# FRM Notes

Michal Mackanic

2015

# Contents

# Chapter 1

# Notes

## 1.1 Statistics

- binominal density function can be approximated by Poisson density function when $np = \lambda$ is large, i.e. when $n \to \infty$ and $p \to 1$
- log-normal distribution is skewed to the right (e.g. positively skewed) with

$$f(x) = e^{\frac{1}{2}\sigma^2-\mu}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\ln x-(\mu-\sigma^2)}{\sigma}\right)^2}$$
$$E[X] = e^{\mu+\frac{1}{2}\sigma^2}$$
$$D[X] = (e^{\sigma^2}-1)e^{2\mu+\sigma^2}$$

- equatility of two population variances - assumes that the two populations are normally distributed; always construct F-statistics such that $s_1^2 > s_2^2$

$$F_{n_1-1,n_2-1} = \frac{s_1^2}{s_2^2}$$

- $A = \beta_1^A EQ + \beta_2^A BND$ and $B = \beta_1^B EQ + \beta_2^B BND$

$$cov_{A,B} = \sum_{i,j}\beta_i, \beta_j\sigma_i\sigma_j = \beta_1^A\beta_1^B\sigma_{EQ}^2 + \beta_2^A\beta_2^B\sigma_{BND}^2 + \beta_1^A\beta_2^B cov_{EQ,BND} + \beta_2^A\beta_1^B cov_{EQ,BND}$$

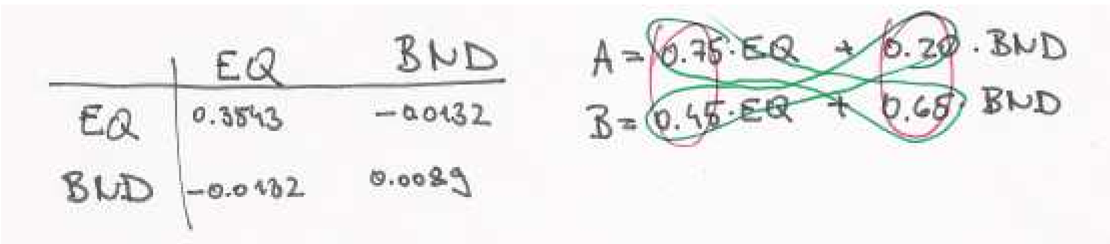


Figure 1.1: covariance in regression model

- mixture distribution

$$f(x; a_1, ..., a_n, b_1, ..., b_n) = \sum_i^n w_i p_i(x; a_i, b_i)$$
$$F(x; a_1, ..., a_n, b_1, ..., b_n) = \sum_i^n w_i F_i(x; a_i, b_i)$$

- goodness-of-fit - chi-square and Kolmogorov-Smirnov test

## 1.2 Regression and OLS

- Regardless homo/heteroskedasticity OLS estimator is unbiased, consistent and assymptotically normal. Homoskedasticity ensures that OLS estimators are efficient (i.e. the lowest volatility).
- adjusted $R^2$ penalizes for number $k$ of slopes in the regression model

$$\overline{R^2} = 1 - \frac{n-1}{n-k-1} R^2$$

- standard error of regression

$$SER = \sqrt{\frac{SSR}{n-k-1}}$$

- confidence interval of $E[Y|X]$ could be calculated as $E[Y|X] \pm \Phi(\alpha/2)\sigma_\epsilon$

$$Y = \beta_0 + \beta_1 X + \epsilon$$
$$E[Y|X] = \beta_0 + \beta_1 X$$

- error term and independent variable have constant volatility
- We have a linear model $Y = \beta_0 + \beta_1 X + \epsilon$. We know $\beta_1, \sigma_Y^2$ and $\sigma_\epsilon^2$. What is $\rho_{X,Y}$?

$$\sigma_Y^2 = \beta_1^2 \sigma_X^2 + \sigma_\epsilon^2$$
$$\sigma_X^2 = \frac{\sigma_Y^2 - \sigma_\epsilon^2}{\beta_1^2}$$
$$\beta_1 = \rho_{X,Y} \frac{\sigma_Y}{\sigma_X}$$
$$\rho_{X,Y} = \frac{\beta_1 \sigma_X}{\sigma_Y}$$

