Lesson 1: Univariate and Multi-Variate Models

Module: Advanced Time-Series Models

Introduction

- Univariate models of prediction and forecasting
- White noise iid process
- Stationarity of AR processes
- Unconditional mean of the AR process
- ACF and PACF plots: AR process
- Autocorrelation coefficient
- Autocorrelation coefficient: numerical example

Introduction

- Moving average (MA) processes
- A simple example of MA(2) process
- ACF and PACF plots: MA process
- ARMA process
- Building ARMA model

Introduction

- Determining the accuracy of the forecast
- Introduction to time-series stationarity
- Time series and stationarity
- Stochastic non-stationarity
- Deterministic non-stationarity
- Time-series and stationarity: visual examination

Introduction

- Testing for unit-roots
- Introduction to mean reversion
- Cointegration
- Error correction models (ECM)
- Engle-Granger approach to ECM and cointegration

Univariate Models of Prediction and Forecasting

Univariate Models of Prediction and Forecasting

- Univariate time-series models are a class of specifications where one attempts to model and predict financial variables using only information contained in their own past values and possibly current and past values of an error term
- Time series models are usually a-theoretical. (HFT trading)
- An important class of models is autoregressive integrated moving average (ARIMA) models.
- Often in HFT, fundamental variables affecting P_t are not available in the same frequency as P_t.

Introduction to Time-Series Stationarity

Time-Series and Stationarity

- A strictly stationary process
  - A time series process {Xt} is said to be strictly stationary if the joint probability distribution of a sequence {yt1 to ytn} is the same as that for {yt1+m to ytn+m} ∀ m: i.e.,

$$F\{y_{t_1}, \ldots, y_{t_n}\} = F\{y_{t_{1+m}}, \ldots, y_{t_{n+m}}\}$$

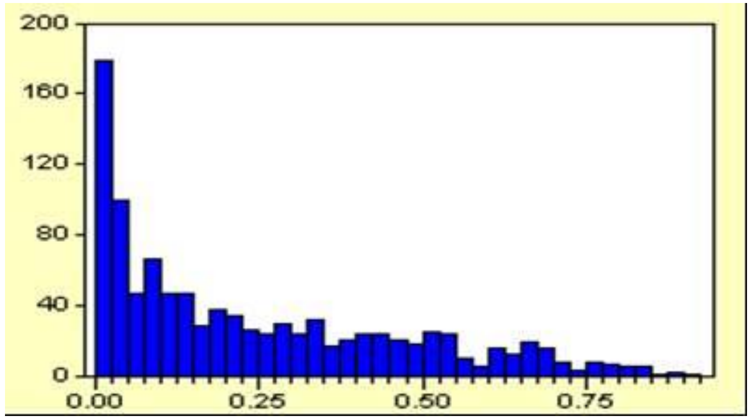
Time-Series and Stationarity

- A weakly stationary process
  - A weakly stationary (or covariance stationary) process satisfies the following three equations:
    - $E(y_t) = u, t = 1,2,3 \dots \infty$
    - $E(y_t - u)(y_t - u) = \sigma^2 < \infty$
    - $E\left(y_{t_1} - u\right)\left(y_{t_2} - u\right) = \gamma_{t_2 - t_1}$ $for\ all\ t_1\ and\ t_2$
  - To summarize, a weak stationary process should have a constant mean, constant variance, and a constant auto-covariance structure.
  - What is auto-covariance: $E(y_t - u)(y_{t+s} - u) = \gamma_s$; $s = 0,1,2, \dots$
  - Autocorrelation: $\tau_s = \frac{\gamma_s}{\gamma_0}$, s = 0,1,2, … ; τ_s lies in the interval ±1

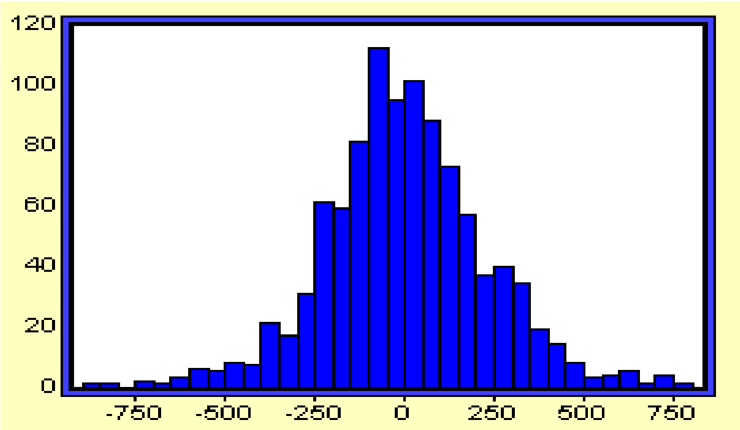
Time-Series and Stationarity

- Problems with non-stationary series
  - Shocks to non-stationary systems do not die away and are highly persistent
  - Estimation with non-stationary data can lead to spurious regressions, i.e., artificially high t-stat. and adjusted R-sq, even if the variables are not related to each other. t-ratio will not follow t-distribution

Adj. R sq



T-stat



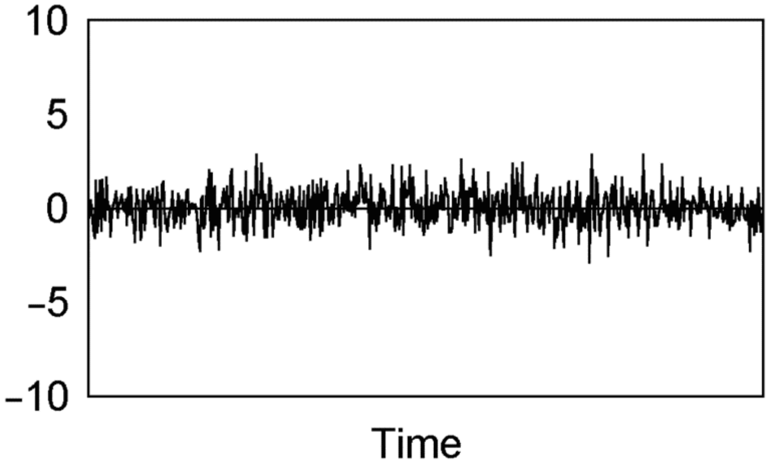
Chris Brooks. *Introductory Econometrics for Finance*. 4th Edition. Chapter 6

White Noise IID Process

White Noise IID Process

- It can be defined as follows

1. E(μ_t) = 0 ,   t = 1,2,...,∞
2. Var(μ_t)= σ^2
3. $E[(\mu_{t_k} - u)(\mu_{t_m} - u)] = \gamma_{t\,k - t_m} = 0\ for\ all\ t_k$ $and\ t_m$ for all $t_k \neq t_m$



- Thus, white noise has a constant (zero) mean, finite constant variance, and is uncorrelated across time
- Often, we assume a fix distribution (e.g., normal iid), then it is a special case of white noise process

Introduction to AR Processes

Autoregressive (AR) Processes

- Current value of the variable y_t depends upon its own previous values
- An AR model of order p, AR(p) can be expressed as below
  - $y_t = \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + u_t$
  - $y_t = \mu + \sum_{i=1}^{p} \varphi_i y_{t-i} + u_t$ or $y_t = \mu + \sum_{i=1}^{p} \varphi_i L^i y_t + u_t$
  - $\varphi(L) y_t = \mu + u_t$; where $\varphi(L) = \left(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p\right)$
- Here u_t is the white noise iid error term [distributed as N $(0, \sigma^2)$]

Stationarity of AR Processes

Stationarity of AR Processes

- $\varphi(L)y_t = \mu + u_t$; where $\varphi(L) = \left(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p\right)$
- $y_t = \varphi(L)^{-1}u_t$ [Ignore the drift]
- As the lag length is increased, the function $\varphi(L)^{-1}$ should converge to zero
- If we compute the roots of this characteristic equation
- $\varphi(L) = \left(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p\right) = 0$
- What if a root is more or less than 1, what if it is equal to 1
- This is non-stationarity

Stationarity of AR Processes: A simple Example

Stationarity of AR Processes

- For example, consider the process below
  - $y_t = y_{t-1} + u_t$
  - $(1 - L)Y_t = u_t$
- Characteristic equation =1-z=0
- What do we infer?

