1. Definitions

- Stationary Time Series- A time series is stationary if the properties of the process such as the mean and variance are constant throughout time.
    i. If the autocorrelation dies out quickly the series should be considered stationary
    ii. If the autocorrelation dies out slowly this indicates that the process is non-stationary
- Nonstationarity- A time series is nonstationary if the properties of the process are not constant throughout time
    i. Unit Root Nonstationarity-
    ii. Random Walk with Drift-
- White Noise- A time series is called a white noise if a sequence of independent and identically distributed random variables with finite mean and variance, usually WN(0, σ^2). White noise has covariance
- Backward shift operator – a short hand for shift backward in the time series.

$$\beta Y_t = Y_{t-1} \qquad \beta^p Y_t = Y_{t-p}$$

2. Autocorrelation

- Measures the linear dependence or the correlation between r_t and r_{t-p}. (summarizes serial dependence)
- $$\rho_l = \frac{Cov(r_t, r_{t-l})}{\sqrt{Var(r_t)Var(r_{t-l})}} = \frac{Cov(r_t, r_{t-l})}{Var(r_t)}$$

    where $Var(r_t) = Var(r_{t-1})$ for weakly stationary process

- A way to check randomness in the data
- Lag 0 of the autocorrelation is 1 by definition
    i. If the autocorrelation dies out slowly this indicates that the process is non-stationary.
    ii. If all the ACFs are close to zero, then the series should be considered white noise.
- No Memory Series
    i. Autocorrelation function is zero
- Short Memory Series
    i. Autocorrelation function decays exponentially as a function of lag
- Long Memory Series
    i. Autocorrelation function decays at polynomial rate
    ii. The "differencing" exponent is between -½ and ½.

3. Partial Autocorrelation

- Correlation between observations X_t and X_{t+h} after removing the linear relationship of all observations in that fall between X_t and X_{t+h}.

- $$r_t = \phi_{0,1} + \boldsymbol{\phi_{1,1}} r_{t-1} + e_{1,t}$$
  $$r_t = \phi_{0,2} + \phi_{1,2} r_{t-1} + \boldsymbol{\phi_{2,2}} r_{t-2} + e_{2,t}$$
  $$r_t = \phi_{0,3} + \phi_{1,3} r_{t-1} + \phi_{2,3} r_{t-2} + \phi_{3,3} r_t + e_{3,t}$$
  $$\vdots$$

Each $\hat{\phi}_{p,p}$ is the lag-p PACF

- The PACF shows the added contribution of r_{t-p} to predicting r_t .

4. Diagnostics and Model Selection

- **Residual Diagnostics**
  - **i.** The residuals should be stationary – white noise
  - **ii.** The ACF and PACF should all be zero
    - **a.** If there is a long memory in the residuals then the assumptions are violated – nonstationarity of residuals

- **AIC (Akaike's Information Criterion)**
  - i. A measure of fit plus a penalty term for the number of parameters
  - ii. Corrected AIC- stronger penalty term ~ makes a difference with smaller sample sizes
  - iii. Choose the model that minimizes this adjusted measure fit
  - iv. AIC_k = log(MLE estimate of the noise variance) + 2k/T, where T is the sample size and k is the number of parameters in the model

- **Portmanteau Test**
  - i. Tests whether the first m correlations are zero vs. the alternative that at least one differs from zero.
  - ii. The sum of the first m squared correlation coefficients
  - iii. $H_0 : \rho_1 = ... = \rho_m = 0$, $H_a : \rho_i \neq 0$ where ρ_i is the autocorrelation
  - iv. Box and Pierce

    $$Q^*(m) = T \sum_{l=1}^{m} \hat{p}_l^{\,2}$$

    Q*(m) is asymptotically a chi-squared random variable with m degrees of freedom
  - v. Ljung and Box

    $$Q(m) = T(T+2) \sum_{l=1}^{m} \frac{\hat{\rho}_l^{\,2}}{T-l}$$

    Modified Box & Pierce statistic to increase power

- **Unit Root Test**
  - i. Derived in 1979 by Dickey and Fuller to test the presence of a unit root vs. a stationary process
  - ii. $\rho_t = \phi_1 \rho_{t-1} + e_t$ $\qquad$ $\rho_t = \phi_0 + \phi_1 \rho_{t-1} + e_t$