- homoskedasticity-only F-statistics

$$F = \frac{(SSR_{rest} - SSR_{unrest})/q}{SSR_{unrest}/(n-k-1)}$$
$$F = \frac{(R^2_{unrest} - R^2_{rest})/q}{(1 - R^2_{unrest})/(n-k-1)}$$

## 1.3 EWMA, GARCH

- GARCH model $\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2$
  - $u_{n-i}^2$ is weighted with $\alpha\beta^{i-1}$
  - persistence $\alpha + \beta$ - the larger the persistence the longer to revert to mean
  - volatility prediction $E[\sigma_{n+t}^2] = V_L + (\alpha+\beta)^t(\sigma_n^2 - V_L)$
  - mean reversion introduces autocorrelation into volatility estimates
    * $VaR_{n-days} = VaR_{1-day}\sqrt{n + n\rho}$
    * if $\sigma_n^2 < V_L \Rightarrow$ volatility is expected to grow $\Rightarrow$ positive autocorrelation $\Rightarrow$ square-root-of-time VaR underestimate the true VaR
    * if $\sigma_n^2 = V_L \Rightarrow$ volatility is expected to be stable $\Rightarrow$ no autocorrelation $\Rightarrow$ square-root-of-time VaR is a good estimate of the true VaR
    * if $\sigma_n^2 > V_L \Rightarrow$ volatility is expected to decrease $\Rightarrow$ negative autocorrelation $\Rightarrow$ square-root-of-time VaR overestimate the true VaR
- EWMA model $\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1-\lambda)u_{n-1}^2$
  - $u_{n-i}^2$ is weighted with $(1-\lambda)\lambda^{i-1}$

## 1.4 VaR

- historical simulation VaR is not subjected to model risk vs. may not recognize changes in volatility and correlations from structural changes
- VaR increases with increasing speed in quantile and increases with decreasing speed in holding period (the speed is inverse to expected gain on the underlying portfolio)
- operational VaR - basic and standardized approach - only positive income is taken into account
- expected credit loss: commitment = outstanding + (commitment - outstanding) * usage given default
- bootstraping method of simulation - you randomly draw from historical data

## 1.5 Ratings

- ratings and quantitative scoring methods cannot be compared
- default rates show statistically significant variations based on industry (vs. region)

## 1.6 CAPM and Related Measures

- $R_P = R_F + \beta(PR_{market} + PR_{country})$
- expected rate of return vs rate of return implied by CAPM

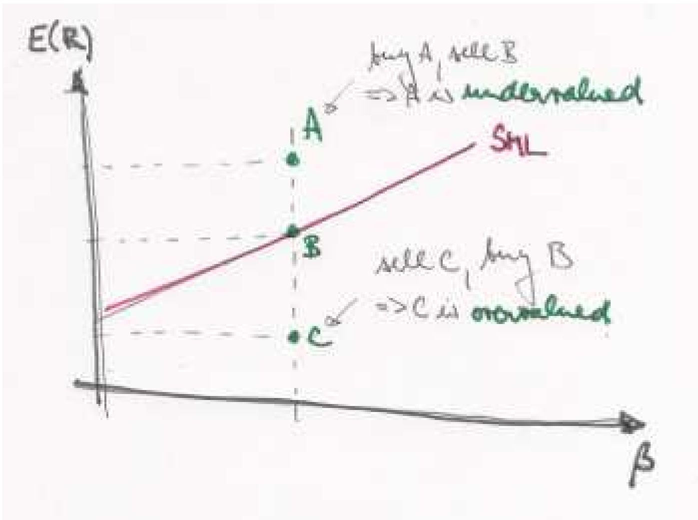


Figure 1.2: CAPM

## 1.7 Bonds

- government bonds are priced in 1/32; corporate and municipal bonds are priced in 1/8
- numerical delta and gamma calculation

$$D = \frac{1}{PV^0}\frac{PV^+ - PV^-}{\Delta y}$$
$$C = \frac{1}{PV^0}\frac{PV^+ + PV^- - 2PV^0}{(\Delta y)^2}$$