Unconditional Mean of the AR process

- If the AR (p) process is stationary the unconditional mean of the process is given below
- $E(y_t) = \frac{\mu}{1-\varphi_1-\varphi_2-\cdots-\varphi_p}$
- For example, for AR(1) process: $y_t = \mu + \varphi_1 y_{t-1} + u_t$
- $E(y_t) = \frac{\mu}{1-\varphi_1}$

Autocorrelation Coefficient

Autocorrelation Coefficient

- $E[(y_t - E(y_t))(y_{t+s} - E(y_{t+s}))] = \gamma_s$; s = 0,1,2, ….
- When s = 0, the autocovariance at lag zero is obtained, which is the variance of y_t; γ_0
- Autocovariances also depend on the units of measurement and are not easy to interpret directly
- Autocorrelation measure $\tau_s = \frac{\gamma_s}{\gamma_0}$
- These autocorrelations (τ_s) lie in the interval ± 1

Autocorrelation Coefficient

- τ_s are ~ (approximately or asymptotically) normally distributed. $\hat{\tau_s} \sim N(0,1/T\ )$
- For example, a 95% non-rejection region would be given by the following interval: $-1.96 * 1/\sqrt{T}$ to $1.96 * 1/\sqrt{T}$
- If the sample autocorrelation coefficient, $\hat{\tau_s}$, falls outside this region, for a given lag s, then the null hypothesis that the true value is zero (at lag s) is rejected

Autocorrelation Coefficient

- Box–Pierce (BP) test stat.: $Q = T \sum_{k=1}^{m} \tau_k^2$
- Ljung–Box (LB) statistic $Q^* = T(T+2) \sum_{k=1}^{m} \frac{\tau_k^2}{T-k}$
- These statistics are asymptotically distributed as χ_m^2
- What happens if T increases towards infinity

Autocorrelation Coefficient: Numerical Example

Autocorrelation Coefficient

- Consider a series of 100 observations. At lags, 1, 2, 3, 4, 5, you observe the following auto-correlation coefficients: $\tau_1 = 0.207; \tau_2 = -0.013;\ \tau_3 = 0.086;\ \tau_4 = 0.005, and\ \tau_5 = -0.022$. We will test these coefficients individually, using Box-Pierce (BP) test and jointly Ljung-Box (LB) test.
- (1) 95% confidence interval : =$\pm 1.96 * \frac{1}{\sqrt{T}}$
- (2) Using BP and LB test conduct the joint hypothesis

Autocorrelation Coefficient

- (1) 95% confidence interval
- For each coefficient, we can construct the confidence interval $=\pm 1.96 * \frac{1}{\sqrt{T}}$; here T=100. Thus, the confidence interval lies in between -0.196 to +0.196.
- Let us come to the joint test. The null hypothesis is that all the first five coefficients are jointly zero. That is, $H_0\text{:}\ \tau_1 = 0; \tau_2 = 0;\ \tau_3 = 0;\ \tau_4 = 0, and\ \tau_5 = 0$.

Autocorrelation Coefficient

- Let us come to the joint test. The null hypothesis is that all the first five coefficients are jointly zero. That is, H_0: $\tau_1 = 0; \tau_2 = 0;\ \tau_3 = 0;\ \tau_4 = 0, and\ \tau_5 = 0$. The relevant critical values are obtained from χ^2 distribution with five degrees of freedom. These are 11.1 at 5% level, 15.1 at 1% level.
- The test statistic for BP test is given below. $Q = T\sum_{k=1}^{m}\tau_k^2$
- The LB statistics is computed as follows. $T(T+2)\sum_{k=1}^{m}\frac{\tau_k^2}{T-k}$

Autocorrelation Coefficient

- Let us come to the joint test. The null hypothesis is that all the first five coefficients are jointly zero. That is, H_0: $\tau_1 = 0; \tau_2 = 0;\ \tau_3 = 0;\ \tau_4 = 0, and\ \tau_5 = 0$. The relevant critical values are obtained from χ^2 distribution with five degrees of freedom. These are 11.1 at 5% level, 15.1 at 1% level.

- The test statistic for BP test is given below.
  $BP = 100 * (0.207^2 + (-0.013)^2 + 0.086^2 + 0.005^2 + (-0.022)^2)$ = 5.09.

- The LB statistics is computed as follows. $LB = 100 * (100 + 2) * \left(\frac{0.207^2}{100-1} + \frac{(-0.013)^2}{100-2} + \frac{0.086^2}{100-3} + \frac{0.005^2}{100-4} + \frac{(-0.022)^2}{100-5}\right) = 5.26$

Autocorrelation Coefficient

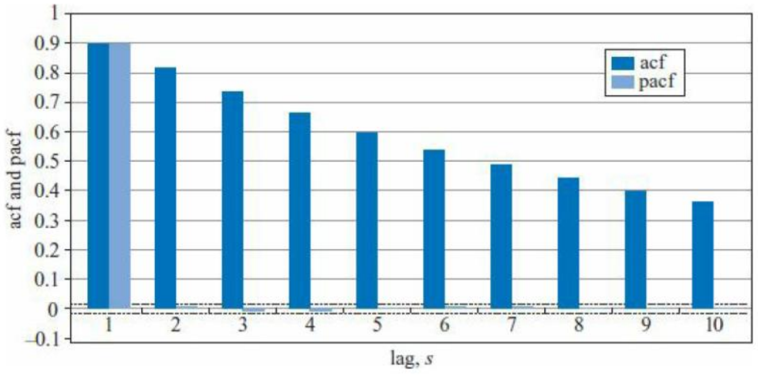
- So, in both cases, the joint null hypothesis that all of the first five autocorrelation coefficients are zero cannot be rejected. Note that, in this instance, the individual test caused a rejection while the joint test did not.
- Thus, the effect of the significant autocorrelation coefficient is diluted in the joint test by the insignificant coefficients.

ACF and PACF Plots: AR Process

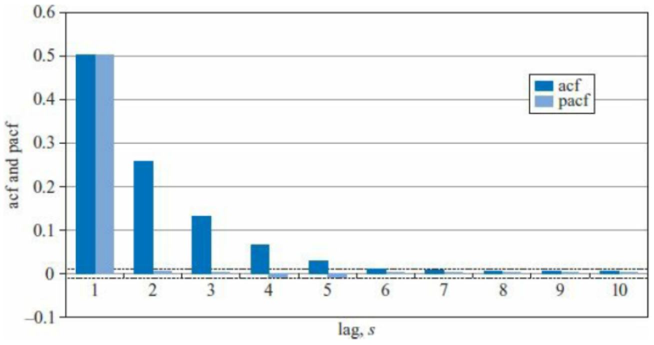
ACF and PACF Plots: AR Process

- Auto-correlation function (ACF) plot and partial ACF plot
- An AR process has
  - A geometrically decaying acf
  - A number of non-zero points of pacf = AR order

$$y_t = 0.9y_{t-1} + u_t$$



$$y_t = 0.5y_{t-1} + u_t$$

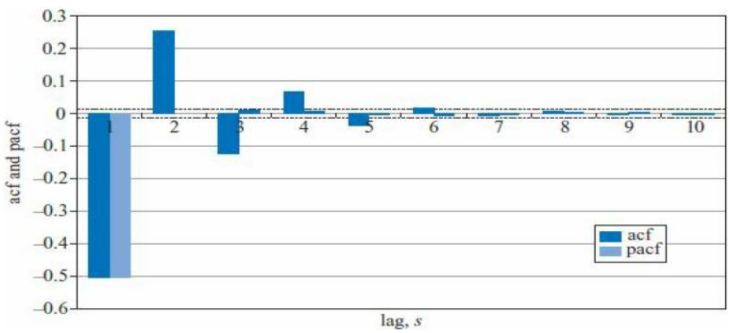


Chris Brooks; Introductory Econometrics for Finance. 4th Edition. Chapter 6

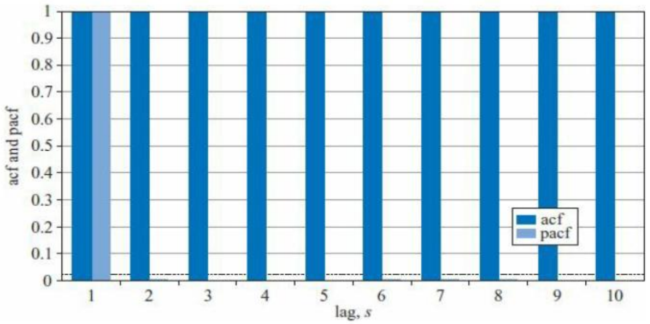
ACF and PACF Plots: AR Process

- Auto-correlation function (ACF) plot and partial ACF plot
- An AR process has
  - A geometrically decaying acf
  - A number of non-zero points of pacf = AR order

$$y_t = -0.5y_{t-1} + u_t$$



$$y_t = y_{t-1} + u_t$$



Chris Brooks; Introductory Econometrics for Finance. 4th Edition. Chapter 6

Moving Average (MA) Processes

Moving Average (MA) Processes

- Current value of the variable y_t depends upon the error terms
- An MA model of order q, MA(q) can be expressed as below
- $y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$
- $y_t = \mu + \sum_{i=1}^{q} \theta_i u_{t-i} + u_t$ or in terms of lag operator
- $y_t = \mu + \theta(L)u_t; where\ \theta(L) = \left(1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q\right)$
- Here u_t is the white noise IID error term [distributed as N (0, σ^2)]
- A moving average model is simply a linear combination of white noise processes

