{t+z} = -\mu - \sigma_{t+z} \mathcal{G}_{CF}^{-1}(\alpha) \quad \mathcal{J} = E \left[\frac{\partial l(y; \psi)}{\partial \psi} \frac{\partial l(y; \psi)}{\partial \psi'} \right]$$

$$\hat{f}(x; \rho) = \rho^* (1 - \rho)^{1-x}, \quad \rho \geq 0$$

$$\hat{f}(\rho|x) \propto \rho^* (1 - \rho)^{1-x} \times \frac{\rho^{\alpha-1} (1 - \rho)^{\beta-1}}{B(\alpha, \beta)} = \frac{\rho^{\alpha-1+x} (1 - \rho)^{\beta-1}}{B(\alpha, \beta)}$$

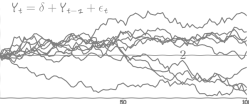
$$\ell(\lambda; y) = -n\lambda + \ln(\lambda) \sum_{i=1}^n y_i - \sum_{j=1}^J \ln(y_j) \quad \hat{S}^{AW} = \hat{\Gamma}_\sigma + \sum_{i=1}^I \frac{1 + i - 1}{1 + 1} (\hat{\Gamma}_i + \hat{\Gamma}'_i)$$

$$Y_i = \beta_1 X_i + \beta_2 X_i I_{[X_i > \kappa]} + \epsilon_i \quad \beta \approx \frac{\partial Y_i}{\partial X_i} \frac{X_i}{Y_i} = E_{y,x}$$

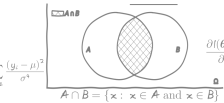


$$\mu_r \equiv E[(X - \mu)^r] = \int_{-\infty}^{\infty} (x - \mu)^r \hat{f}(x) dx \quad \Delta y_t = \phi_0 + \delta_1 t + \gamma y_{t-1} + \sum_{p=2}^P \phi_p \Delta y_{t-p} + \epsilon_t$$

$$\frac{\mu_k}{(\sigma^2)^k} = \frac{E[(X - E[X])^k]}{E[(X - E[X])^2]^k} = E[Z^k] \quad \mathcal{N}(\mu_1 + \beta'(x_2 - \mu_2), \Sigma_{22} - \beta' \Sigma_{22} \beta)$$



$$\sqrt{T}(R(\hat{\theta}) - R(\theta_\theta)) \xrightarrow{d} \mathcal{N} \left(\rho, \frac{\partial R(\theta_\theta)}{\partial \theta'} \Sigma \frac{\partial R(\theta_\theta)}{\partial \theta} \right)$$

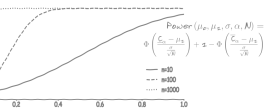
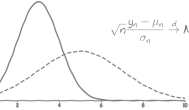


$$\hat{f}(x_2 | x_1 \in B) = \frac{\int_B \hat{f}(x_1, x_2) dx_2}{\int_B \hat{f}_2(x_2) dx_2} \quad \lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^k \ln(1 - \hat{\lambda}_i)$$



$$KS = \max_{\tau} \left| \sum_{i=2}^T \hat{F}_{[y_i < \frac{\tau}{2}]} - \frac{1}{T} \right| \quad \sqrt{n}(\hat{S} - S) \xrightarrow{d} \mathcal{N} \left(\rho, 1 - \frac{\mu \mu_3}{\sigma^4} + \frac{\mu^2 (\mu_4 - \sigma^4)}{4\sigma^6} \right)$$

$$\argmin_{\beta} (y - X\beta)'(y - X\beta) + \lambda \sum_{j=1}^k |\beta_j|$$



$$c(u_1, u_2, \dots, u_k) = \frac{\partial^k C(u_1, u_2, \dots, u_k)}{\partial u_1 \partial u_2 \dots \partial u_k} \quad f(x_1 | x_2 \in B) = \frac{\int_B f(x_1, x_2) dx_2}{\int_B f_2(x_2) dx_2}$$

$$\sigma_t^2 = \omega + \alpha \hat{y}_{t-1}^2 + \beta \sigma_{t-1}^2 \quad \Sigma_t = CC' + AA' \odot \epsilon_{t-1} \epsilon_{t-1}' + BB' \odot \Sigma_{t-1}$$

Time Series Analysis

- Introduction to Time Series Analysis
- Key Concepts in Time Series Analysis
- Autoregressive Moving-Average Processes
- Properties of ARMA Processes
- Autocorrelations and Partial Autocorrelations
- Estimating Autocorrelations and Partial Autocorrelations
- Parameter Estimation
- Model Building
- Forecasting
- Forecast Evaluation
- Nonstationary Time Series
- Random Walks, Unit Roots and Stochastic Trend
- Non-linear Models

Hilary 2021 Teaching Structure

- Viewing pre-recorded content is **mandatory** before the lecture
 - ▶ Restricted to less than 2 hours per week
 - ▶ Prerecorded videos are as short as possible and limited to a single topic
- **Alternatively** read the corresponding section of the course notes
- Lecture focuses on application and problems
- Expanded review section
 - ▶ Review sections appear at content section breaks
 - ▶ Key concepts, questions and problems
 - ▶ Solutions to review problems covered in detail
- Two sets of office hours on Wednesdays
 - ▶ 8.00-9.00 and 16.30-17.30 (UK Local Time)
 - ▶ Weeks 0 to 9

Stochastic Processes

Definition (Stochastic Process)

A stochastic process is a collection of random variables $\{Y_t\}$ defined on a common probability space indexed by a set \mathcal{T} usually defined as \mathbb{N} for discrete time processes or $[0, \infty)$ for continuous time processes.

Basic Example: An i.i.d. time series

$$Y_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$$

More Complex Examples

- Random Walk

$$Y_t = Y_{t-1} + \epsilon_t, \epsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$$

- ARMA(1,1)

$$Y_t = \phi_1 Y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t$$

- Series focuses on ARMA

- GARCH(1,1)

$$Y_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2$$

- GARCH and other non-linear processes later

- Ornstein-Uhlenbeck Process

$$Y(t) = e^{-\beta t} Y(0) + \sigma \int_0^t e^{-\beta(t-s)} dW(s)$$

Review

Stochastic Processes

Key Concepts

Stochastic Process

Questions

- What are the requirements for a sequence of random variables to be a stochastic process?
- Are cross-sectional random variables indexed by i a stochastic process?
- Are the observations of stochastic processes always regularly spaced in time?

Autocovariance

Definition (Autocovariance)

The autocovariance of a covariance stationary scalar process $\{Y_t\}$ is defined

$$\gamma_s = E[(Y_t - \mu)(Y_{t-s} - \mu)]$$

where $\mu = E[Y_t]$. Note that $\gamma_0 = E[(Y_t - \mu)(Y_t - \mu)] = V[Y_t]$.

- Covariance of a process at different points in time
- Otherwise identical to usual covariance

Stationarity

The future resembles the past

Key concept

- Stationarity is a statistically meaningful form of regularity
- First type:

Definition (Covariance Stationarity)

A stochastic process $\{Y_t\}$ is covariance stationary if

$$E[Y_t] = \mu \quad \text{for } t = 1, 2, \dots$$

$$V[Y_t] = \sigma^2 < \infty \quad \text{for } t = 1, 2, \dots$$

$$E[(Y_t - \mu)(Y_{t-s} - \mu)] = \gamma_s \quad \text{for } t = 1, 2, \dots, s = 1, 2, \dots, t - 1$$

- *Unconditional* mean, variance and autocovariance do *not* depend on time

Stationarity

Second type (stronger):

Definition (Strict Stationarity)

A stochastic process $\{Y_t\}$ is strictly stationary if the joint distribution of $\{Y_t, Y_{t+1}, \dots, Y_{t+h}\}$ only depends only on h and not on t .