- modified duration of ZCB is slightly lower than its maturity; Macaulay duration of ZCB equals its maturity
- effective duration - takes into consideration changes in cash-flows resulting from changes in yield ⇒ more convenient for bonds with embedded options (vs. modified duration)
- perpetuity duration is $D = \frac{1}{y}$ ⇒ does not depend on coupon
- convexity grows with square of time and falls with coupon
- duration is relative measure ⇒ two bonds with the same duration but different NPV ⇒ the same relative price change but different USD price change

## 1.8 Mortgages

- recourse loan - allows the lender to go after the debtor's assets that were not used as loan collateral in case of default

## 1.9 Futures

- order types
  - market order - carried out immediately at current market price
  - limit order - executed at specified or more favorable price; waiting for low price to buy
  - stop order / stop-loss order - executed at specified or less favorable price; minimize loss
  - stop-limit order - current stock price is 50 USD, sell at 48 USD with limit of 47 USD $\Rightarrow$ conditional on bid or ask falling to 48 USD, sell order is is filled in if the sale can be exercised at 47 USD or higher (but may not be filled at all)
  - market-if-touched order - market order if a specified price is reached
  - discretionary / market-not-held order - market order with broker's discretion in an attempt to get a better price
  - time-of-day - specifies a particular time period of trading day
  - open order - can be executed during a given trading day
  - fill-or-kill - has to be executed immediately or not at all
- hedge effectivness - proportion of variance eliminated by hedging

$$effectivness = R^2 = \rho^2 = 1 - \frac{variation\ of\ basis}{variation\ of\ spot} = 1 - \frac{\sigma^2_{S-F}}{\sigma^2_S}$$

- forward price adjustmend

$$fwd = fut - \frac{1}{2}\rho\sigma^2 T_1 T_2$$

- long Eurodolar futures position = short bond position
- short hedge position, i.e. selling futures = long position in basis $\Rightarrow$ gain if basis streghtens

$$BS = S_T - F_0$$

## 1.10 Options

- rho vs moneyness
- American call can be exercised prior (not on) ex-dividend day
- American call might be exercised for $q > r$ and small $T$

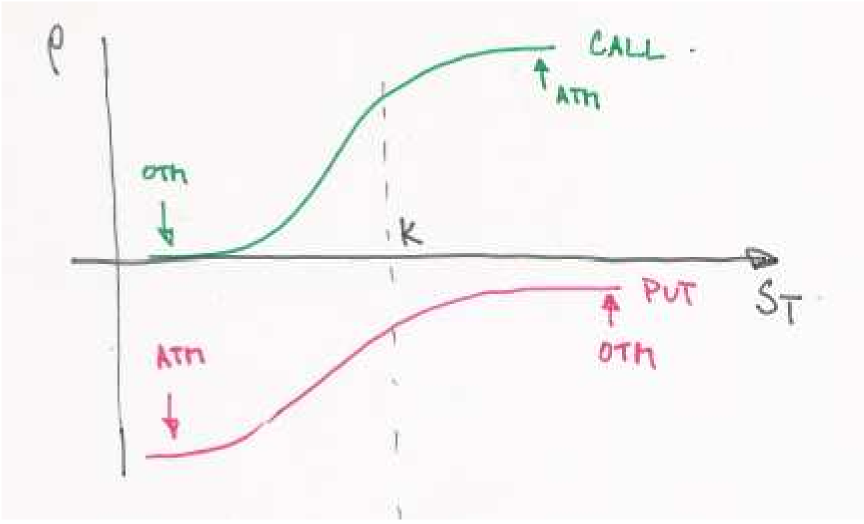


Figure 1.3: rho

- American put might be exercised if $q$ is small and $r$ increases (because stock prices tend to fall)
- limits on option prices are

$$max(S_T - Xe^{-rT}, 0) \leq c \leq S_0$$
$$max(S_T - Xe^{-rT}, 0) \leq C \leq S_0$$
$$max(Xe^{-rT} - S_T, 0) \leq p \leq Xe^{-rT}$$
$$max(X - S_T, 0) \leq P \leq X$$

- vega is the greatest for ATM options with long maturities
- horizontal spread = calendar spread
- collar

$$collar = p - c$$
$$collar = covered\ call\ +\ protective\ put\ =\ p - c + S_0$$