If $\phi_1 = 1$ then he series is said to have unit root and is not stationary. The unit root test determines if ϕ is significantly close to 1.

$$H_0 : \phi_1 = 1$$
$$H_A : \phi_1 < 1$$

iii. The behavior of the test statistics differs if it is a random walk with drift or if it is a random walk without drift.

5. Unit Root Nonstationary Process

- **Random Walk**
    i. The equation for a random walk is $\rho_t = \rho_{t-1} + a_t$, where ρ_0 denotes the starting values and a_t is white noise.
    ii. A random walk is not predictable and this can not be forecasted.
    iii. All forecasts of a random-walk model are simply the value of the series at the forest origin.
    iv. The series has a strong memory

- **Random Walk with Drift**
    i. $\rho_t = \mu + \rho_{t-1} + a_t$, where $\mu = E(\rho_t - \rho_{t-1})$

$$\rho_1 = \mu + \rho_0 + a_1$$
$$\rho_2 = \mu + \rho_1 + a_2 = 2\mu + \rho_0 + a_2 + a_1$$
$$\vdots$$
$$\rho_t = t\mu + \rho_0 + a_t + a_{t-1} + \ldots + a_1$$

A positive μ implies that the series eventually goes to infinity.

6. Differencing

- Reasons why the Difference is taken
    i. To transform non-stationary data into a stationary time series
    ii. To remove seasonal trends
        a. take 4th difference for quartly data
        b. take 12th difference for monthly data
- First Difference- The first difference of a time series is $z_t = y_t - y_{t-1}$
    i. A way to handle strong serial correlation of ACF is to take the first difference
- Second Difference- The second difference is $z_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$

7. Log Transformation

- Reasons to take log transformation
    i. Used to handle exponential growth of a series
    ii. Used to stabilize the variability
- Values must all be positive before the log is taken

i. If not all values are positive a positive constant can be added to every data point

8. Autoregressive Model

- A regression model in which r_t is predicted using past values, $r_{t-1}, r_{t-2}, \ldots$
  - i. AR(1) : $r_t = \phi_0 + \phi_1 r_{t-1} + a_t$, where a_t is a white noise series with zero mean and constant variance
  - ii. AR(p): $r_t = \phi_0 + \phi_1 r_{t-1} + \ldots + \phi_p r_{t-p} + a_t$
- Weak stationary is the sufficient and necessary condition of an AR model
  - i. For an AR model to be stationary all of its characteristic roots must be less than 1 in modulus
- ACF for Autoregressive Model
  - i. The ACF decays exponentially to zero
    1) For $\phi > 0$, the plot of ACF for AR(1) should decay exponentially
    2) For $\phi < 0$, the plot should consist of two alternating exponential decays with rate ϕ_1^2.
  - ii. The ACF for AR(1) $\rho_l = \phi_1 \rho_{l-1}$, because $\rho_0 = 1$ then $\rho_l = \phi_1^2$. So the ACF for the AR(1) should decay to exponentially with rate ϕ_1 starting at $\rho_0 = 1$
- PACF for Autoregressive Model
  - ii. The PACF is zero after the lag of the AR process
  - iii. $\hat{\phi}_{l,l}$ converges to zero for all $l > p$. Thus for AR(p) the PACF cuts off at lag p.