- *Entire joint distribution* does not depend on time.
- Examples of stationary time series:
 - ▶ i.i.d.: Always strict, covariance if $\sigma^2 < \infty$
 - ▶ i.i.d. sequence of t_2 random variables, strict only
 - ▶ Multivariate normal, both
 - ▶ AR(1): $Y_t = \phi_1 Y_{t-1} + \epsilon_t$, covariance if $|\phi_1| < 1$ and $V[\epsilon_t] < \infty$, strict is ϵ_t is i.i.d.
 - ▶ ARCH(1): $Y_t \sim N(0, \sigma_t^2)$, $\sigma_t^2 = \omega + \alpha Y_{t-1}^2$ both if $\alpha < 1$.

What processes are not stationary?

Nonstationary time series

- Seasonalities, Diurnality, Hebdomadality: $Y_t = \mu + \beta I_{[\text{Quarter}(t) = Q1]} + \epsilon_t$
 - ▶ $E[Y_t]$ is different in Q1 than in other quarters
- Time trends: $Y_t = t + \epsilon_t$
 - ▶ $E[Y_t] = t$
- Random walks: $Y_t = Y_{t-1} + \epsilon_t$
 - ▶ $V[Y_t] = t\sigma^2$
- Processes with structural breaks: $Y_t = \mu_1 + \epsilon_t$ if $t < 1974$, $Y_t = \mu_2 + \epsilon_t$, $t \geq 1974$.
 - ▶ $E[Y_t] = \mu_1 + (\mu_2 - \mu_1)(1 - I_{t < 1974})$

Ergodicity

- Measure of “asymptotic independence”

Theorem (Ergodic Theorem)

If $\{Y_t\}$ is ergodic and the r^{th} moment μ_r is finite, then $T^{-1} \sum_{t=1}^T Y_t^r \xrightarrow{p} \mu_r$.

- Asymptotic independence ensures that averages that use points far apart in time converge to their expected value
- Example of a nonergodic process:

$$Y_t = \mu + \epsilon_t$$

- ▶ $\mu \sim N(0, 1)$ and $\epsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$
- ▶ $E[Y_t] = 0$
- ▶ $T^{-1} \sum_{t=1}^T Y_t \xrightarrow{p} \mu \neq 0$
- ▶ μ has a permanent effect on all Y_t

Review

Stationarity and Ergodicity

Key Concepts

Covariance Stationarity, Strict Stationarity, Ergodicity

Questions

- Why is stationarity important when modeling and forecasting a time series?
- What is the difference between strict and covariance stationarity?
- Why does asymptotic independence help to ensure that a LLN will apply?
- What are the four main sources of non-stationarity in a time series?

Problems

1. Why are the two processes below non-stationary when $\epsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$?
 - a. $Y_t = 0.3t + \epsilon_t$
 - b. $Y_t = 0.7 + 0.2I_{[t > 2020]} + \epsilon_t$.

White noise

Essential Building Block of Time Series

Definition (White Noise)

A process $\{\epsilon_t\}$ is known as white noise if

$$\begin{aligned} E[\epsilon_t] &= 0 && \text{for } t = 1, 2, \dots \\ V[\epsilon_t] &= \sigma^2 < \infty && \text{for } t = 1, 2, \dots \\ E[\epsilon_t \epsilon_{t-j}] &= 0 && \text{for } t = 1, 2, \dots, j \neq 0 \end{aligned}$$

- Not necessarily independent

- ▶ ARCH(1) process $Y_t \sim N(0, \sigma_t^2)$, $\sigma_t^2 = \omega + \alpha Y_{t-1}^2$
- ▶ **Variance** is dependent, mean is not

Linear Time-series Processes

Standard tool of time-series analysis

- *Linear* time series process can always be expressed as

$$Y_t = \delta_t + Y_0 + \sum_{i=0}^t \theta_i \epsilon_{t-i}$$

- ▶ Linear in the errors
- ▶ δ_t is a purely deterministic process
- ▶ $\{\epsilon_t\}$ is a White Noise process
- Example of non-linear processes
 - ▶ GARCH(1,1)

$$Y_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2$$

- ▶ Threshold Autoregression

$$Y_t = \phi_s Y_{t-1} + \epsilon_t, \quad \phi_s = 1 \text{ if } L < Y_{t-1} < U \text{ otherwise } 0.9$$

ARMA Processes

- Inclusive class of all linear time-series processes

Definition (Autoregressive-Moving Average Process)

An Autoregressive Moving Average process with orders P and Q , abbreviated $ARMA(P,Q)$, has dynamics which follow

$$Y_t = \phi_0 + \sum_{p=1}^P \phi_p Y_{t-p} + \sum_{q=1}^Q \theta_q \epsilon_{t-q} + \epsilon_t$$

where ϵ_t is a white noise process with the additional property that $E_{t-1} [\epsilon_t] = 0$.

- $ARMA(1,1)$

$$Y_t = \phi_1 Y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$$

Special case: Moving Average

- ARMA family comprises two sub-classes

Definition (Moving Average Process of Order Q)

A Moving Average process of order Q , abbreviated MA(Q), has dynamics which follow

$$Y_t = \phi_0 + \sum_{q=1}^Q \theta_q \epsilon_{t-q} + \epsilon_t$$

where ϵ_t is white noise series with the additional property that $E_{t-1} [\epsilon_t] = 0$.

- 1st order Moving Average (MA(1))

$$Y_t = \phi_0 + \theta_1 \epsilon_{t-1} + \epsilon_t$$

- Simplest non-degenerate time series process

Special cases of ARMA processes: Autoregression

- Other sub-class of ARMA

Definition (Autoregressive Process of Order P)

An Autoregressive process of order P , abbreviated $AR(P)$, has dynamics which follow

$$Y_t = \phi_0 + \sum_{p=1}^P \phi_p Y_{t-p} + \epsilon_t$$

where ϵ_t is white noise series with the additional property that $E_{t-1}[\epsilon_t] = 0$.

- 1st order Autoregression ($AR(1)$)

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \epsilon_t$$

Moments and Autocovariances

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \epsilon_t$$

- *Unconditional* Mean

$$E[Y_t]$$

- *Unconditional* Variance

$$\gamma_0 = V[Y_t]$$

- Autocovariance

$$\gamma_s = E[(Y_t - E[Y_t])(Y_{t-s} - E[Y_{t-s}])]$$

- *Conditional* Mean

$$E_t[Y_{t+1}] = E[Y_{t+1}|\mathcal{F}_t]$$

- *Conditional* Variance

$$V_t[Y_{t+1}] = E_t[(Y_{t+1} - E_t[Y_{t+1}])^2]$$

Review

Linear Time Series Processes

Key Concepts

White Noise, Linear Stochastic Process, Autoregression, Moving Average, ARMA, Conditional Moment

Questions

- Is White Noise covariance stationary?
- Is White Noise homoskedastic?
- Is an i.i.d. sequence White Noise?
- Is an i.i.d. normal sequence White Noise?
- In what sense is a linear process *linear*?
- Why are linear processes important in the context of covariance stationary time series?
- What is the difference between a conditional and an unconditional moment?
- What is the difference between an AR and an MA model?

How to work with ARMA processes: AR(1)

The MA(∞) Representation

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \epsilon_t$$

- Use backward substitution (assume $|\phi_1| < 1$)

$$\begin{aligned} Y_t &= \phi_0 + \phi_1 Y_{t-1} + \epsilon_t \\ &= \phi_0 + \phi_1(\phi_0 + \phi_1 Y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \phi_0 + \phi_1 \phi_0 + \phi_1^2 Y_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t \\ &= \phi_0 + \phi_1 \phi_0 + \phi_1^2(\phi_0 + \phi_1 Y_{t-3} + \epsilon_{t-2}) + \phi_1 \epsilon_{t-1} + \epsilon_t \\ &= \phi_0 \sum_{j=0}^{\infty} \phi_1^j + \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i} \\ &= \frac{\phi_0}{1 - \phi_1} + \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i} \end{aligned}$$

- $\lim_{s \rightarrow \infty} \sum_{i=0}^s \phi_1^i = 1/(1 - \phi_1)$

Properties of an AR(1)

$$\begin{aligned} E[Y_t] &= E \left[\frac{\phi_0}{1 - \phi_1} + \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i} \right] \\ &= \frac{\phi_0}{1 - \phi_1} + \sum_{i=0}^{\infty} \phi_1^i E[\epsilon_{t-i}] \\ &= \frac{\phi_0}{1 - \phi_1} + \sum_{i=0}^{\infty} \phi_1^i 0 \\ &= \frac{\phi_0}{1 - \phi_1} \end{aligned}$$