Moving Average (MA) Processes

- The following properties define this moving average [MA(q)] process
  - $y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$
- $E(y_t) = \mu$
- Var$(y_t) = \gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \theta_3^2 + \cdots + \theta_q^2)\,\sigma^2$
- Auto-Cov.=(γ_s)=$\begin{cases} (\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_2 + \cdots + \theta_q\theta_s)\sigma^2; for\ s = 1,2,3..q \\ 0 \quad for\ s > q \end{cases}$
- An MA (q) process has constant mean, constant variance, and auto-covariances that are non-zero up to lag 'q' and zero thereafter.

A Simple Example of MA(2) Process: Part I

Example MA (2) Processes: Variance

- The following properties define this moving average [MA(q)] process
  - $y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$
- $E(y_t) = E[u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}]$=$E(u_t) + \theta_1 E(u_{t-1}) + \theta_2 E(u_{t-2})$=0
- Var(y_t) $= E\big[[y_t - E(y_t)][y_t - E(y_t)]\big] = E(y_t^2)$
- $E(y_t^2) = E[(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})(u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2})]$
- $E(y_t^2) = E[(u_t^2 + \theta_1^2 u_{t-1}^2 + \theta_2^2 u_{t-2}^2 + Cross\ products]$
- $E(Cross - product) =?\ [Hint: E[u_t * u_{t-s}] = Cov(u_t, u_{t-s}) =?$ For s≠0
- For s=0; $E[u_t * u_t]$=$Cov(u_t, u_t)$=$Var(u_t, u_t) = \sigma^2$

Example MA (2) Processes: Variance

- The following properties define this moving average [MA(q)] process
  - $y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$
- $E(y_t) = E[u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}] = E(u_t) + \theta_1 E(u_{t-1}) + \theta_2 E(u_{t-2}) = 0$
- Var$(y_t) = E\big[[y_t - E(y_t)][y_t - E(y_t)]\big]$
- $E(y_t^2) = E[(u_t^2 + \theta_1^2 u_{t-1}^2 + \theta_2^2 u_{t-2}^2] = (1 + \theta_1^2 + \theta_2^2)\ \sigma^2$

A Simple Example of MA(2) Process: Part 2

Example MA (2) Processes: Auto-Covariance

- The following properties define this moving average [MA(q)] process
  - $u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}$
- $\gamma_1 = E\big[[y_t - E(y_t)][y_{t-1} - E(y_{t-1})]\big] = E[y_t * y_{t-1}]$
- $\gamma_1 = E\big[[u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2}][u_{t-1} + \theta_1 u_{t-2} + \theta_2 u_{t-3}]\big]$
- $\gamma_1 = E\left[[\theta_1 u_{t-1}^2 + \theta_1 \theta_2 u_{t-2}^2] + Cross\ Product\ terms\right]$
- $\gamma_1 = (\theta_1 + \theta_1 \theta_2)\ \sigma^2$
- $\gamma_2 = (\theta_2)\ \sigma^2$
- $\gamma_3 = 0$

Example MA (2) Processes: Auto-Correlation

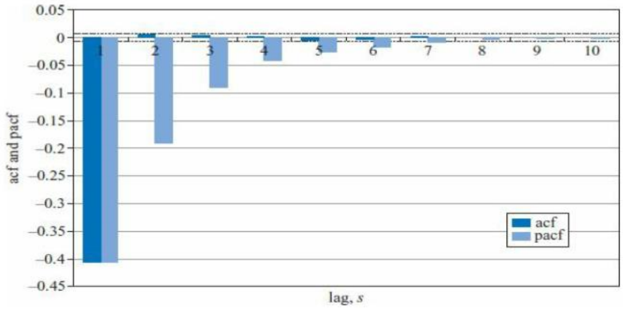
- $\tau_0 = \frac{\gamma_0}{\gamma_0} = 1$
- $\tau_1 = \frac{\gamma_1}{\gamma_0} = \frac{(\theta_1 + \theta_1\theta_2)\ \sigma^2}{(1+\theta_1^2+\theta_2^2)\ \sigma^2} = \frac{(\theta_1+\theta_1\theta_2)}{(1+\theta_1^2+\theta_2^2)}$
- $\tau_2 = \frac{\gamma_s}{\gamma_0} = \frac{(\theta_2)}{(1+\theta_1^2+\theta_2^2)}$

ACF and PACF Plots: MA Process

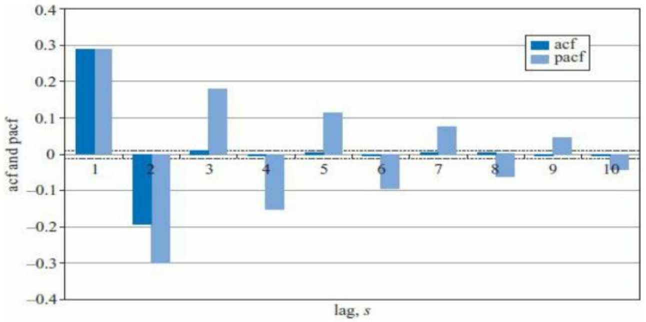
ACF and PACF Plots: MA Process

- Auto-correlation function (ACF) plot and partial ACF plot
- An MA process has
  - Number of non-zero points of acf = MA order
  - A geometrically decaying pacf.

$$y_t = -0.5u_{t-1} + u_t$$



$$y_t = 0.5u_{t-1} - 0.25u_{t-2} + u_t$$



Chris Brooks; Introductory Econometrics for Finance. 4th Edition. Chapter 6

ARMA Process

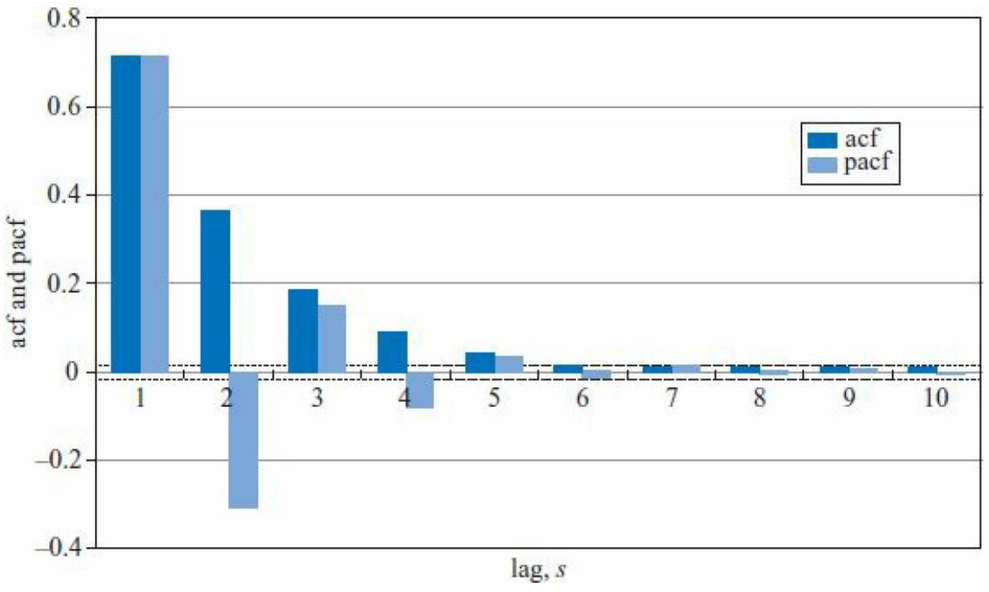
ARMA Process

- By combining the AR(p)and MA(q)models, an ARMA(p, q)model is obtained
- $y_t = \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$
- $\varphi(L) y_t = \mu + \theta(L) u_t$
- Where $\theta(L) = \left(1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q\right)$
- $\varphi(L) = \left(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p\right)$
- Also, $E[u_t] = 0; E[u_t^2] = \sigma^2; E[u_t u_s] = 0; t \neq s$