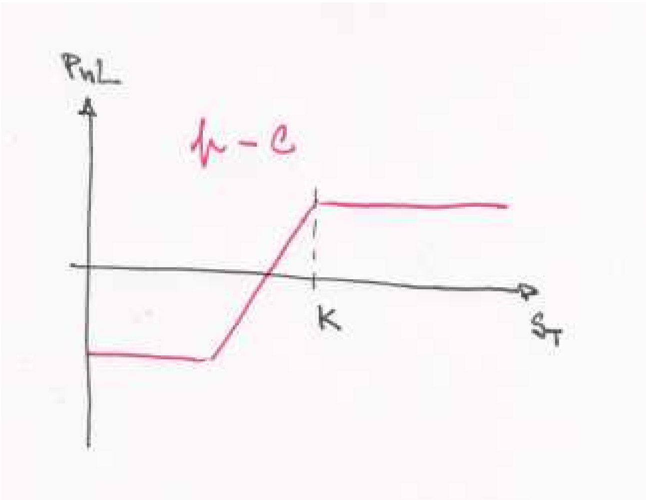


Figure 1.4: collar

# Chapter 2

# Time Series

## 2.1 Model Selecting

- to find a model with the smallest out-of-the-sample 1-ahead mean squared prediction error
- mean squared error - corresponds to selecting a model with the highest $R^2$; does not penalize for number of predictive variables $\Rightarrow$ in-sample overfitting
$$MSE = \frac{\sum_{t=1}^{T}}{T}$$
- MSE corrected for degrees of freedom - corresponds to selecting a model with the highest $\overline{R^2}$
$$s^2 = \frac{T}{T-k} MSE$$
- Akaike information criterion - heavier penalization for degrees of freedom than $s^2$
$$AIC = e^{\frac{2k}{T}} MSE$$
- Schwarz information criterion - heavier penalization for degrees of freedom than AIC
$$SIC = T^{\frac{k}{T}} MSE$$
- the lower the selection criterion the better
- consistency - probability of the true DGP (or DGP closest to the true DGP) selection increases with sample size
- effectivness - DGP 1-step-ahead forcast error variances approach the one that would be obtained using the true DGP with increasing sample size
- SIC is consistent but not efficient; AIC is efficient but not consistent
- if SIC and AIC select different DGP, we prefer SIC over AIC

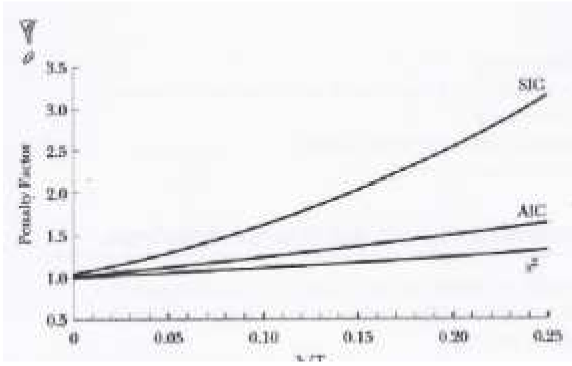


Figure 2.1: $s^2$, AIC and SIC "punishment" for number of degrees of freedom

## 2.2 Characterizing Cycles

- covariance stationarity - necesary if we want to predict time series
  - mean - $E[y_t] = \mu$
  - autocovariance - $\gamma(t, \tau) = cov(y_t, y_{t-\tau}) = \gamma(\tau)$

$$\begin{aligned} \gamma(\tau) &= \gamma(-\tau) \\ \gamma(0) &= cov(y_t, y_t) = D[y_t] \\ 0 &< \gamma(0) < \infty \end{aligned}$$

- many economic time series is not stationary but can be converted to stationary ones - from absolute to relative changes, remove seasonalities and trends
- autocorerlation function - ordinary correlation between $y_t$ and $y_{t-\tau}$

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$

- partial autocorerlation function - correlation between $y_t$ and $y_{t-\tau}$ with other $y_{t-i}$ being constant $\Rightarrow$ coeficient of autoregression function

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$

- white noise

$$y_t = \epsilon_t \quad \epsilon_t \sim (0, \sigma^2)$$

  - shock $\epsilon_t$ is serially uncorellated $Rightarrow$ serially uncorrelated white noise
  - if shock $\epsilon_t$ is normally distributed $Rightarrow$ independent white noise
  - autocovariance $\gamma(0) = \sigma^2 \quad \gamma(\tau \geq 0) = 0$
  - autocorrelation $\rho(0) = 1 \quad \rho(\tau \geq 0) = 0$