- In general AR(P): $E[Y_t] = \frac{\phi_0}{1 - \phi_1 - \phi_2 - \dots - \phi_P}$
- Only sensible if $\phi_1 + \phi_2 + \dots + \phi_P < 1$
- Variance can be shown in same manner
 - ▶ AR(1): $V[Y_t] = \frac{\sigma^2}{1 - \phi_1^2}$
 - ▶ AR(P): $V[Y_t] = \frac{\sigma^2}{1 - \rho_1 \phi_1 - \rho_2 \phi_2 - \dots - \rho_P \phi_P}$
 - ▷ ρ s are autocorrelations

Autocovariance of an AR(1)

$$\begin{aligned} E[(Y_t - E[Y_t])(Y_{t-s} - E[Y_{t-s}])] &= E \left[\left(\sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i} \right) \left(\sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-s-j} \right) \right] \\ &= E \left[\left(\underbrace{\sum_{i=0}^{s-1} \phi_1^i \epsilon_{t-i}}_{\text{After } t-s} + \underbrace{\sum_{k=s}^{\infty} \phi_1^k \epsilon_{t-k}}_{t-s \text{ and later}} \right) \left(\sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-s-j} \right) \right] \\ &= \phi_1^s \frac{\sigma^2}{1 - \phi_1^2} \end{aligned}$$

- Full details in notes
- The autocovariance *function*

$$\gamma_s = \phi_1^{|s|} \left\{ \frac{\sigma^2}{1 - \phi_1^2} \right\}$$

- Autocovariance declines geometrically with the lag length
- Requires $\phi_1^2 < 1$ to exist
 - ▶ Same condition as the mean

Stationarity of ARMA processes

- Primarily interested in covariance stationarity
- Stationarity depends on parameters of *AR* portion
- AR(0) or finite order MA: always stationary
- AR(1) or ARMA(1,Q): $Y_t = \phi_1 Y_{t-1} + \text{MA} + \epsilon_t$
 - ▶ $|\phi_1| < 1$
- AR(P) or ARMA(P,Q) $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_P Y_{t-P} + \text{MA} + \epsilon_t$
- Rewrite $Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_P Y_{t-P} = \text{MA} + \epsilon_t$
- Easy to determine using the characteristic equation and corresponding characteristic roots

The characteristic equation

Definition (Characteristic Equation)

Let Y_t follow a P^{th} order linear difference equation

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_P Y_{t-P} + x_t$$

which can be rewritten as

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_P Y_{t-P} = \phi_0 + x_t$$

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_P L^P) Y_t = \phi_0 + x_t$$

The characteristic equation of this process is

$$z^P - \phi_1 z^{P-1} - \phi_2 z^{P-2} - \dots - \phi_{P-1} z - \phi_P = 0$$

- Key is in the forming of the characteristic equation and its roots
- L is known as “lag operator”

Characteristic roots

Definition (Characteristic Root)

Let

$$z^P - \phi_1 z^{P-1} - \phi_2 z^{P-2} - \dots - \phi_{P-1} z - \phi_P = 0$$

be the characteristic polynomial associated with some P^{th} order linear difference equation. The P characteristic roots, c_1, c_2, \dots, c_P are defined as the solution to this polynomial

$$(z - c_1)(z - c_2) \dots (z - c_P) = 0.$$

- The roots are c_1, c_2, \dots, c_P
- AR(P) or ARMA(P,Q) is covariance stationary if $|c_j| < 1$ for all j
- If complex, $|c_j| = |a_j + b_j i| = \sqrt{a^2 + b^2}$ (complex modulus)

Characteristic roots example

- Difficult to determine by inspection

Example 1

$$Y_t = .1Y_{t-1} + .7Y_{t-2} + .2Y_{t-3} + \epsilon_t$$

- Characteristic equation

$$z^3 - .1z^2 - .7z^1 - .2$$

- Roots: 1, $-.5$, and $-.4 \Rightarrow$ nonstationary

Example 2

$$Y_t = 1.7Y_{t-1} - .72Y_{t-2} + \epsilon_t$$

- Characteristic equation

$$z^2 - 1.7z^1 + .72$$

- Roots: $.9$ and $.8 \Rightarrow$ stationary

Review

Properties of ARMA Models

Key Concepts

Backward Substitution, Characteristic Equation, Characteristic Root

Questions

- What role so the MA component play in determining stationarity?
- What is the key condition for stationarity of an ARMA model?
- What is complex modulus and why is it needed?

Problems

1. Which of the models listed below are covariance stationary?
 - a. $Y_t = 1.8Y_{t-1} - 0.8Y_{t-2} + \epsilon_t$
 - b. $Y_t = 0.4 - 0.75Y_{t-1} - 0.25Y_{t-2} + \epsilon_t$
 - c. $Y_t = 10 + \sum_{j=1}^{100} 0.01Y_{t-j} + \epsilon_t$
2. Write the ARMA(1,1) $Y_t = \phi_1 Y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$ as a function of $\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-h}$ and Y_{t-h} using backward substitution.
3. Use backward substitution to write the model $Y_t = -0.5\epsilon_{t-1} + \epsilon_t$ as an AR(∞) using the relationship that $Y_{t-1} = -0.5\epsilon_{t-2} + \epsilon_{t-1}$ implies $\epsilon_{t-1} = Y_{t-1} + 0.5\epsilon_{t-2}$.

Autocorrelations and the ACF

- Autocorrelations are a **key element** of model building

Definition (Autocorrelation)

The autocorrelation of a covariance stationary scalar process is defined

$$\rho_s = \frac{\gamma_s}{\gamma_0}$$

where $\gamma_s = \text{E}[(Y_t - \mu)(Y_{t-s} - \mu)]$.

- Measures the correlation of a process at different points in time
- AR(1):

$$\rho_s = \phi_1^s$$

- One of two possibilities
 - Decay geometrically if $0 < \phi_1 < 1$
 - Oscillate and decay $-1 < \phi_1 < 0$

Partial Autocorrelations (PACF)

- Partial Autocorrelation is the other **key element** of model building
- More complicated than autocorrelations:
- Regression interpretation of s^{th} partial autocorrelation:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_{s-1} Y_{t-s+1} + \varphi_s Y_{t-s} + \epsilon_t$$

- φ_s is the s^{th} partial autocorrelation
 - ▶ Population (not sample) value of φ_s
- AR(1):

$$\varphi_s = \begin{cases} \phi_1^{|s|} & \text{for } s = -1, 0, 1 \\ 0 & \text{otherwise} \end{cases}$$

- Partial autocorrelation function maps the parameters of a process to the s^{th} autocorrelation, $\varphi(s)$

Using the ACF and PACF to categorize processes

- ACF and PACF are useful when choosing models

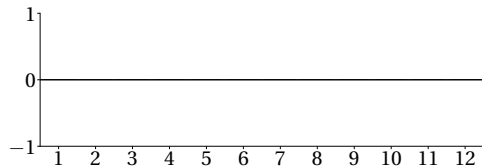
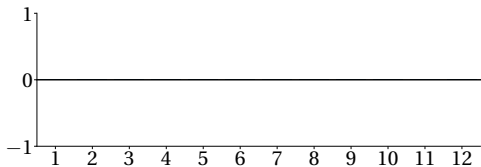
Process	ACF	PACF
White Noise	All 0	All 0
AR(1)	$\rho_s = \phi_1^s$	0 beyond lag 1
AR(P)	Decays toward zero exponentially	Non-zero through lag P, 0 thereafter
MA(1)	$\rho_1 \neq 0, \rho_s = 0, s > 1$	Decays toward zero exponentially
MA(Q)	$\rho_s \neq 0, s \leq Q, \rho_s = 0, s > Q$	Decays toward zero exponentially, possible oscillating
ARMA(P,Q)	Exponential Decay	Exponential Decay