ACF and PACF Plots: ARMA Process

- Auto-correlation function (ACF) plot and partial ACF plot
- An ARMA process has
  - A geometrically decaying acf
  - A geometrically decaying pacf

$$y_t = 0.5y_{t-1} + 0.5u_{t-1} + u_t$$



Building ARMA Model

Building the ARMA Model for Price Prediction

- Identification: Determine the order of the process, that is, p and q values for ARMA (p,q)
- Estimation: Estimate the parameters with OLS/MLE
- Model Diagnostics: Testing the model. Residual diagnostics: Clean iid residuals
- Objective: A parsimonious model removes irrelevant lags of AR and MA terms.
- Information criteria for model/lag selection

Information Criteria for ARMA Model Selection

- Akaike's (1974) information criterion (AIC)=$-2\,log[L] + 2K$
- Schwarz's (1978) Bayesian information criterion (SBIC)=$-2\,log[L] + K * log(T)$
- Hannan–Quinn criterion (HQIC)=$-2\,log[L] + 2Klog(log(T))$
- Where $K = \frac{k}{T}$ = k is number of parameters, T is the sample size, and - log(L) is log-likelihood of observing the parameters obtained from the model

Forecasting with time-series models

Time-series forecasting

- Time-series models use conditional expectations $E(y_{t+1}|\Omega_t)$: expected value of ‘y’ at ‘t+1’ given all the information up to ‘t’ (Ω_t)
- For a zero mean white noise process $E(\mu_{t+1}|\Omega_t) = 0 \; \forall \; s > 0$
- Naïve Forecasting: $E(y_{t+1}|\Omega_t)$= y_t; or random walk process (no change forecast
- The unconditional expectation is the unconditional mean of ‘y’ with out any time reference (long-term mean)
- For mean-reverting stationary process, the long-term average becomes the forecast

Problems with structural models

- $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_k x_{kt} + u_t$: conditional forecasts
- $E(y_t|\Omega_{t-1}) = \beta_1 + \beta_2 E(x_{2t}) + \beta_3 E(x_{3t}) + \cdots + \beta_k E(x_{kt})$
- To forecasts conditional expectations for y, one needs forecasts of x's, i.e., $E(x_{kt})$
- This makes the process cumbersome and complex
- One may have to look the historical values of x's to forecast the current values; however, this can be directly captured in the historical time-series values of y_t itself

Forecasting with ARMA Models: Part I

Forecasting with ARMA models

- Forecast from ARMA(p,q) model at time 't' for 's' steps into the future is given as
- $f_{t+s} = \Sigma_{i=1}^{p} a_i f_{t+s-i} + \Sigma_{k=1}^{q} b_k u_{t+s-k}$
- Here, a_i and b_k are the autoregressive and moving average coefficients

Forecasting MA(q)

- Let us look at MA(3) model: $y_t = \mu + \theta_1\mu_{t-1} + \theta_2\mu_{t-2} + \theta_3\mu_{t-3} + \mu_t$
- Assuming parameter constancy (i.e., the relationship holds), then
- $y_{t+1} = \mu + \theta_1\mu_t + \ \theta_2\mu_{t-1} + \ \theta_3\mu_{t-2} + \mu_{t+1}$
- $f_{t,1} = E(y_{t+1}|\Omega_t) = E(\mu + \theta_1\mu_t + \ \theta_2\mu_{t-1} + \ \theta_3\mu_{t-2} + \mu_{t+1}|\Omega_t)$
- The values of error terms up to time ‘t’ is known, but after that we have to take their conditional expectation, which is zero
- That is $E(\mu_{t+1}|\Omega_t)$=0

Forecasting MA(q)

- That is $E(\mu_{t+1}|\Omega_t)$=0; therefore
- $f_{t,1} = E(y_{t+1}|\Omega_t) = \mu + \theta_1\mu_t + \theta_2\mu_{t-1} + \theta_3\mu_{t-2}$
- $f_{t,2} = E(y_{t+2}|\Omega_t) = E(\mu + \theta_1\mu_{t+1} + \theta_2\mu_t + \theta_3\mu_{t-1}|\Omega_t) = u + \theta_2\mu_t + \theta_3\mu_{t-1}$
- $f_{t,3} = E(y_{t+3}|\Omega_t) = E(\mu + \theta_1\mu_{t+2} + \theta_2\mu_{t+1} + \theta_3\mu_t|\Omega_t) = u + \theta_3\mu_t$
- $f_{t,4} = E(y_{t+4}|\Omega_t) = E(\mu + \theta_1\mu_{t+3} + \theta_2\mu_{t+2} + \theta_3\mu_{t+1}|\Omega_t) = u$

Forecasting MA(q)

- Since the MA(3) process has a memory of only three periods, any forecast of four or more steps ahead converge to long-term unconditional mean (i.e., the intercept term)

Forecasting with ARMA Models: Part II

Forecasting AR(p) process

- Consider AR(2) process
- $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$; unlike MA process, AR process has infinite memory
- The 't+1' forecast is obtained as:
- $y_{t+1} = \mu + \phi_1 y_t + \phi_2 y_{t-1} + u_{t+1}$, then
- $f_{t,1} = E(y_{t+1}|\Omega_t) = E(\mu + \phi_1 y_t + \phi_2 y_{t-1} + u_{t+1}|\Omega_t) = \mu + \phi_1 y_t + \phi_2 y_{t-1}$ (since actual values of y_t and y_{t-1} are observed)

Forecasting AR(p) process

- $f_{t,1} = E(y_{t+1}|\Omega_t) = E(\mu + \phi_1 y_t + \phi_2 y_{t-1} + u_{t+1}|\Omega_t)$=$\mu + \phi_1 y_t + \phi_2 y_{t-1}$ (since actual values of y_t and y_{t-1} are observed)
- Similarly, for next steps 2 and 3
- $f_{t,2} = E(y_{t+2}|\Omega_t) = E(\mu + \phi_1 y_{t+1} + \phi_2 y_t + u_{t+2}|\Omega_t)$=$\mu + \phi_1 f_{t,1} + \phi_2 y_t$
- $f_{t,3} = E(y_{t+3}|\Omega_t) = E(\mu + \phi_1 y_{t+2} + \phi_2 y_{t+1} + u_{t+3}|\Omega_t)$=$\mu + \phi_1 f_{t,2} + \phi_2 f_{t,1}$

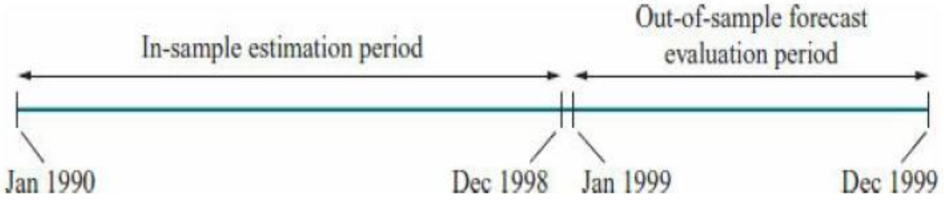
Forecasting AR(p) process

- Hence, the generic 's' step ahead forecast becomes
- $f_{t,s} = \mu + \phi_1 f_{t,s-1} + \phi_2 f_{t,s-2}$
- These steps can be used to generate ARMA(p,q) order forecast