9. Moving Average Model

- A linear regression of the current value of the series against the white noise or random shocks of one or more prior values of the series.
  - i. $X_t = \mu + a_t - \theta_1 a_{t-1}$, where μ is the mean of the series, a_{t-i} are white noise, and θ_1 is a model parameter.
- The MA model is always stationary as it is the linear function of uncorrelated or independent random variables.
- The first two moments are time-invariant
- MA model can be viewed as a infinite order AR model
- ACF for Moving Average Model
  - ii. The ACF is zero after the largest lag of the process
- PACF for Moving Average Model
  - i. The PACF decays to zero

10. ARMA [p,q]

- The series r_t is a function of past values plus current and past values of the noise.
  - Combines an AR(p) model with a MA(q) model
- The equation for a ARMA(1,1) is $r_t = \phi_1 r_{t-1} + a_t + \theta_1 a_{t-1}$
- ACF for ARMA
  i. The ACF begins to decay exponentially to zero after the largest lag of the MA component.

11. ARIMA

- r_t is an ARIMA model if the first difference of r_t is an ARMA model.
- In an ARMA model, if the AR polynomial has 1 as the characteristic root, then the model is a ARIMA
- Unit-root nonstationary because it's AR has unit root.
- ARIMA has strong memory

12. ARFIMA

- A process is a fractional ARMA (ARFIMA) process if the fractional differenced series follows an ARMA(p,q) process. Thus if a series $(1-B)^d x_t$ follows ARMA(p,q) model, then the series is an ARFIMA(p,d,q).

13. Forecasting

- The multistep forecast converges to the mean of the series and the variances of forecast errors converge to the variance of the series.
- For AR Model
  i. The 1-step ahead forecast is the conditional expectation

  $$\hat{r}_h(1) = E(r_{h+1} \mid r_h, r_{h-1}, ...) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i}$$

  ii. For multistep ahead forecast: $\hat{r}_h(l) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+l-i}$

  iii. The forecast error for 1 –step ahead: $e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}$

  iv. Mean reverting. For a stationary AR(p) model, long –term point forecasts approach then unconditional mean. Also, the variance of the forecast approaches the unconditional variance of r_t.
- For MA Model
  i. Because the model has finite memory, its point forecasts go to the mean of the series quickly.
  ii. The 1-step ahead forecast for MA(1) is the conditional expectation
  $\hat{r}_h(1) = E(r_{h+1} \mid r_h, r_{h-1}, ...) = c_o - \theta_1 a_h$
  The 2-step ahead forecast for MA(1)
  $\hat{r}_h(1) = E(r_{h+1} \mid r_h, r_{h-1}, ...) = c_o$

iii. For a MA(q) model, the multistep ahead forecasts go to the mean after the first q steps.

14. Spectral Density

- A way of representing a time series in terms of harmonic components at various frequencies. Tells the dominant cycles or periods in the series
- Spectral Density is only appropriate for stationary time series data.
- A <u>Periodogram</u> at a particular frequency ω is proportional to the squared amplitude of the corresponding cosine wave, $\alpha\cos(\omega t)+\beta\sin(\omega t)$, fitted to the data using least squares.
- For a Covariance stationary time series(CSTS) with autocovariance function $\gamma(v)$, v = 0, ±1, ±2… the spectral density is given by

$$f(v)=\sum_{n=-\infty}^{\infty}\gamma(h)e^{-2\pi i v h}$$

where $v\in$ [-1/2, 1/2]

$$\gamma(h)=\int_{-1/2}^{1/2}e^{2\pi i v h}f(v)\,dv$$

15. VaR – Value at Risk

- Estimates the amount which an institution's position in a risk category could decline due to general market movements during a given holding period.
- Concerned with market risk
- In reality, used to assess risk or set margin requirements
  i. Ensures that financial institutions can still be in business after a catastrophic event
- Determined via forecasting
- If multivariate:
  i. $VaR=\sqrt{VaR_1^2+VaR_2^2+2\rho VaR_1 VaR_2}$