Autocorrelation for ARMA processes

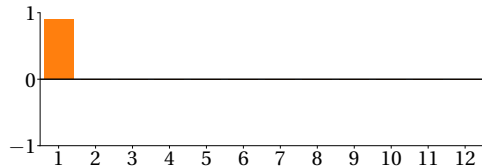
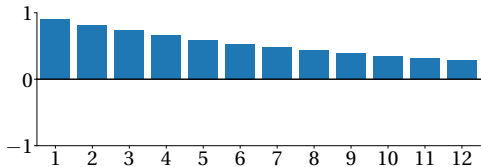
ACF

PACF

White Noise

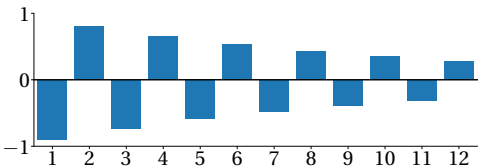


AR(1), $\phi_1 = 0.9$



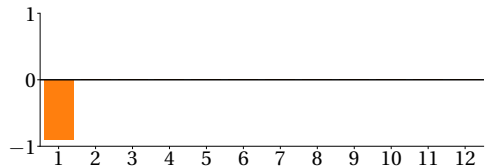
Autocorrelation for ARMA processes

ACF

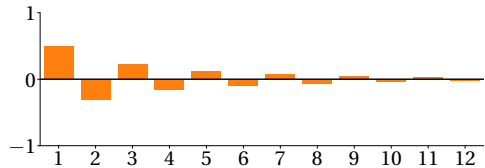
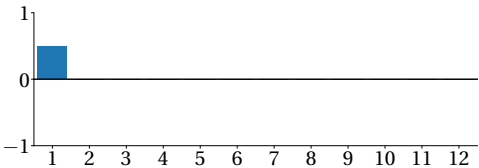


AR(1), $\phi_1 = -0.9$

PACF



MA(1), $\theta_1 = 0.8$

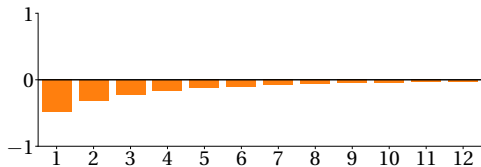
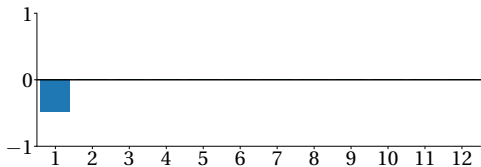


Autocorrelation for ARMA processes

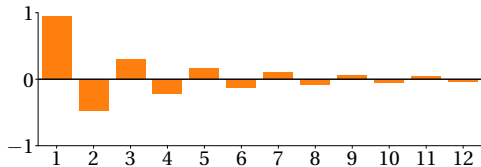
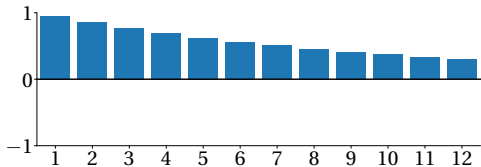
ACF

PACF

MA(1), $\theta_1 = -0.8$

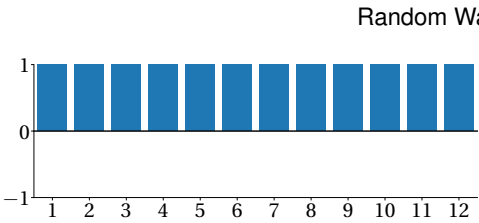


ARMA(1,1), $\phi_1 = 0.9$, $\theta_1 = -0.8$

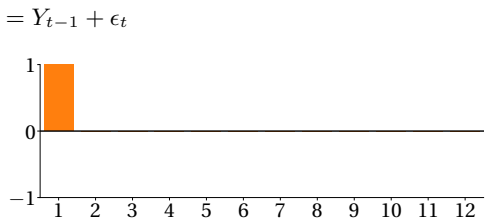


Autocorrelation for ARMA processes

ACF



PACF



Review

Autocorrelation and Partial Autocorrelation

Key Concepts

Autocorrelation, Partial Autocorrelation

Questions

- What is the difference between the h -lag autocorrelation and the h -lag partial autocorrelation?
- When are the autocorrelation and partial autocorrelation always the same for any DGP?
- What shape would you expect in the ACF and PACF of an AR(3)?
- What shape would you expect in the ACF and PACF of an MA(12)?

Problems

1. What is the ACF and PACF of an AR(1) $Y_t = \phi_1 Y_{t-1} + \epsilon_t$?
2. What is the ACF of an MA(2) $Y_t = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \epsilon_t$?

Sample ACF and PACF

- Sample autocorrelations

$$\hat{\rho}_s = \frac{\sum_{t=s+1}^T Y_t^* Y_{t-s}^*}{\sum_{t=1}^T Y_t^{*2}} = \frac{\hat{\gamma}_s}{\hat{\gamma}_0}$$

- ▶ $Y_t^* = Y_t - \bar{Y}$ where $\bar{Y} = T^{-1} \sum_{t=1}^T Y_t$

- Some prefer the small-sample-size corrected version

$$\hat{\rho}_s = \frac{\sum_{t=s+1}^T Y_t^* Y_{t-s}^*}{\sqrt{\sum_{t=s+1}^T Y_t^{*2} \sum_{t=1}^{T-s} Y_t^{*2}}}.$$

- Sample partial autocorrelations

- ▶ Run regression to estimate $\hat{\phi}_s$

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_s Y_{t-s} + \epsilon_t$$

- More efficient ways to compute PACF using Yule-Walker (see notes)

Testing autocorrelations and partial ACs

- Inference on autocorrelations:

$$V[\hat{\rho}_s] = T^{-1} \quad \text{for } s = 1$$

$$= T^{-1} \left(1 + 2 \sum_{j=1}^{s-1} \hat{\rho}_j^2 \right) \quad \text{for } s > 1$$

- Standard t -stats

$$\frac{\hat{\rho}_s}{\sqrt{V[\hat{\rho}_s]}} \stackrel{A}{\sim} N(0, 1).$$

- Inference on partial autocorrelations:

$$V[\hat{\varphi}_s] \approx T^{-1}$$

- Standard t -stats

$$T^{\frac{1}{2}} \hat{\varphi}_s \stackrel{A}{\sim} N(0, 1)$$

Testing multiple autocorrelations

- Testing multiple autocorrelations: Ljung-Box Q , $H_0 : \rho_1 = \dots = \rho_s = 0$

$$Q = T(T+2) \sum_{k=1}^s \frac{\hat{\rho}_k^2}{T-k} \sim \chi_s^2$$

- **Note:** Not heteroskedasticity robust, use LM test for serial correlation

Definition (LM test for serial correlation)

Under the null, $E[Y_t^* Y_{t-j}^*] = 0$ for $1 \leq j \leq s$. The LM-test for serial correlation is constructed by defining the score vector $\mathbf{s}_t = Y_t^* [Y_{t-1}^* \ Y_{t-2}^* \ \dots \ Y_{t-s}^*]'$,

$$LM = T \bar{\mathbf{s}}' \hat{\mathbf{S}}^{-1} \bar{\mathbf{s}} \xrightarrow{d} \chi_s^2$$

where $\bar{\mathbf{s}} = T^{-1} \sum_{t=1}^T \mathbf{s}_t$, $\hat{\mathbf{S}} = T^{-1} \sum_{t=1}^T \mathbf{s}_t \mathbf{s}_t'$ and $Y_t^* = Y_t - \bar{Y}$ where $\bar{Y} = T^{-1} \sum_{t=1}^T Y_t$.

Review

Sample Autocorrelations and Partial Autocorrelations

Key Concepts

Sample Autocorrelation, Sample Partial Autocorrelation, Ljung-Box Test, LM Test for Serial Correlation

Questions

- What is the asymptotic distribution of estimated autocorrelations and partial autocorrelations?
- Where does the rule-of-thumb $2/\sqrt{T}$ come from when plotting sample autocorrelations?
- What is the difference between the Q -test and an LM test for serial correlation?
- If you computed a sample autocorrelation in Excel using the `correl` function by copying and shifting a variable by h places, would you get the usual sample autocorrelation estimator?