Forecasting with ARMA Models: Part III

One-Step-Ahead vs. Multi-Step-Ahead Forecasts and Rolling vs. Recursive Samples

Objective: to produce	Data used to estimate model parameters	
1-, 2-, 3-step-ahead forecasts for:	Rolling window	Recursive window
1999*M*1, *M*2, *M*3	1990*M*1–1998*M*12	1990*M*1–1998*M*12
1999*M*2, *M*3, *M*4	1990*M*2–1999*M*1	1990*M*1–1999*M*1
1999*M*3, *M*4, *M*5	1990*M*3–1999*M*2	1990*M*1–1999*M*2
1999*M*4, *M*5, *M*6	1990*M*4–1999*M*3	1990*M*1–1999*M*3
1999*M*5, *M*6, *M*7	1990*M*5–1999*M*4	1990*M*1–1999*M*4
1999*M*6, *M*7, *M*8	1990*M*6–1999*M*5	1990*M*1–1999*M*5
1999*M*7, *M*8, *M*9	1990*M*7–1999*M*6	1990*M*1–1999*M*6
1999*M*8, *M*9, *M*10	1990*M*8–1999*M*7	1990*M*1–1999*M*7
1999*M*9, *M*10, *M*11	1990*M*9–1999*M*8	1990*M*1–1999*M*8
1999*M*10, *M*11, *M*12	1990*M*10–1999*M*9	1990*M*1–1999*M*9



Chris Brooks. *Introductory Econometrics for Finance*. 4th Edition. Chapter 6

Determining the Accuracy of Forecast

Determining the Accuracy of Forecast

- Conditional expectations: The expression $E(y_{t+1}|\Omega_t)$ states the expected value of y at 't+1' conditional upon information available up to time t, i.e., Ω_t.
- Naïve forecasting: Forecast or expectation of y, s steps into future is the current value of y, i.e., $E(y_{t+1}|\Omega_t) = y_t$. Then the process follows random walk. For a mean reverting series some long-term unconditional average is the best forecast for the series in future.

Forecasting with Time-Series Models

Steps ahead	Forecast (F)	Actual (A)	Squared error $(F - A)$^2	Absolute error |F-A|
1	0.2000	-0.4000	?	?
2	0.1500	0.2000	?	?
3	0.1000	0.1000	?	?
4	0.0600	-0.1000	?	?
5	0.0400	-0.0500	?	?

- Mean Squared Error (MSE)=?
- Mean Absolute Error (MAE)=?
- Root Mean Square Error (RMSE)=?

Forecasting with Time-Series Models

Steps ahead	Forecast	Actual	Squared error	Absolute error
1	0.2000	-0.4000	0.3600	0.6000
2	0.1500	0.2000	0.0025	0.0500
3	0.1000	0.1000	0.0000	0.0000
4	0.0600	-0.1000	0.0256	0.1600
5	0.0400	-0.0500	0.0081	0.0900

- MSE=(0.3600+0.0025+0.0000+0.0256+0.0081)/5=0.08
- MAE=(0.6000+0.0500+0.0000+0.1600+0.0900)/5=0.16
- RMSE=SQRT(MSE)=sqrt(0.08)=0.28

Time-Series Stationarity: Recap

Time-Series and Stationarity

- A strictly stationary process
  - A time series process {yt} is said to be strictly stationary if the joint probability distribution of a sequence {yt1 to ytn} is the same as that for {yt1+m to ytn+m} ∀ m: i.e.,

$$F\{y_{t_1}, \ldots, y_{t_n}\} = F\{y_{t_{1+m}}, \ldots, y_{t_{n+m}}\}$$

- A weakly stationary process
  - A weakly stationary (or covariance stationary) process satisfies the following three equations
  - $E(y_t) = u, t = 1,2,3 \ldots \infty$
  - $E(y_t - u)(y_t - u) = \sigma^2 < \infty$
  - $E(y_{t_1} - u)(y_{t_2} - u) = \gamma_{t_2 - t_1}\ for\ all\ t_1\ and\ t_2$

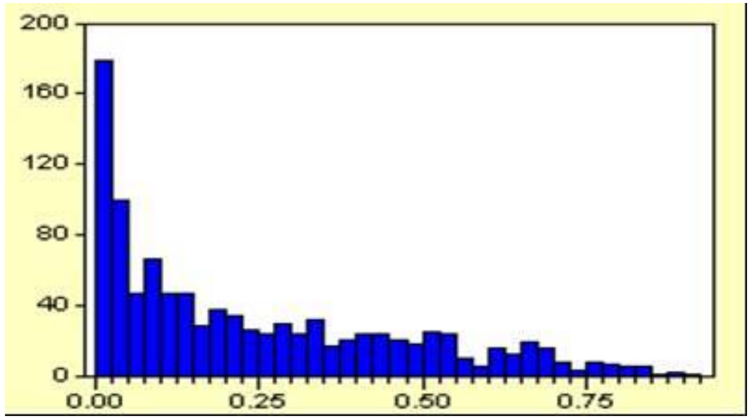
Time-Series and Stationarity

- A weakly stationary process
  - A weakly stationary (or covariance stationary) process satisfies the following three equations:
    - $E(y_t) = u, t = 1,2,3 \ldots \infty$
    - $E(y_t - u)(y_t - u) = \sigma^2 < \infty$
    - $E\left(y_{t_1} - u\right)\left(y_{t_2} - u\right) = \gamma_{t_2 - t_1}\ for\ all\ t_1\ and\ t_2$
  - To summarize, a weak stationary process should have a constant mean, constant variance, and a constant auto-covariance structure.
  - What is auto-covariance: $E(y_t - u)(y_{t+s} - u) = \gamma_s;\ s = 0,1,2, \ldots$
  - Autocorrelation: $\tau_s = \frac{\gamma_s}{\gamma_0},$ s = 0,1,2, ... ; τ_s lies in the interval ± 1

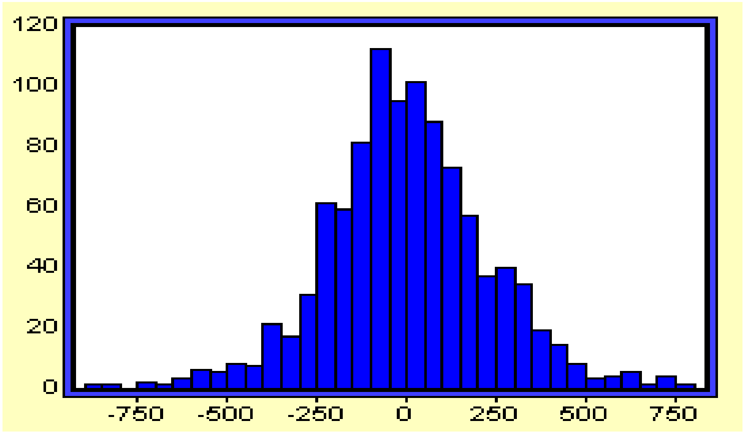
Time-Series and Stationarity

- Problems with non-stationary series
  - Shocks to non-stationary systems do not die away and are highly persistent
  - Estimation with non-stationary data can lead to spurious regressions, i.e., artificially high t-stat. and adjusted R-sq, even if the variables are not related to each other. t-ratio will not follow t-distribution

Adj. R sq



T-stat



Chris Brooks. *Introductory Econometrics for Finance*. 4th Edition. Chapter 6

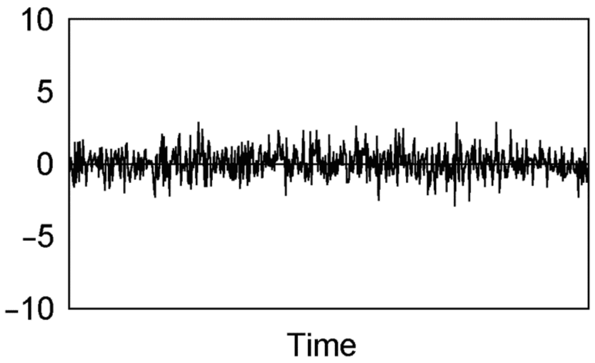
Types of Non-Stationarity

Types of Non-Stationarity

- Two types of non-stationarity
  - In financial econometrics, two kinds of non-stationarity become important

    1. Random walk model with drift: $y_t = \mu + y_{t-1} + u_t$ (1)
    2. Deterministic trend process: $y_t = \alpha + \beta t + u_t$ (2)

  - Here u_t is a white noise iid process distributed as $N(0, \sigma^2)$



Recall: White Noise IID Process

- It can be defined as follows

1. E(yt) = 0 , t = 1,2,...,∞
2. Var(yt)= σ^2
3. $E\left(y_{t_1} - u\right)\left(y_{t_2} - u\right) = \gamma_{t\,2-t_1} = 0\ for\ all\ t_1\ and\ t_2$ for all t1 ≠ t2

- Thus, white noise has a constant (zero) mean, finite constant variance, and is uncorrelated across time.
- Often, we assume a fix distribution (e.g., Normal), then it is a special case of IID process.