- partial autocorrelation $p(0) = 1 \quad p(\tau \geq 0) = 0$
- uncoditional mean and variance are constant by covariance stationarity vs. conditional mean and variance

- lag operator

$$Ly_t = y_{t-1}$$
$$L^m y_t = y_{t-m}$$
$$B(L)y_t = b_0 + b_1 L + b_2 L^2 + ...$$
$$B(L)\epsilon_y = b_0\epsilon_t + b_1\epsilon_{t-1} + b_2\epsilon_{t-2} + ... = \sum_{i=0}^{\infty} b_i \epsilon_{t-i}$$

- Wold's theorem - any zero-mean covariance-stationary process can be written as

$$y_t = B(L)\epsilon_t = \sum_{i=0}^{\infty} b_i \epsilon_{t-i} \quad b_0 = 1, \sum_{i=0}^{\infty} b_i^2 < \infty$$

- the above is called general linear process
- shock $\epsilon_t$ is serially uncorrelated but not necessarily independent
- if a covariance stationary series $y_t$ has mean of $\mu$ we define $y_t^* = y_t - \mu$
- unconditional moments - stable in time

$$E[y_t] = E\left[\sum_{i=0}^{\infty} b_i \epsilon_{t-i}\right] = 0$$
$$D[y_t] = D\left[\sum_{i=0}^{\infty} b_i \epsilon_{t-i}\right] = \sigma^2 \sum_{i=0}^{\infty} b_i^2$$

- conditional moments - move over time in response to the evolving information set

$$E[y_t|\Omega_{t-1}] = E\left[\sum_{i=0}^{\infty} b_i \epsilon_{t-i}|\Omega_{t-1}\right] = \sum_{i=1}^{\infty} b_i \epsilon_{t-i}$$
$$D[y_t|\Omega_{t-1}] = D\left[\sum_{i=0}^{\infty} b_i \epsilon_{t-i}|\Omega_{t-1}\right] = b_0\sigma^2 = \sigma^2$$

- infinite order polynomial $B(L)$ (and therefore Wold's theorem as well) can be approximated by rational polynomials

$$B(L) \approx \frac{\Theta(L)}{\Phi(L)} = \frac{\sum_{i=0}^{p} \theta_i L^i}{\sum_{i=0}^{q} \phi_i L^i}$$

- sample autocorrelation
  - sample mean

$$\overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

- * sample autocorrelation

$$\hat{\rho}(\tau) = \frac{\frac{1}{T}\sum_{t=\tau+1}^{T}(y_t - \overline{y})(y_{t-\tau} - \overline{y}))}{\frac{1}{T}\sum_{t=1}^{T}(y_t - \overline{y})^2}$$

- * if a particular autocorrelation $\hat{\rho}(\tau)$ is 0, then

$$\hat{\rho}(\tau) \sim N(0, 1/T)$$

- * if all autocorrelations are 0, then

$$T\hat{\rho}^2(\tau) \sim \chi_1^2$$
$$T\sum_{\tau=1}^{m}\hat{\rho}^2(\tau) \sim \chi_m^2$$

- * Box-Pierce Q-statistics $Q_{BP} = T\sum_{\tau=1}^{m}\hat{\rho}^2(\tau)$
- * Ljung-Box Q-statistics $Q_{LB} = T(T+2)\sum_{\tau=1}^{m}\left(\frac{1}{T-\tau}\right)\hat{\rho}^2(\tau)$
- * $m$ is usually selected in neighbordhood of $\sqrt{T}$

– sample partial autocorrelation - if the series is white noise, approximately 95% of the sample partial autocorrelations should fall in the intereval $\pm 2/\sqrt{T}$

$$\hat{y}_t = \hat{c} + \hat{\beta}_1 y_{t-1} + ... + \hat{\beta}_\tau y_{t-\tau}$$
$$\hat{p}(\tau) \equiv \hat{\beta}_\tau$$

– correlogram analysis - plot autocorrelations and partial autocorrelations for individual displacements $\Rightarrow$ tells with type process is the most sutaible