16. VAR – Vector Autoregressive Model

- A vector model used for multivariate time series
  i. VAR(1): $\mathbf{r}_t=\phi_0+\mathbf{\Phi}_1\mathbf{r}_{t-1}+\mathbf{a}_t$; where ϕ_0 is a k-dim vector, $\mathbf{\Phi}$ is a *k* x *k* matrix, and {$\mathbf{a}_t$} is a sequence of serially uncorrelated random vectors with mean zero and covariance matrix $\mathbf{\Sigma}$. $\mathbf{\Sigma}$ positive definite.
  ii. VAR(p): $\mathbf{r}_t=\phi_0+\mathbf{\Phi}_1\mathbf{r}_{t-1}+\ldots+\mathbf{\Phi}_p\mathbf{r}_{t-p}+\mathbf{a}_t$
- Can also model VMA and VARMA models
  i. One issue, VARMA has an *identifiability* problem (i.e. may not be uniquely defined
  ii. When VARMA models are used, you should only entertain lower order models.

17. Volatility Models

- ARCH
  i. Only an AR term
  ii. ARCH(m): $a_t = \sigma_t \varepsilon_t$
  $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + ... + \alpha_m a_{t-m}^2$
  iii. Weaknesses:

▪ Assume +ve & -ve shocks have same effects on volatility (i.e. use square of previous shocks to determine order) ➔ use "leverage to account for the fact that –ve shocks (i.e. "bad news") have larger impact on volatility than +ve shocks (i.e. "good news").
  ▪ Model is restrictive (see p.86, 3.3.2(2))

▪ Only *describes* the behavior of the conditional variance. Does not *explain* the source of the variations.

▪ Likely to over-predict the volatility since the respond slowly to large isolated shocks to the return series.

- GARCH – generalized ARCH
  i. Mean structure can be described by an ARMA model
  ii. GARCH(m,s):
  $$a_t = \sigma_t \varepsilon_t$$
  $$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2$$
  iii. Same weaknesses as the ARCH
  iv. If the AR component has a unit root, then we have an IGARCH model (i.e. Integrated GARCH; a.k.a. unit-root GARCH model)
  v. EGARCH (i.e. Exponential GARCH) allows for asymmetric effects between +ve & -ve asset returns. Models the log(cond. variance) as an ARMA. PRO: variances are guaranteed to be positive.

- GARCH-M - GARCH *in mean*

i. Used when the return of a security depends on its volatility

$$a_t = \sigma_t \varepsilon_t$$

ii. GARCH(1,1)-M: $r_t = \mu + c\sigma_t^2 + a_t$; where μ, c constant. A +ve c

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

indicates that the return is positively related to its past volatility.

iii. Cross-Correlation: series correlated against series^2; used to determine whether there exists volatility in the mean structure.

- Alternative GARCH models

1) CHARMA – Conditional heteroscedastic ARMA uses random coefficients to produce conditional heteroscedasticity.
2) RCA – Random Coefficient Autoregressive model accounts for variability among different subjects under study. Better suited for modeling the conditional mean as it allows for the parameters to evolve over time.
3) SV – Stochastic Volatility model is similar to an EGARCH but incorporates an innovation to the conditional variance equation.

4) LMSV – Long-Memory SV model allows for long memory in the volatility.

NOTE: Differencing ONLY effects mean structure, Log Transformation effects volatility structure.

18. MCMC Methods (Markov Chain Monte Carlo)

- Markov chain simulation creates a Markov process on Θ, which converges to a stationary transition distribution, $P(\theta, X)$.
- GIBBS SAMPLING (p.397)
    - Likelihood unknown, conditional dist$^{\underline{n}}$s known.
    - Need starting values
    - Sampling from cond. dist$^{\underline{n}}$s converges to sampling from the joint dist$^{\underline{n}}$.
    - PRO: Compared to MCMC, Gibbs can decompose a high-dim estimation problem into several lower-dim ones.
    - CON: When parameters are highly correlated, you should draw them jointly.
    - In practice, repeat several times with different starting values to ensure the algorithm has converged.
- BAYESIAN INFERENCE (p. 400)
    - Combines prior belief with data to obtain posterior dist$^{\underline{n}}$s on which statistical inference is based.