Stochastic Non-Stationarity

Stochastic Non-Stationarity

- A generalized model of the autoregressive process with drift.
  - $y_t = \mu + \phi y_{t-1} + u_{t;}$ this is an explosive process if $\phi > 1$
- This does not describe real economic data because shocks to the system persist and explode.
- So, the only non-stationarity case considered in that of $\phi = 1$.

Stochastic Non-Stationarity

- Consider the general case of an AR(1) with no drift: $y_t = \phi y_{t-1} + \mu_t$
  - $y_{t-1} = \phi y_{t-2} + \mu_{t-1}$ and $y_{t-2} = \phi y_{t-3} + \mu_{t-2}$
  - $y_t = \phi(\phi y_{t-2} + \mu_{t-1}) + \mu_t = \phi^2 y_{t-2} + \phi \mu_{t-1} + \mu_t$
- T successive substitutions $y_t = \phi^T y_0 + \phi \mu_{t-1} + \phi^2 \mu_{t-2} + \phi^3 \mu_{t-3} + \cdots + \phi^T \mu_0 + \mu_t$
  - $\phi<1 \Rightarrow \phi^T \rightarrow 0$ as $T \rightarrow \infty$: Shocks to system die away
  - $\phi=1 \Rightarrow \phi^T = 1 \forall$ T: Shocks persist in the system and never die away: $y_t = y_0 + \sum_{t=0}^{\infty} u_t$ as $T \rightarrow \infty$

Stochastic Non-Stationarity (NS)

- $y_t = \mathrm{u} + y_{t-1} + \mu_t$:Random walk non-stationary (NS) process with drift
- NS can be removed using first differences
- $\Delta y_t = y_t - y_{t-1} = y_t - L\ y_t = (1\text{-}L)\ y_t = \mathrm{u} + \mu_t$
- First differencing has introduced stationarity, so the process is called integrated to the order of one I(1)
- Here (1-L) is called the characteristic equation of the process. Since the equation has a root/solution of 1, it is also called unit root process.

Stochastic Non-Stationarity

- First differencing has introduced stationarity, so the process is called integrated to the order of one: I(1).
- If a non-stationary series, y_t must be differenced d times before it becomes stationary, then it is said to be integrated of order d: I(d).
- This would be written y_t ~ I(d). So if y_t ~ I(d) then $\Delta^d\, y_t$ ~ I(0).
- Applying the difference operator, Δ, d times, leads to an I(0) process, i.e., a process with no unit roots.
- I(1) and I(2) series can wander a long way from their mean value and cross this mean value rarely, while I(0) series should cross the mean frequently.

Deterministic Non-Stationarity

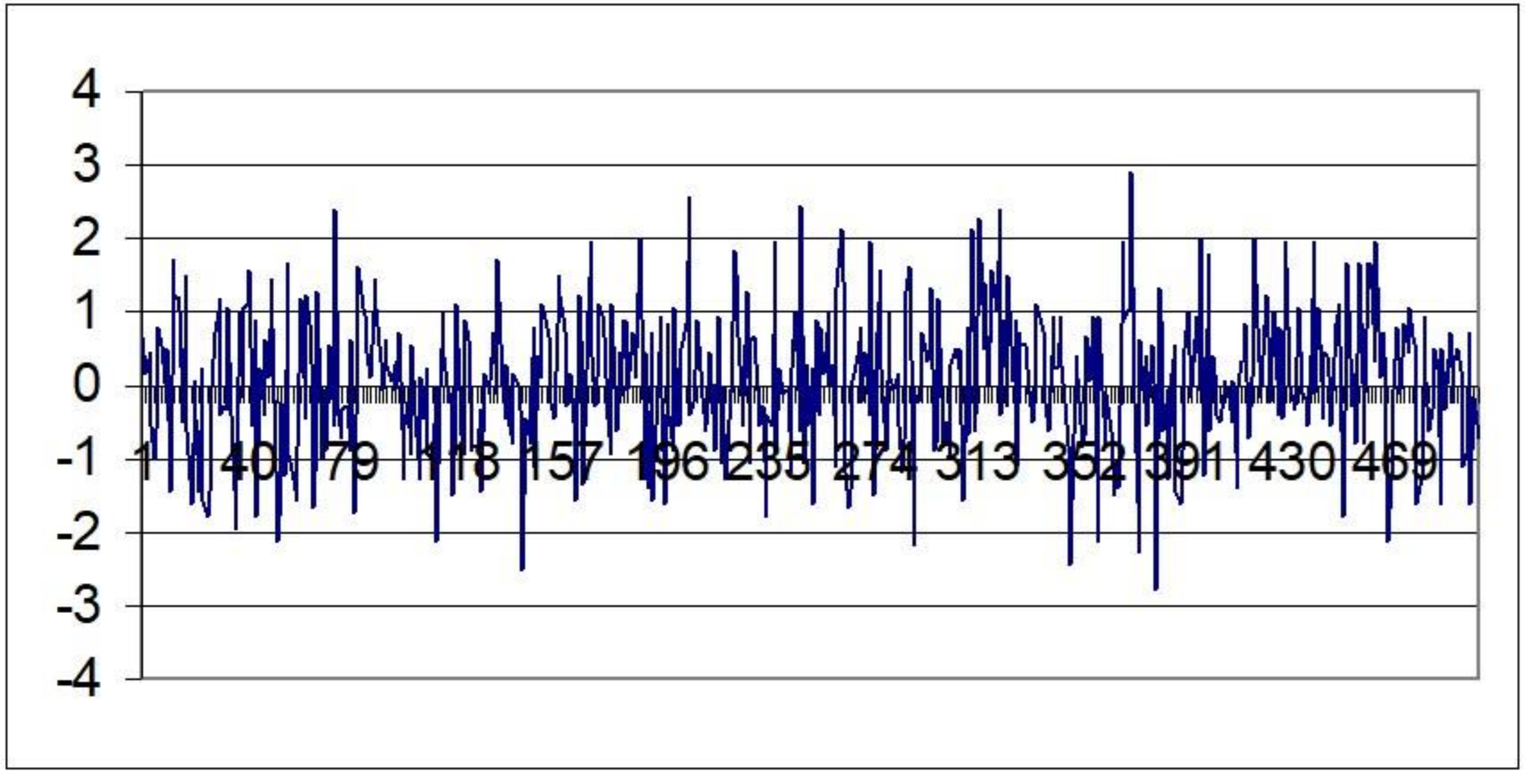
Deterministic Non-Stationarity

- Deterministic trend process: $y_t = \alpha + \beta t + u_t$
- This is a rather simple case of non-stationarity
- Only de-trending is required, and a time ‘t’ term is added to the model to remove the trend
- Thus, residuals obtained from the series, u_t ‘s have their trend removed
- Thus, any subsequent estimation can be done on these error terms