## 2.3 MA, AR and ARMA Models

### 2.3.1 MA Models

#### MA(1) Model

- MA(1): $y_t = \epsilon_t + \theta\epsilon_{t-1} = (1 + \theta L)\epsilon_t$ where $\epsilon_t$ is zero-mean white noise (not necessarily normaly distributed)
- always covariance stationary but not always invertible (requires $|\theta| < 1$)
- autocorrelation function was cut-off beyond $\tau > 0$
- partical autocorrelation function is oscilating
- unconditional moments

$$E[y_t] = E[\epsilon_t] + \theta[\epsilon_{t-1}] = 0$$
$$D[y_t] = D[\epsilon_t] + \theta^2 D[\epsilon_{t-1}] = (1 + \theta^2)\sigma^2$$

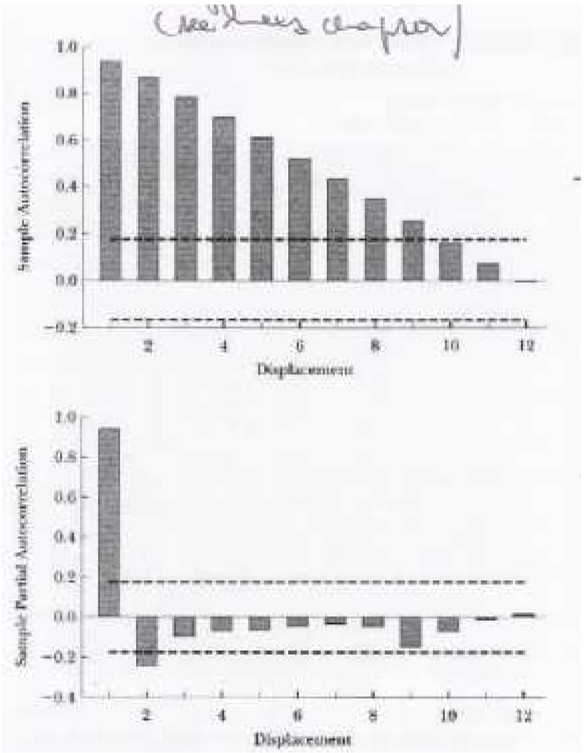


Figure 2.2: correlogram for autocorrelation and partial autocorrelation function

- conditional moments

$$E[y_t|\Omega_{t-1}] = E[\epsilon_t|\Omega_{t-1}] + \theta[\epsilon_{t-1}|\Omega_{t-1}] = \theta\epsilon_{t-1}$$
$$D[y_t|\Omega_{t-1}] = D[\epsilon_t|\Omega_{t-1}] + \theta^2 D[\epsilon_{t-1}|\Omega_{t-1}] = \sigma^2$$

- autocorrelation function $\gamma(0) = \theta\sigma^2 \quad \gamma(\theta > 0) = 0$
- MA(1) is invertible is $|\theta| < 1 \Rightarrow$ MA(1) can be expressed in autoregressive form (the below sum converge)

$$y_t = \epsilon_t + \theta\epsilon_{t-1}$$
$$\epsilon_{t-1} = y_{t-1} - \theta\epsilon_{t-2}$$
$$y_t = \epsilon_t + \theta y_{t-1} - \theta^2 y_{t-2} + \theta^3 y_{t-3} - ...$$