Conditional MLE

- Conditional MLE assuming distribution of $Y_t|Y_{t-1}, \epsilon_{t-1}, Y_{t-2}, \epsilon_{t-2}, \dots$ is $N(0, \sigma^2)$
- If $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-Q}$ are observable, identical to least squares

$$\underset{\phi, \theta}{\operatorname{argmin}} \sum_{t=P+1}^T (Y_t - \phi_0 - \phi_1 Y_{t-1} - \dots - \phi_P Y_{t-P} - \theta_1 \epsilon_{t-1} - \dots - \theta_Q \epsilon_{t-Q})^2$$

- Ignore distribution of Y_1, \dots, Y_P in fit
 - ▶ Finite sample effects, asymptotically irrelevant
- If $\epsilon_{P-1}, \dots, \epsilon_{P-Q}$ are observable, can recursively compute $\epsilon_P, \dots, \epsilon_T$ for a set of parameters ϕ, θ
- Overcome missing initial shocks by assuming $\epsilon_{P-1} = \dots = \epsilon_{P-Q} = 0$

Ordinary Least Squares

- If $Q = 0$, conditional MLE simplifies

$$\underset{\phi}{\operatorname{argmin}} \sum_{t=P+1}^T (Y_t - \phi_0 - \phi_1 Y_{t-1} - \dots - \phi_P Y_{t-P})^2$$

- Conditional MLE is identical to OLS
- Inference is identical
- Use classical or White's covariance estimator as appropriate
- Can also incorporate deterministic terms such as time trends while maintaining simplicity of OLS

Exact MLE

- Define the vector of data

$$\mathbf{y} = [Y_1, Y_2, \dots, Y_{T-1}, Y_T]'$$

- $\mathbf{\Gamma}$ be the T by T covariance matrix of \mathbf{y}

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \gamma_3 & \dots & \gamma_{T-2} & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{T-3} & \gamma_{T-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{T-4} & \gamma_{T-3} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \gamma_{T-2} & \gamma_{T-3} & \gamma_{T-4} & \gamma_{T-5} & \dots & \gamma_0 & \gamma_1 \\ \gamma_{T-1} & \gamma_{T-2} & \gamma_{T-3} & \gamma_{T-4} & \dots & \gamma_1 & \gamma_0 \end{bmatrix}$$

- The joint likelihood of \mathbf{y}

$$f(\mathbf{y}|\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) = (2\pi)^{-\frac{T}{2}} |\mathbf{\Gamma}|^{-\frac{T}{2}} \exp\left(-\frac{\mathbf{y}'\mathbf{\Gamma}^{-1}\mathbf{y}}{2}\right)$$

- Log-likelihood

$$l(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2; \mathbf{y}) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln |\mathbf{\Gamma}| - \frac{1}{2} \mathbf{y}'\mathbf{\Gamma}^{-1}\mathbf{y}$$

Review

Parameter Estimation

Key Concepts

Conditional Maximum Likelihood, Exact Maximum Likelihood

Questions

- How are missing initial innovations addressed in conditional MLE?
- What is the key advantage of exact MLE over conditional MLE?
- When does conditional MLE reduce to OLS?
- How is the autocovariance matrix computed in exact MLE?

Model building the Box-Jenkins way

- Model building is similar to cross-section regression
- Can use same techniques
 - ▶ General to Specific or Specific to General
 - ▶ Information criteria: AIC, BIC
- Box-Jenkins is dominant methodology, 2-steps
 - ▶ Identification: Use ACF and PACF to choose model
 - ▶ Estimation: Estimate model and do diagnostic checks
- Two principles
 - ▶ Parsimony
 - ▶ Invertibility

Strategies

■ General to Specific

- ▶ Fit largest specification
- ▶ Drop regressor with largest p-value
- ▶ Refit
- ▶ Stop if all p-values indicate significance using a size of α
 - ▷ α is the econometrician's choice

■ Specific to General

- ▶ Fit all specifications with a single variable
- ▶ Retain variable with smallest p-value
- ▶ Extend this model adding on additional variables one at a time
- ▶ Stop if the p-values of all excluded variables are larger than α

Information Criteria

- Information Criteria

- ▶ Akaike Information Criterion (AIC)

$$AIC = \ln \hat{\sigma}^2 + k \frac{2}{T}$$

- ▶ Schwartz (Bayesian) Information Criterion (SIC/BIC)

$$BIC = \ln \hat{\sigma}^2 + k \frac{\ln T}{T}$$

- Both have versions suitable for likelihood based estimation
- Reward for better fit: Reduce $\ln \hat{\sigma}^2$
- Penalty for more parameters: $k \frac{2}{T}$ or $k \frac{\ln T}{T}$
- Choose model with smallest IC
 - ▶ AIC has fixed penalty \Rightarrow inclusion of extraneous variables
 - ▶ BIC has larger penalty if $\ln T > 2$ ($T > 7$)

Model Diagnostics

- Important to assess whether your model “fits”
 - ▶ Are the residuals white noise?
 - ▷ Eye-ball test
 - ▷ Ljung-Box Q stat or LM serial correlation test of $H_0 : \rho_1 = \dots = \rho_s = 0$.
 - ▷ SACF/SPACF of the residuals
 - ▶ Are there any large outliers?
 - ▷ Eye-ball test
- What to do if there are problems?
 - ▶ Use SPACF/SACF to repeat Box-Jenkins and augment your model with correct dynamics to pick up problem
 - ▶ Repeat diagnostics
- Concern: Repeated testing may render critical values misleading

Review

Model Selection

Key Concepts

Invertibility, Parsimony, AIC, BIC

Questions

- How are the ACF and PACF used to identify candidate models?
- How does GtS differ in an ARMA from application to a linear regression?
- Which chooses a larger model, AIC or BIC, and why?
- What property should residuals have from a well specified model?
- What use is the parsimony principle?
- What does invertibility ensure?

The information set and the law of iterated expectations

- Information set: \mathcal{F}_t
- Contains a lot of information!
 - ▶ Every time t *measurable* event
 - ▶ Observed variables: prices, returns, GDP, interest rates, FX rates
 - ▶ Functions of these
 - ▶ Excludes variables which are latent: volatility
- Conditional expectation:

$$E[Y_{t+1}|\mathcal{F}_t]$$

Conditional Variance

$$V[Y_{t+1}|\mathcal{F}_t]$$

- ▶ Shorthand $E_t[Y_{t+1}]$ and $V_t[Y_{t+1}]$
- Law of Iterated Expectation (LIE):

$$E_t[E_{t+1}[Y_{t+2}]] = E_t[Y_{t+2}]$$

- ▶ Monday's belief about what Tuesday's belief about Wednesday is the same as Monday's belief of Wednesday

Forecasting

- A h -step ahead forecast, $\hat{Y}_{t+h|t}$, is designed to minimize a loss function
 - ▶ MSE: $(Y_{t+h} - \hat{Y}_{t+h|t})^2$
 - ▶ MAD: $|Y_{t+h} - \hat{Y}_{t+h|t}|$
 - ▶ Quad-Quad: $\alpha_1(Y_{t+h} - \hat{Y}_{t+h|t})^2 + \alpha_2 I_{[Y_{t+h} - \hat{Y}_{t+h|t} < 0]}(Y_{t+h} - \hat{Y}_{t+h|t})^2$
 - ▷ Asymmetric if $\alpha_1 \neq \alpha_2$