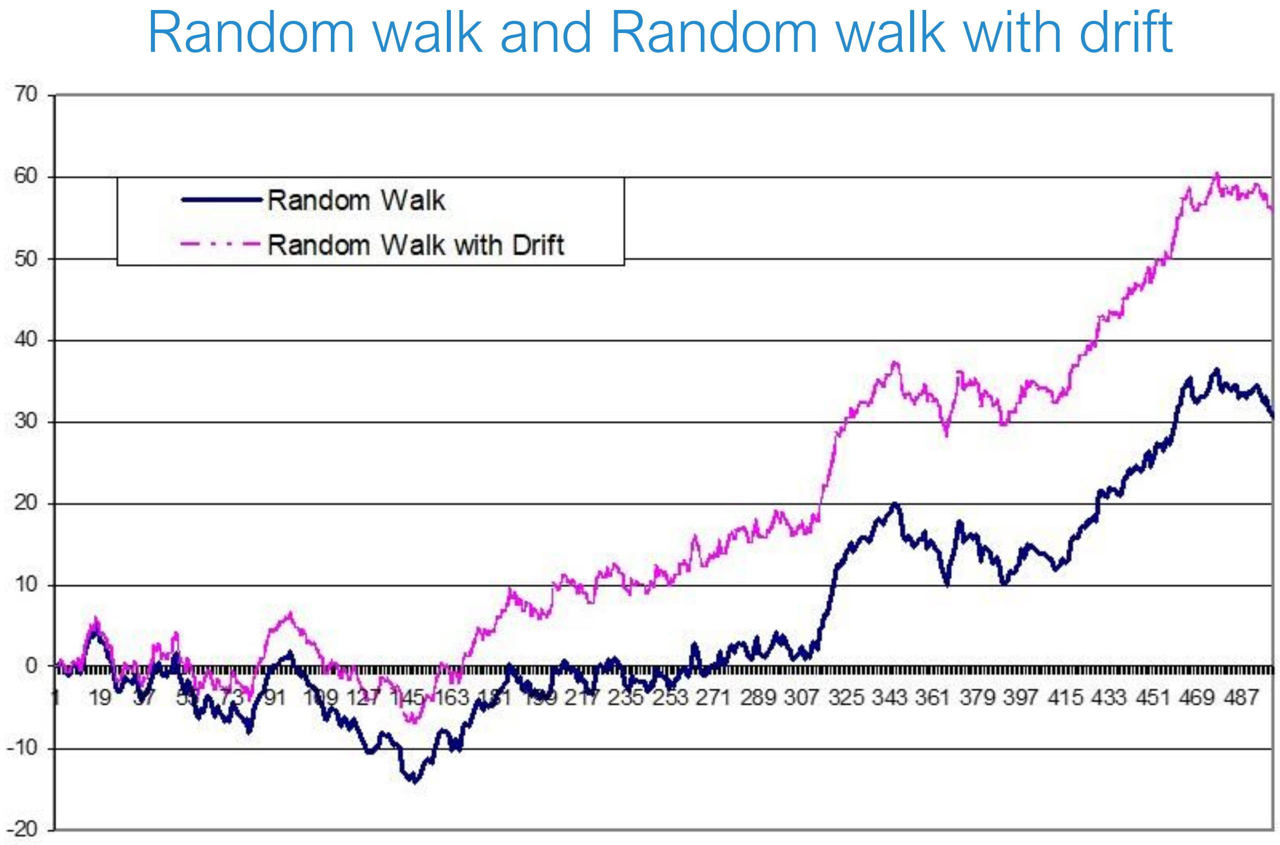
Time-Series and Stationarity: Visual Examination

White Noise IID Process with Zero Mean

White noise IID process with zero mean

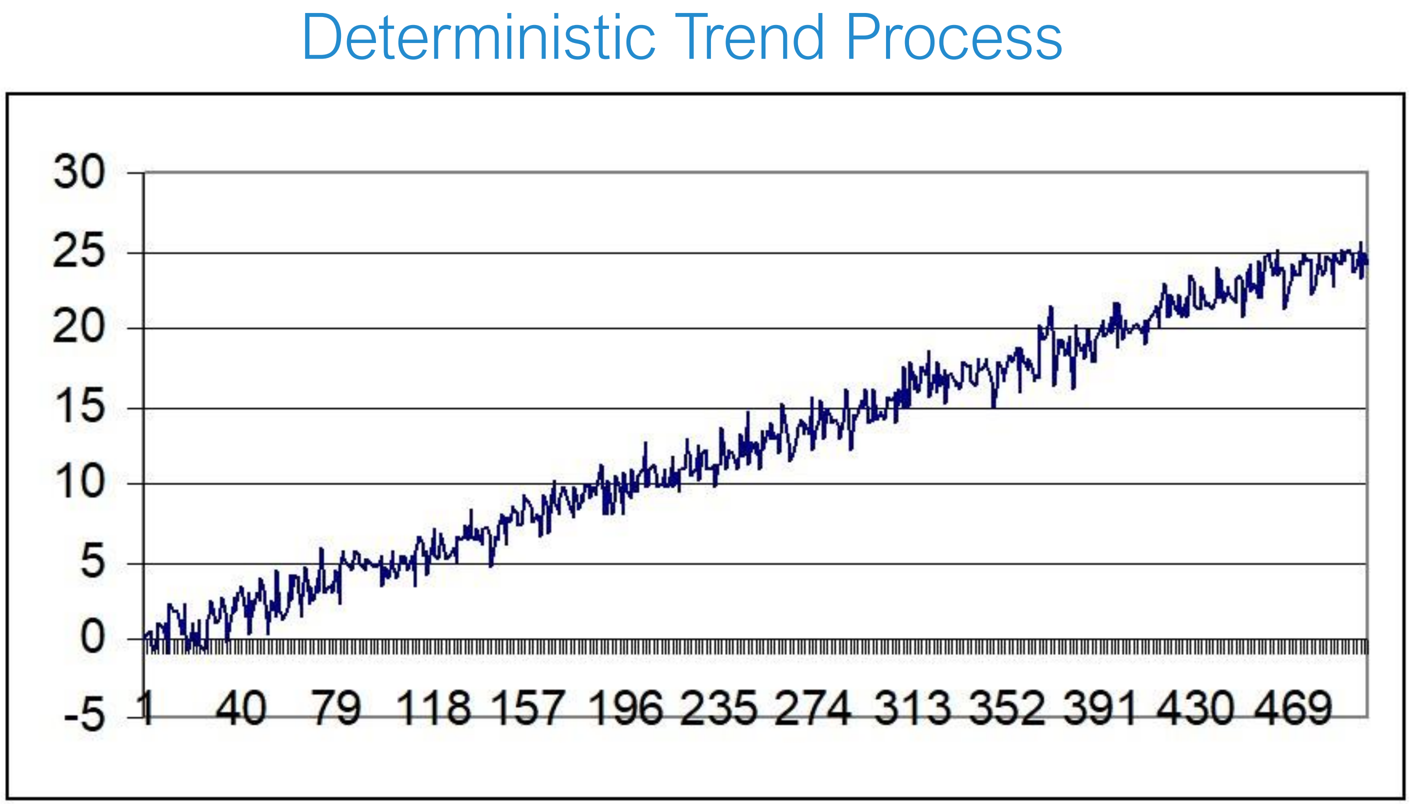


Random Walk and Random Walk with Drift



Chris Brooks. *Introductory Econometrics for Finance*. 4th Edition. Chapter 6

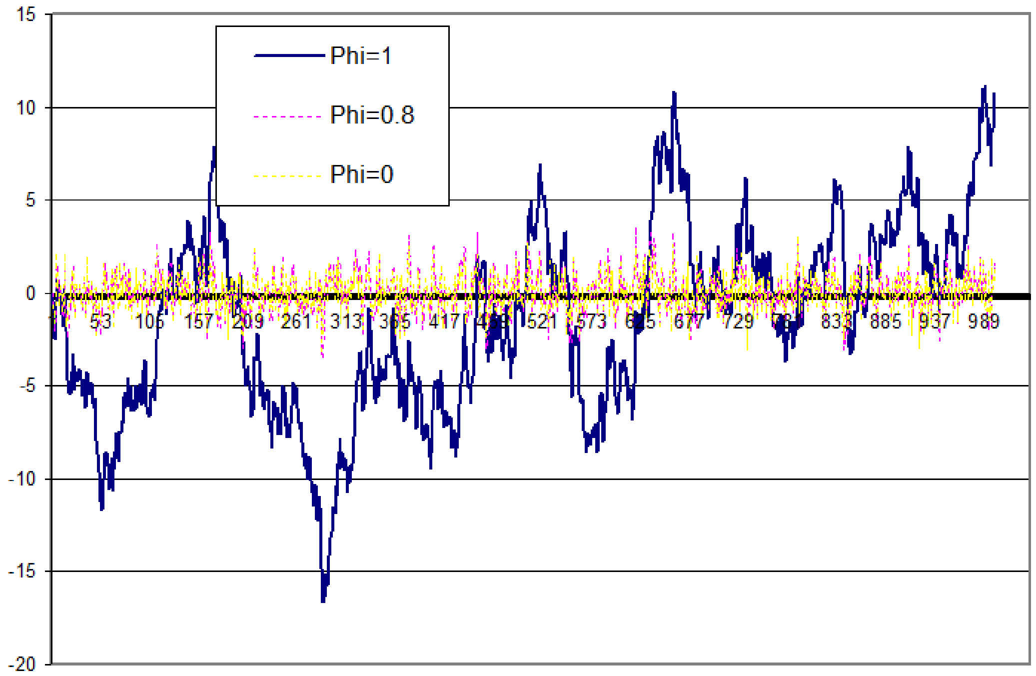
Deterministic Trend Process



Chris Brooks. *Introductory Econometrics for Finance*. 4th Edition. Chapter 6

Autoregressive Processes with Differing Values of (0, 0.8, 1)

Autoregressive Processes with Differing Values of ϕ (0, 0.8, 1)



Time-Series and Stationarity: Visual Inspection

- A white noise process visibly has no trending behavior, and it frequently crosses its mean value of zero.
- The random walk (thick line) and random walk with drift (faint line) process exhibit ‘long swings’ away from their mean value, which they rarely cross.
- A comparison of the two lines in this graph reveals that the positive drift leads to a series that is more likely to rise over time than to fall; obviously, the effect of the drift on the series becomes greater and greater the further the two processes are tracked.
- Finally, the deterministic trend process clearly does not have a constant mean and exhibits completely random fluctuations about its upward trend. If the trend were removed from the series, a plot similar to the white noise process would result.