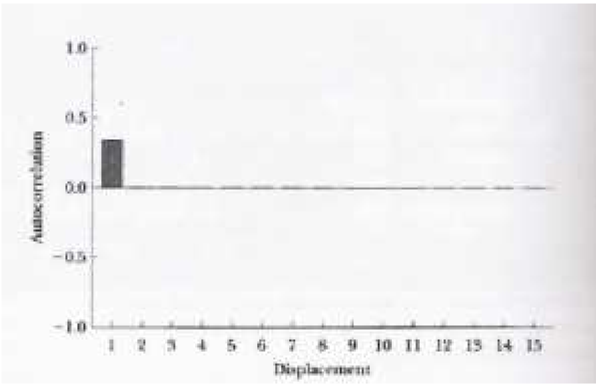


Figure 2.3: MA(1) - correlogram of autocorrelation function

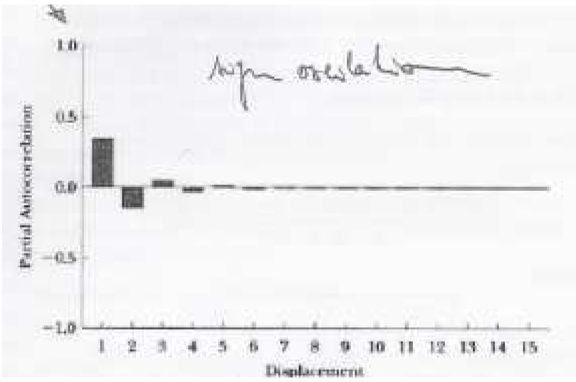


Figure 2.4: MA(1) - correlogram of partial autocorrelation function

**MA(q) Model**

- MA(q): $y_t = \epsilon_t + \theta_1\epsilon_{t-1} + ... + \theta_q\epsilon_{t-q} = \Theta(L)\epsilon_t$ where $\epsilon_t$ is zero-mean white noise (not necessarily normaly distributed)

- autocorrelation function cut-off is at displacement $q$
- partial autocorrelation function dacays gradually
- can capture richer dynamic patterns than MA(1) because of its higher order polynomial
- is always covariance stationary
- is invertible only if roots of $\Theta(L)$ are within a unit circle

### 2.3.2 AR Models

**AR(1) Model**

- AR(1): $y_t = \varphi y_{t-1} + \epsilon_t \Leftrightarrow (1 - \varphi L)y_t = \epsilon_t$ where $\epsilon_t$ is zero-mean white noise (not necessarily normaly distributed)
- always invertible but not always covariance stationary (requires $|\varphi| < 1$)
- lag operator in form of
$$y_t = \frac{1}{1 - \varphi L}\epsilon_t$$
- special case of MA($\infty$) with $\theta_i = \theta^i$
$$y_t = \epsilon_t + \varphi\epsilon_{t-1} + \varphi^2\epsilon_{t-2} + ...$$
- autocorrelation function decays gradually
- partial autocorrelation functions has cut-off at displacement 0
- fluctuation is much more persistent than in case of MA(1) $\Rightarrow$ able to capture much more persistent dynamics than MA(1)
- unconditional moments
$$E[y_t] = E[\epsilon_t + \varphi\epsilon_{t-1} + \varphi^2\epsilon_{t-2} + ...] = 0$$
$$D[y_t] = D[\epsilon_t + \varphi\epsilon_{t-1} + \varphi^2\epsilon_{t-2} + ...] = \frac{1}{1 - \varphi^2}\sigma^2$$
- conditional moments
$$E[y_t|\Omega_{t-1}] = E[\varphi y_{t-1} + \epsilon_t|\Omega_{t-1}] = \varphi y_{t-1}$$
$$D[y_t|\Omega_{t-1}] = D[\varphi y_{t-1} + \epsilon_t|\Omega_{t-1}] = \sigma^2$$
- covariance function - Yule-Walker equation
$$y_t y_{t-\tau} = \varphi y_{t-1} y_{t-\tau} + \epsilon y_{t-\tau}$$
$$E[y_t y_{t-\tau}] = E[\varphi y_{t-1} y_{t-\tau} + \epsilon y_{t-\tau}]$$
$$\gamma(\tau) = \varphi\gamma(\tau - 1)$$
$$\gamma(\tau) = \varphi^\tau \frac{\sigma^2}{1 - \varphi^2}$$
- autocorrelation function $\rho(\tau) = \varphi^\tau \Rightarrow$ autocorrelation function decays gradually; if $\varphi > 0$ the decay is one-sided
- partial autocorrelation function $p(1) = \varphi \quad p(\tau > 1) = 0$