The MSE Optimal Forecast is the conditional mean

- Let $Y_{t+h}^* = E_t[Y_{t+h}]$
- Let \tilde{Y}_{t+h} be any other value

$$\begin{aligned} E_t[(Y_{t+h} - \tilde{Y}_{t+h})^2] &= E_t\left[\left((Y_{t+h} - Y_{t+h}^*) + (Y_{t+h}^* - \tilde{Y}_{t+h})\right)^2\right] \\ &= E_t[(Y_{t+h} - Y_{t+h}^*)^2 + 2(Y_{t+h} - Y_{t+h}^*)(Y_{t+h}^* - \tilde{Y}_{t+h}) + (Y_{t+h}^* - \tilde{Y}_{t+h})^2] \\ &= V_t[Y_{t+h}] + 2E_t[(Y_{t+h} - Y_{t+h}^*)(Y_{t+h}^* - \tilde{Y}_{t+h})] + E_t[(Y_{t+h}^* - \tilde{Y}_{t+h})^2] \\ &= V_t[Y_{t+h}] + 2(Y_{t+h}^* - \tilde{Y}_{t+h})E_t[(Y_{t+h} - Y_{t+h}^*)] + E_t[(Y_{t+h}^* - \tilde{Y}_{t+h})^2] \\ &= V_t[Y_{t+h}] + 2(Y_{t+h}^* - \tilde{Y}_{t+h}) \cdot 0 + E_t[(Y_{t+h}^* - \tilde{Y}_{t+h})^2] \\ &= V_t[Y_{t+h}] + (Y_{t+h}^* - \tilde{Y}_{t+h})^2 \end{aligned}$$

Forecasting

- MSE optimal forecast for an AR(1):

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t$$

$$\begin{aligned} E_t[Y_{t+1}] &= E_t[\phi_1 Y_t + \epsilon_{t+1}] \\ &= \phi_1 E_t[Y_t] + E_t[\epsilon_{t+1}] \\ &= \phi_1 Y_t + 0 \end{aligned}$$

$$\begin{aligned} E_t[Y_{t+2}] &= E_t[\phi_1 Y_{t+1} + \epsilon_{t+2}] \\ &= \phi_1 E_t[Y_{t+1}] + E_t[\epsilon_{t+2}] \\ &= \phi_1 (\phi_1 Y_t) + 0 \\ &= \phi_1^2 Y_t + 0 \end{aligned}$$

Note: Long-run forecast is always $E[Y_t]$ for a covariance stationary process

Forecast Errors

$$\begin{aligned}V_t[Y_{t+1}] &= E_t \left[(Y_{t+1} - E_t [Y_{t+1}])^2 \right] \\&= E_t \left[(\phi Y_t + \epsilon_{t+1} - \phi Y_t)^2 \right] \\&= E_t [\epsilon_{t+1}^2] = \sigma^2 \text{ if homoskedastic}\end{aligned}$$

$$\begin{aligned}V_t[Y_{t+2}] &= E_t \left[(Y_{t+2} - E_t [Y_{t+2}])^2 \right] \\&= E_t \left[(\phi^2 Y_t + \phi \epsilon_{t+1} + \epsilon_{t+2} - \phi^2 Y_t)^2 \right] \\&= E_t \left[(\phi \epsilon_{t+1} + \epsilon_{t+2})^2 \right] \\&= \phi^2 E_t [\epsilon_{t+1}^2] + E_t [\epsilon_{t+2}^2] = (1 + \phi^2) \sigma^2 \text{ if homoskedastic}\end{aligned}$$

Note: Long-run forecast error variance is always $V[Y_t]$ for a covariance stationary process

Review

Forecasting

Key Concepts

Mean Square Error, Conditional Expectation

Questions

- How is the MSE optimal forecast related to the conditional mean? What about the conditional median?
- What is the key principle for producing multi-step forecasts?
- What does the long-run forecast for a covariance stationary time series always converge to? What is the long-run variance of the error?

Problems

1. What are the first three forecasts from the model $Y_t = \phi_0 + \phi_1 Y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$?
2. What are the first three forecasts errors?
3. What is the variance of the first three forecast errors?

Forecast evaluation

Mincer-Zarnowitz regressions

- Objective Forecast Evaluation

$$Y_{t+h} = \alpha + \beta \hat{Y}_{t+h|t} + \eta_t$$

- $H_0 : \alpha = 0, \beta = 1, H_1 : \alpha \neq 0 \cup \beta \neq 1$
 - ▶ Use any test: Wald, LR, LM
- Can be generalized to include any variable available when the forecast was produced

$$Y_{t+h} = \alpha + \beta \hat{Y}_{t+h|t} + \gamma \mathbf{x}_t + \eta_t$$

- $H_0 : \alpha = 0, \beta = 1, \gamma = \mathbf{0}, H_1 : \alpha \neq 0 \cup \beta \neq 1 \cup \gamma_j \neq 0$
- \mathbf{x}_t *must* be in the time t information set
- Important when working with macro data

Relative evaluation: Diebold-Mariano

- Two forecasts, $\hat{Y}_{t+h|t}^A$ and $\hat{Y}_{t+h|t}^B$
- Two losses, $l_t^A = (Y_{t+h} - \hat{Y}_{t+h|t}^A)^2$ and $l_t^B = (Y_{t+h} - \hat{Y}_{t+h|t}^B)^2$
 - ▶ Losses do not need to be MSE
- If equally good or bad, $E[l_t^A] = E[l_t^B]$ or $E[l_t^A - l_t^B] = 0$
- Define $\delta_t = l_t^A - l_t^B$

Relative evaluation: Diebold-Mariano

- Implemented as a t -test that $E[\delta_t] = 0$
- $H_0 : E[\delta_t] = 0$, $H_1^A : E[\delta_t] < 0$, $H_1^B : E[\delta_t] > 0$
 - ▶ Composite alternative
 - ▶ Sign indicates which model is favored

$$DM = \frac{\bar{\delta}}{\sqrt{\widehat{V[\delta]}}} = \frac{T^{-1} \sum_{t=1}^T \delta_t}{\sqrt{\frac{\hat{\sigma}_{NW}^2}{T}}}$$

- One complication: $\{\delta_t\}$ cannot be assumed to be uncorrelated, so a more complicated variance estimator is required
- Newey-West covariance estimator:

$$\hat{\sigma}_{NW}^2 = \hat{\gamma}_0 + 2 \sum_{l=1}^L \left[1 - \frac{l}{L+1} \right] \hat{\gamma}_l$$

Implementing a Diebold-Mariano Test

$$DM = \frac{\bar{\delta}}{\sqrt{\widehat{V[\bar{\delta}]}}}$$

Algorithm (Diebold-Mariano Test)

1. *Using the two forecasts, $\hat{Y}_{t+h|t}^A$ and $\hat{Y}_{t+h|t}^B$, compute $\delta_t = l_t^A - l_t^B$*
2. *Run the regression*

$$\delta_t = \beta + \eta_t$$

3. *Use a Newey-West covariance estimator (`cov_type="HAC"`)*
4. *T-test $H_0 : \beta = 0$ against $H_1^A : \beta < 0$, and $H_1^B : \beta > 0$*
5. *Reject if $|t| > C_\alpha$ where C_α is the critical value for a 2-sided test using a normal distribution with a size of α . If significant, reject in favor of model A if test statistic is negative or in favor of model B if test statistic is positive.*

Review

Forecast Evaluation

Key Concepts

Objective Forecast Evaluation, Relative Forecast Evaluation, Mincer-Zarnowitz Test, Diebold-Mariano Test, Newey-West Variance Estimator

Questions

- What is the difference between objective and relative forecast evaluation?
- Why is a Newey-West covariance estimator used in Diebold-Mariano test?
- How is rejection of the null in a Newey-West test different from most tests?
- Why is a multi-step forecast be sensitive to a future realization of the time series between the current period and the forecast horizon?
- How is a MZ regression transformed to an Augmented MZ regression?