Testing for Unit-Roots: Part I

Testing for a Unit Root : Simple Dicky–Fuller Test

- Consider the model: $y_t = \phi y_{t-1} + u_t$
- The basic objective of the test is to test the null hypothesis that $\phi = 1$, against the one-sided alternative $\phi < 1$.
- So, we have: H0: series contains a unit root ; vs. H1: series is stationary.
- We usually use the regression: $\Delta y_t = \psi y_{t-1} + u_t$ so that a test of $\phi=1$ is equivalent to a test of $\psi=0$ (since $\phi-1=\psi$).
- Why not original regression.

Testing for a Unit Root: Simple Dicky–Fuller Test

- The model can be generalized by allowing for a drift (intercept) and deterministic trend or neither.
- Consider the model: $y_t = \phi y_{t-1} + \mu + \lambda t + u_t$
- First Differencing the series: $\Delta y_t = \psi y_{t-1} + \mu + \lambda t + u_{t,}$
  Where $\Delta y_t = y_t - y_{t-1}$
- Then H_0: Series contains a unit root $\phi = 1$ and H_1: $\phi < 1$
- This is a test for a random walk against a stationary AR(1) with drift and a time trend.

Testing for a Unit Root: Simple Dicky–Fuller Test

- Dickey Fuller (DF) tests are also known as τ tests: τ, τ_{μ}, τ_{τ}
- This test statistic is computed as follows: Tau=$\frac{\widehat{\psi}}{SE(\widehat{\psi})}$
- The test statistic does not follow the usual t-distribution under the null since the null is one of non-stationarity but rather follows a non-standard distribution.
- Tau values are larger (negative) than regular t-stat. indicating more evidence required against null.

Testing for Unit-Roots: Part II

Testing for a Unit Root: Augmented Dicky–Fuller (ADF) Test

- The null hypothesis of a unit root is rejected in favor of the stationary alternative in each case if the test statistic is more negative than the critical value.
- Also, this test requires u_t to be white noise IID.
- u_t will be autocorrelated if there was autocorrelation in the dependent variable of the regression (Δy_t).
- The solution is to "augment" the test using p lags of the dependent variable.

Testing for a Unit Root: Augmented Dicky–Fuller (ADF) Test

- The solution is to “augment” the test using p lags of the dependent variable

$$\Delta y_t = \psi y_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + u_t$$

- This is referred to as the augmented DF test, as the lags absorb any autocorrelation in the structure
- Again, the optimum lag selection can be done through information criteria [What is the problem of too few and too many lags?].
- There are other tests Phillips Perron (PP) and KPSS test

Testing for Unit-Roots: Part III

Summary

- ADF/PP: $\mathrm{H_0: y_t \sim I(1), i.e, nonstationarity}$ and $\mathrm{H_1: y_t \sim I(0), i.e, stationarity}$
- KPSS: $\mathrm{H_0: y_t \sim I(0), i.e., stationarity}$ and $\mathrm{H_1: y_t \sim I(i), i.e, nonstationarity}$

Outcome	ADF/PP	KPSS
1	Reject $\boldsymbol{H_0}$	Do not Reject $\boldsymbol{H_0}$
2	Do not Reject $\boldsymbol{H_0}$	Reject $\boldsymbol{H_0}$
3	Reject $\boldsymbol{H_0}$	Reject $\boldsymbol{H_0}$
4	Do not Reject $\boldsymbol{H_0}$	Do not Reject $\boldsymbol{H_0}$

Introduction to Cointegration and Mean Reversion

Introduction

- A price series or a combination of two-series returns back to its long-run mean value
- This mean value is driven by some fundamental factor (cash/future relation)
- This leads to a property called stationarity
- Price series are rarely stationary; it is the returns that are most often of stationary

Cointegration

- Most of the time if two variables are I(1), then their combination (error term!) would also be I(1)
- Similarly, if two or more variables with different orders are combined, then the combination will have the integration order equal I(d) to the largest order variable I(d)
- Consider a simple regression model: $y_1 = \alpha_0 + \alpha_1 * x_1 + \alpha_2 * x_2 + u_t$ or alternatively
- $y_1 - \alpha_0 - \alpha_1 * x_1 - \alpha_2 * x_2 = u_t$: here the error term is a combination of three variables y1, x1, and x2.

Cointegration

- If these variables are I(1), chances are that error term would also be I(1).
- It would be desirable to have an error term as I(0): This is called cointegration.
- But what would be these economic and financial scenarios where an error would be I(0).
- A set of variables are called cointegrated if a linear combination is stationary.
- Many time series (spot vs. futures, exchange rates) in finance and economics move together.
- While in the short-run, they can go apart, in the long-run fundamental forces of the market bring them together.
- It is a long-term equilibrium phenomenon.

Error Correction Models (ECM)

Error Correction Models (ECM)

- For univariate series, the approach to deal with non-stationarity is to use first differences and then use the I(0) series for further analysis.
- However, in this manner, we lose the original long-term relationship between the variables at I(1) [Why we are only discussing I(1)].
- For example, if two variables $y_t \ and \ x_t$ are I(1), then one may consider estimating the following relation $\Delta y_t = \alpha + \beta \Delta x_t + u_t$ [Think of doing this for cash and futures data].

Error Correction Models (ECM)

- Often long-run is defined as a steady state equilibrium when values have converged to steady and are not changing, i.e., $y_t = y_{t-1} = y \text{ } and \text{ } x_t = x_{t-1} = x$
- Thus, Δy_t and Δx_t are not changing and close to 0. So, estimating equation above is meaningless
- Then the model $\Delta y_t = \alpha + \beta \Delta x_t + u_t$ has no solution
- However, there is a solution to this problem if the variables are cointegrated.

Error Correction Models (ECM)

- Consider the following class of model, that employs lagged cointegrated variables as well as their first differences. $[y_{t-1} - \Upsilon x_{t-1} = e_t]$
- $\Delta y_t = \alpha + \beta_1 \Delta x_t + \beta_2 (y_{t-1} - \Upsilon x_{t-1}) + u_t;$
- This is called error correction model (ECM) or vector error correction model (VECM).
- $[y_{t-1} - \Upsilon x_{t-1}]$ is the error correction term, provided that y_t *and* x_t *are cointegrated*, this term would be I(0) even though y_t *and* x_t are I(1) [May an intercept also]
- So, it is perfectly valid to use the OLS procedure here

Error Correction Models (ECM)

- Of course, the error correction term appears with lags (t-1); otherwise, it would indicate that changes in y from t-1 to t are driven by disequilibrium in the long-term relation at t.
- Gamma (Υ) coefficient here defines the long-run relationship between x and y.
- β_1 describes the short-run relationship (first differences), and β_2 describes the speed of adjustment toward equilibrium.
- This discussion can be extended to a system of more than two variables.

Engle-Granger Approach to ECM and Cointegration

Engle-Granger Approach

- Consider the following equilibrium model of k cointegrated variables
- $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \cdots + \beta_k x_{kt} + u_t$
- If the variables y_t's and x_t's are cointegrated, then u_t can be I(0)
- One can set-up tests of stationarity on $\mathrm{u_t}$ to confirm this

Error Correction Models (ECM)

Two-step Engle-Granger approach to parameter estimation:

Step 1

- Use the I(1) variables to estimate the cointegrating model and extract the residuals u_t
- Test these errors for I(0) stationarity

Step 2

- If the residuals are stationary, then estimate the following VECM model
- $\Delta y_t = \alpha_0 + \alpha_1 \Delta x_t + \alpha_2 (u_t) + v_t$
- If the relationship is $y_t - \beta_2 x_{2t} - \beta_3 x_{3t}$, then $[1 - \beta_2 - \beta_3]$ is the cointegrating vector
- All the problems related to OLS discussed earlier are also applicable here

Thanks!