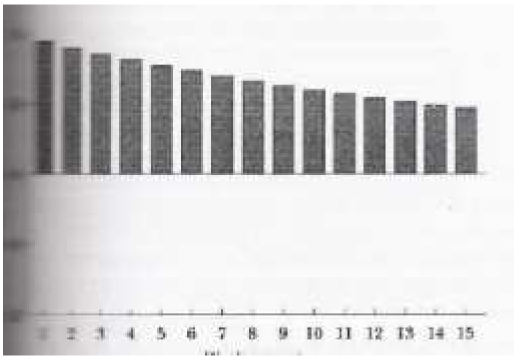


Figure 2.5: AR(1) - correlogram of autocorrelation function

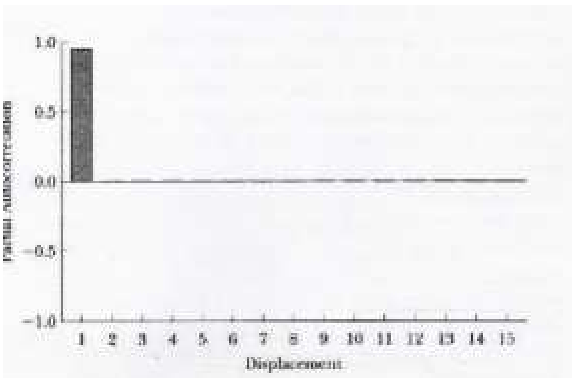


Figure 2.6: AR(1) - correlogram of partial autocorrelation function

**AR(q) Model**

- AR(q): $y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + ... + \varphi_q y_{t-q} + \epsilon_t \Leftrightarrow \epsilon_t = \Phi(L)y_t$ where $\epsilon_t$ is zero-mean white noise (not necessarily normaly distributed)

- autocorrelation function decays gradually (may oscilate)

- partial autocorrelation function has a cut-off at displacement $q$

- is always invertible

- is invertible only if roots of $\Phi(L)$ are withing a unit circle

### 2.3.3 ARMA Models

- neither autocorrelation nor partial autocorrelation functions have a cut-off

- more accurate approximation of Wold's theorem than MA or AR processes with the same number of parameters

- ARMA models creation

  - combining MA and AR models
  - random shock ifself is a moving average
  - AR processes subject to measurement error also turn out to be ARMA processes

- ARMA(1,1): $y_t = \varphi y_1 + \epsilon_t + \theta\epsilon_{t-1} \Leftrightarrow (1-\varphi)Ly_t = (1+\theta L)\epsilon_t$

  - is invertible if $|\theta| < 1 \Rightarrow \epsilon_t = \frac{1-\varphi L}{1+\theta L} y_t$
  - is covariance stationary if $|\varphi| < 1 \Rightarrow y_t = \frac{1+\theta L}{1-\varphi L}\epsilon_t$

- ARMA(p,q): $y_t = \varphi_1 y_{t-1} + ... + \varphi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} \Leftrightarrow \Theta(L)y_t = \Phi(L)\epsilon_t$

  - is invertible if inverses of all roots of $\Theta(L)$ are within a unit circle $\Rightarrow y_t = \frac{\Theta(L)}{\Phi(L)}\epsilon_t$
  - is covariance stationary if inverses of all roots of $\Phi(L)$ are within a unit circle $\Rightarrow \epsilon_t = \frac{\Phi(L)}{\Theta(L)} y_t$

### 2.3.4 Application

- select model with the lowest SIC / AIC

- plot collerogram for autocorrelation and partial autocorrelation function

### MA Processes

- non-linear in parameters $\Rightarrow$ parameter estimation using numerical minimalization

$$y_t = \epsilon_t + \theta\epsilon_{t-1}$$
$$\hat{y}_t \approx \theta y_{t-1} - \theta^2 y_{t-2} + ... + (-1)^{m+1}\theta^m y_{t-m} + \epsilon_t$$

- try to find $\theta$ that minimizes $\sum_{t=1}^{T}(\hat{y}_t - y_t)^2$
- if the model fits the underlying process $(\hat{y}_t - y_t)$ is a white noise

### AR Processes

- parameters could be easily estimated through linear regression

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

### ARMA Processes

- parameter estimation is similar to MA models
- AR(2) vs ARMA(3,1) and common factors - arises when $\varphi_i = \theta_i$

$$ARMA(0,0): \; y_t = \epsilon_t$$
$$ARMA(1,1): \; y_t - \alpha y_{t-1} = \epsilon_t - \alpha\epsilon_{t-1} \Rightarrow y_t = \alpha y_{t-1} + \epsilon_t - \alpha\epsilon_{t-1}$$