Nonstationarity defined

- Any series which is not stationary is nonstationary
- Four major types
 - ▶ Seasonality
 - ▷ Only slightly problematic
 - ▷ Can often be analyzed using standard tools and Box-Jenkins
 - ▶ Deterministic trends: growth over time
 - ▷ Linear
 - ▷ Polynomial
 - ▷ Exponential
 - ▶ Random walks or unit roots
 - ▶ Structural breaks

Deterministic trends

- Trending series can be decomposed

$$Y_t = \text{deterministic trend} + \text{stationary component} + \text{noise}$$

- Two major types

- ▶ Polynomial

$$Y_t = \phi_0 + \delta_1 t + \delta_2 t^2 + \dots + \delta_s t^s + \epsilon_t$$

- ▶ Linear (important special case)

$$Y_t = \phi_0 + \delta_1 t + \epsilon_t$$

- ▶ Exponential

$$\ln Y_t = \phi_0 + \delta_1 t + \epsilon_t$$

- Solution is to detrend

- ▶ Detrended series is a stationary process
- ▶ Standard model building on residuals
- ▶ Can directly include time trends in ARMA models

The Lag Operator

- The Lag Operator is a useful tool in time series
- Simplifies expressing complex models with seasonal dynamics
- Key properties
 1. $LY_t = Y_{t-1}$
 2. $L^2Y_t = LY_{t-1} = L(LY_t) = Y_{t-2}$
 3. $L^aL^b = L^{(a+b)}$
 4. $Lc = c$ where c is a constant

Seasonality

- Seasonality is technically a form of non-stationarity
 - Mean explicitly depends on the quarter, month, day or minute
- Three types:

Definition (Seasonality)

Data are said to be seasonal if they exhibit a non-constant deterministic pattern on an annual basis.

Definition (Hebdomadality)

Data which exhibit day-of-week deterministic effects are said to be hebdomadal.

Definition (Diurnality)

Data which exhibit intra-daily deterministic effects are said to be diurnal.

Seasonality

- Simpler to think of processes with seasonality as having two models
 - ▶ Short-run AR and MA dynamics
 - ▶ Seasonal AR and MA dynamics
- Model building is standard with these two goals in mind
- Also consider seasonal deterministic terms
 - ▶ Seasonal dummy variables
 - ▶ Seasonal Fourier series

ARMA Modeling of Seasonality

Four Components

- Observation AR

$$(1 - \phi_1 L) Y_t = \phi_0 + \epsilon_t$$

- Seasonal AR

$$(1 - \phi_s L^s) Y_t = \phi_0 + \epsilon_t$$

- Observation MA

$$Y_t = \phi_0 + (1 + \theta_1 L^1) \epsilon_t$$

- Seasonal MA

$$Y_t = \phi_0 + (1 + \theta_s L^s) \epsilon_t$$

- Combined Model

$$(1 - \phi_1 L) (1 - \phi_s L^s) Y_t = (1 + \theta_1 L^1) (1 + \theta_s L^s) \epsilon_t$$

$$\begin{aligned} Y_t = & \phi_0 + \phi_1 Y_{t-1} + \phi_s Y_{t-s} - \phi_1 \phi_s Y_{t-s-1} \\ & + \theta_1 \epsilon_{t-1} + \theta_s \epsilon_{t-s} + \theta_1 \theta_s \epsilon_{t-s-1} + \epsilon_t \end{aligned}$$

ARMA Modeling of Seasonality

Four Components

- Generalizes to higher orders of each term
- Known as SARIMA($p, 0, q$) \times ($P, 0, Q, s$)
- Imposes restrictions on parameters due to multiplication of terms
- Can estimate unrestricted equivalent

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_s Y_{t-s} + \phi_{s+1} Y_{t-s-1} + \theta_1 \epsilon_{t-1} + \theta_s \epsilon_{t-s} + \theta_{s+1} \epsilon_{t-s-1} + \epsilon_t$$

- Can test $H_0 : \phi_{s+1} = \phi_1 \phi_s \cap \theta_{s+1} = \theta_1 \theta_s$

Review

Seasonality

Key Concepts

Seasonality, Lag Operator, SARIMA, Deterministic Trend, Exponential Trend

Questions

- How can seasonality be modeled in an ARMA model?
- Define diurnality, hebdomadality and seasonality.
- What are seasonal determinist terms and how do they differ from seasonal AR and MA terms?
- What is an exponential trend?
- What do the orders in a SARIMA mean?
- How could a standard AR be used to model a time series with a seasonal AR component?

Stochastic trends

- Stochastic trends are similar to deterministic trends
 - ▶ Dominant feature of a process

$$Y_t = \text{stochastic trend} + \text{stationary component} + \text{noise}$$

- Most common stochastic trend is a unit root
- There are others (generally non-linear)
- Removed using stochastic detrending (differencing)
 - ▶ Meaningfully different than deterministic detrending

Short-run Dynamics in a Unit Root process

- Unit root processes, in the long-run, behave like random walks
- In the short run, can have stationary dynamics

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \epsilon_t$$

- If this process contains a unit root, $\phi_1 + \phi_2 + \phi_3 = 1$
- Can see the SR dynamics by differencing

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-2} - \phi_3 Y_{t-2} + \phi_3 Y_{t-3} + \epsilon_t$$

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-2} - \phi_3 \Delta Y_{t-2} + \epsilon_t$$

$$Y_t - Y_{t-1} = (\phi_1 + \phi_2 + \phi_3 - 1) Y_{t-1} - (\phi_2 + \phi_3) \Delta Y_{t-1} - \phi_3 \Delta Y_{t-2} + \epsilon_t$$

$$\Delta Y_t = \pi_1 \Delta Y_{t-1} + \pi_2 \Delta Y_{t-2} + \epsilon_t$$

What's the problem with unit roots?

- Unit roots cause a number of problems
 - ▶ Exploding variance: $V[Y_t] = t\sigma^2$
 - ▶ Parameter estimates converge at different rates
 - ▶ Hypothesis tests have non-standard distributions
 - ▶ No mean reversion in long-run forecasts
 - ▶ Spurious regression
- Crucial to understand whether a process is stationary or contains a unit root
- Often has large economic consequences
 - ▶ PPP
 - ▶ Covered interest rate parity
 - ▶ Carry trades

Testing for unit roots

- Dickey-Fuller looks like a standard t -test

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t$$

- $H_0 : \phi_1 = 1, H_1 : \phi_1 < 1$
- Impose the null

$$Y_t - Y_{t-1} = \phi_1 Y_{t-1} - Y_{t-1} + \epsilon_t$$

$$\Delta Y_t = (\phi_1 - 1)Y_{t-1} + \epsilon_t$$

$$\Delta Y_t = \gamma Y_{t-1} + \epsilon_t$$

- New $H_0 : \gamma = 0, H_1 : \gamma < 0$
- Test with t -stat
- Augmented Dickey Fuller (ADF) captures short run dynamics as well

$$\Delta Y_t = \gamma Y_{t-1} + \rho_1 \Delta Y_{t-1} + \rho_2 \Delta Y_{t-2} + \dots + \rho_P \Delta Y_{t-P} + \epsilon_t$$

- Lags of ΔY_{t-1} needed to ensure $\epsilon_t \sim WN(0, \sigma^2)$, also reduce variance of residuals

The problem

- t -stat is no longer asymptotically normal
- Requires Dickey-Fuller distribution
 - ▶ Most software packages contain the correct critical value
- Many processes with unit roots also contain deterministic components
- Asymptotic distribution depends on choice of model:

$$\Delta Y_t = \gamma Y_{t-1} + \sum_{p=1}^P \phi_p \Delta Y_{t-p} + \epsilon_t \quad (\text{No trend})$$

$$\Delta Y_t = \delta_0 + \gamma Y_{t-1} + \sum_{p=1}^P \phi_p \Delta Y_{t-p} + \epsilon_t \quad (\text{Constant, linear in } Y_t)$$

$$\Delta Y_t = \delta_0 + \delta_1 t + \gamma Y_{t-1} + \sum_{p=1}^P \phi_p \Delta Y_{t-p} + \epsilon_t \quad (\text{Constant, quadratic in } Y_t)$$

- More deterministic regressors lower the critical value
- Reject null of unit root if t -stat of γ is **negative** and below the critical value

Important considerations

- Unit root tests are well known for having low power
- Power = 1-Pr(type II)
 - ▶ Chance you don't reject when alternative is true
- Some suggestions
 - ▶ Use a loose model selection when choosing the number of lags of ΔY_{t-j} , e.g. AIC
 - ▶ Be conservative in excluding deterministic regressors.
 - ▷ Including a constant or time-trend when absent hurts power
 - ▷ Excluding a constant or time-trend when present results in **no power**
 - ▶ More powerful tests than the ADF are available: DF-GLS
 - ▶ Visually inspect the data and differenced data
 - ▶ Use a general-to-specific search
- Number of differences needed is the *order of integration*
 - ▶ Integrated of Order 1 or $I(1)$: Y_t is nonstationary but ΔY_t is stationary
 - ▶ $I(d)$: Y_t is nonstationary, $\Delta^j Y_t$ also nonstationary when $j < d$, $\Delta^d Y_t$ is stationary

Seasonal Differencing

- Seasonal series should use seasonal differencing

$$\Delta_s Y_t = Y_t - Y_{t-s}$$

- Complete SARIMA(P, D, Q) \times (P_s, D_s, Q_s, s) model
 - ▶ D is order of observational difference
 - ▶ D_s is order of seasonal difference
 - ▶ P and Q are observational AR and MA orders
 - ▶ P_s and Q_s are seasonal AR and MA orders
- Special Cases
 - ▶ ARMA(P, Q): $D = D_s = P_s = Q_s = 0$
 - ▶ ARIMA(P, D, Q): $D_s = P_s = Q_s = 0$
 - ▶ SARMA(P, Q) \times (P_s, Q_s, s): $D = D_s = 0$

Review

Unit Roots and Integration

Key Concepts

Unit Root, Integrated Process, $I(1)$, Augmented Dickey-Fuller Test, Seasonal Difference

Questions

- What happens if a relevant deterministic term is omitted in a ADF test?
- What is the effect of including an unnecessary deterministic in an ADF test?
- How should you decide how many lags of the differenced variable to include in an ADF test?
- When should you use seasonal differencing?
- What is the relationship between a random walk and a unit root process?
- What are the consequences of ignoring a unit root when modeling a time series?

Nonlinear Models for the mean

- *Linear* time series process

$$Y_t = Y_0 + \sum_{i=0}^t \theta_i \epsilon_{t-i}$$

- Alternatives
 - ▶ Markov Switching Autoregression (MSAR)
 - ▶ Threshold Autoregression (TAR) and Self-exciting Threshold Autoregression (SETAR)
 - ▶ Many, many others
- Nonlinear models can capture different dynamics
 - ▶ *State-dependent parameters*

$$Y_t = \phi_0^{s_t} + \phi_1^{s_t} Y_{t-1} + \sigma^{s_t} \epsilon_t$$

- ▶ Models differ in how s_t evolves

Markov Switching Example

- Two states, H and L

$$Y_t = \begin{cases} \phi^H + \epsilon_t \\ \phi^L + \epsilon_t \end{cases}$$

- States evolve according to a 1st order Markov Chain

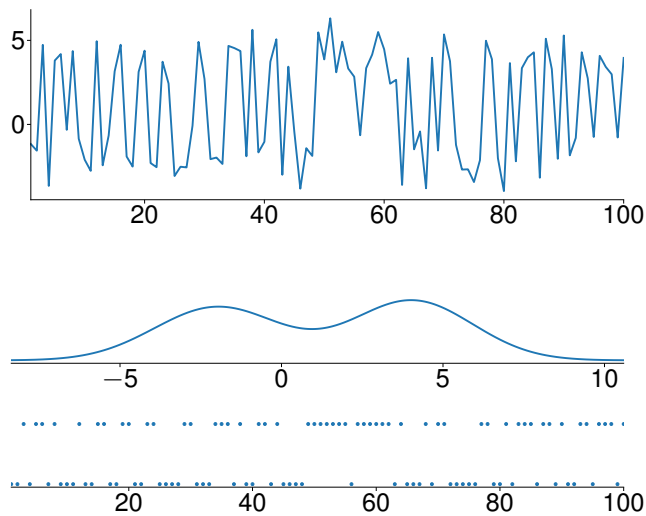
$$\{s_t\} = \{H, H, H, L, L, L, H, L, \dots\}$$

- Transition Probabilities

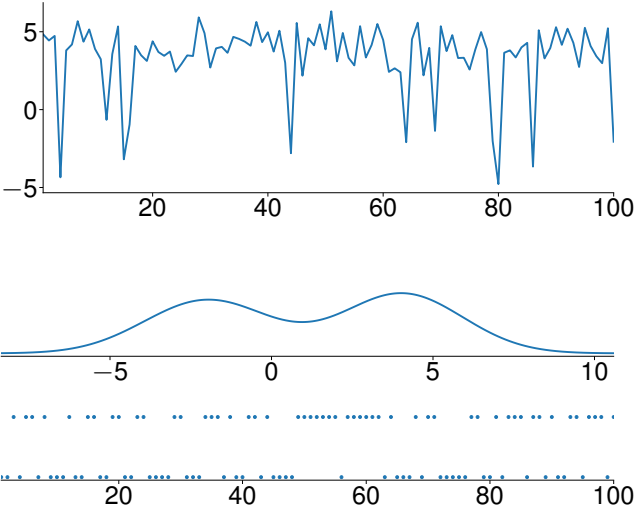
$$\begin{bmatrix} p_{HH} & p_{HL} \\ p_{LH} & p_{LL} \end{bmatrix} = \begin{bmatrix} p_{HH} & 1 - p_{LL} \\ 1 - p_{HH} & p_{LL} \end{bmatrix}$$

- ▶ p_{HH} is the probability $s_{t+1} = H$ given $s_t = H$.
- Model will switch between a high mean state and a low mean state
- Models like this are very flexible and nest ARMA
 - ▶ Successful in financial econometrics for asset allocation, volatility modeling, modeling series with business-cycle length patterns: GDP

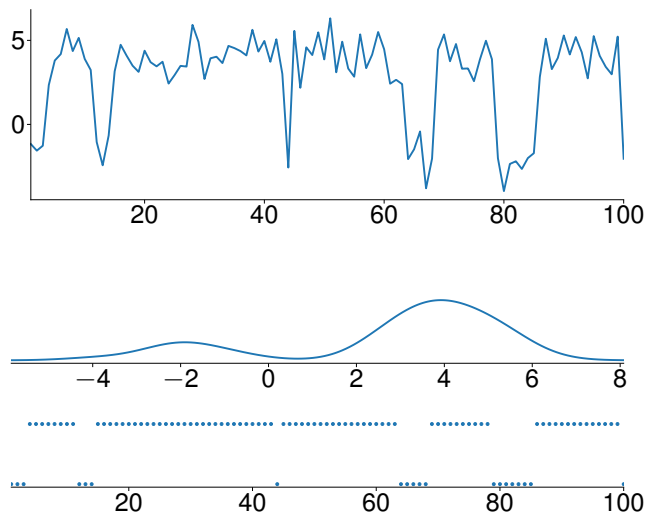
Markov Switching: i.i.d. Mixture



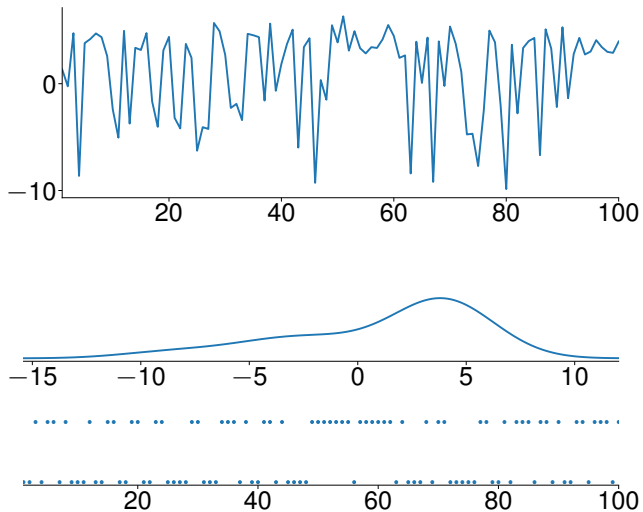
Markov Switching: Symmetric Persistent



Markov Switching: Asymmetric Persistent



Markov Switching: Different Variances



Review

Non-linear Time Series Models

Key Concepts

Self-exciting Threshold Autoregression, Markov Switching Processes

Questions

- It is always necessary to consider nonlinear models to model covariance stationary time series?
- What advantages might a nonlinear model have over a linear model when modeling a covariance stationary time series?