Characteristics of Time Series

Introduction

Time Series
- Set of observations $\{x_t\}$ recorded through time $t \in \{0, 1, 2, \ldots\}$ or $t \in \{0, \pm 1, \pm 2, \ldots\}$ or $t \in [0, \infty)$.
- Interested in the relationship between observations separated by k units of time; i.e. how the present depend on the past.

Process/Objectives
1. Descriptions
    - Graphical/nonparametric (eg. sample autocorrelation).
    - Suggest possible probability models.
    - Identify "unusual" observations.
    - stationary/nonstationary stochastic processes.
2. Modelling and Inference
    - Use data to build a model.
    - Check model fit.
3. Forecasting and Prediction
    - Forecast future values of the series.

Time vs Frequency Domain
1. Time Domain
    - Look at correlation (or some other dependence measure) between lagged values of the series; i.e. can x_{t+1} be predicted by $x_t, x_{t-1}, x_{t-2}, \ldots$?
    - Autoregressive model: $x_{t+1} = \sum_{k \le t} \phi_k x_k + \text{error}$ ..
2. Frequency Domain
    - Look at total variability of the time series.
    - Decompose variance $\operatorname{var}(x_t)$ into components of different frequencies (or periods); i.e. if $x_t = A \sin(2\pi\omega t) + B \cos(2\pi\omega t)$ where A, B are random variables with $\operatorname{var}(A) = \operatorname{var}(B) = \sigma^2$ and $\operatorname{cov}(A, B) = 0$, then $\operatorname{var}(x_t) = \sin^2(2\pi\omega t)\sigma^2 + \cos^2(2\pi\omega t)\sigma^2 = \sigma^2$.

Graphical Methods
Preliminary goal: Describe the overall structure of data; also want to identify memory type of the time series.
1. Short-Memory
    - Immediate past may give some information about immediate future but less information about "long-term" future.
2. Long-Memory
    - Past gives more information about "long-term" future.
    - Examples: series with polynomial trends or cycles/seasonality.

Time Series Statistical Models

Examples of Simple Models
1. $x_t \sim \text{iid}$ with $\mathrm{E}(x_t) = 0$, $\operatorname{var}(x_t) = \sigma^2$.
    - $x_t \sim \text{iid}\, N(0, \sigma^2)$, $t = 1, 2, \ldots$.

- Bernoulli trials: $x_t = \{ \begin{matrix} 1 \text{ wp } p \\ -1 \text{ wp } 1-p \end{matrix}$.

2. Random Walk
   - Start with $w_t \sim \text{iid}$.
   - Set $x_0 = 0$. Let $x_t = w_1 + w_2 + \cdots + w_t$ by cumulatively summing (or integrating) the w_t's.
   - It can be expressed as $x_t = x_{t-1} + w_t$.
   - Can also add a drift $x_t = \delta + x_{t-1} + w_t$.
3. Linear Trend
   - $x_t = a_0 + a_1 t + w_t$ where $w_t \sim \text{iid}$.
   - a_0 and a_1 are constants that can be estimated through least squares.
   - Plot residuals $e_t = x_t - \hat{a}_0 - \hat{a}_1 t$. The should look like the w_t's.
4. Seasonal Model
   - Example: series affected by the weather.
   - May need to include a periodic component with some fixed known period T.
   - $x_t = a_0 + a_1 \cos(2\pi\omega t) + a_2 \sin(2\pi\omega t) + w_t$ where $w_t \sim \text{iid}$. $\omega = \frac{1}{T}$ is a fixed frequency.
5. $w_t \sim WN(0, \sigma_w^2)$, i.e. $\text{E}(w_t) = 0$, $\text{var}(w_t) = \sigma_w^2$, w_t's uncorrelated.
6. First-Order Moving Average or MA(1) Process
   - $x_t = w_t + \theta w_{t-1}$, $t = 0, \pm 1, \ldots$, with $w_t \sim WN(0, \sigma_w^2)$.
   - Then $\text{E}(x_t) = 0$ and $\text{var}(x_t) = \sigma_w^2 (1 + \theta^2)$.
7. First-Order Autoregressive or AR(1) Process
   - $x_t = \phi x_{t-1} + w_t$, $t = 0, \pm 1, \ldots$, with $w_t \sim WN(0, \sigma_w^2)$.

General Modelling Approach

1. Plot the series. Examine the graph for:
   - a trend,
   - a seasonal component,
   - obvious change point,
   - "outliers".
2. Remove trend and seasonal components; this should provide "stationary" residuals. Two main techniques:
   - Estimate the trend and/or seasonal components and subtract them from the data.
   - Differencing the data. Replace x_t by $y_t = x_t - x_{t-1}$, or in general $y_t = x_t - x_{t-d}$ for some positive integer d.
3. Fit a model to the stationary residuals.
4. Forecast the residuals, then invert to forecast x_t.
5. Alternatively, use a frequency domain (or spectral) approach. Express the series in terns if sinusoidal waves of different frequencies, i.e. Fourier components.

Measures of Dependence: Autocorrelation and Cross-Correlation

Joint Distribution Function

Let $\{x_t\}$ $, t = t_1, t_2, \ldots$ be a time series. The joint distribution function is $F(c_1, \ldots, c_n) = P(x_{t_1} \le c_1, \ldots, x_{t_n} \le c_n)$.

- Not useful in practices.
- Special case: If $x_t \sim \text{iid}\, N(0,1)$, then $F(c_1, \ldots, c_n) = \prod_{t=1}^{n} \Phi(c_t)$ where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$.

Definition: Mean Function

The mean function is defined by $\mu_x = E(x_t) = \int_{-\infty}^{\infty} x f_t(x) dx$, where $f_t(x) = \frac{\partial F_t(x)}{\partial x}$ where $F_t(x) = P(x_t \le x)$.

Definition: Autocovariance Function
The autocovariance function of $\{x_t\}$ is defined by $\gamma(s,t) = E((x_s - \mu_s)(x_t - \mu_t))$, $\forall s, t$.
Note: $\gamma(t,t) = E((x_t - \mu_t)^2) = \operatorname{var}(x_t)$.

Definition: Autocorrelation Function (ACF)
The autocorrelation function is defined by $\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)}\sqrt{\gamma(t,t)}} = \frac{\operatorname{cov}(x_s, x_t)}{\sqrt{\operatorname{var}(x_s)}\sqrt{\operatorname{var}(x_t)}}$, $\forall s, t$.

Remark
If $x_t = \beta_0 + \beta_1 x_s$, then $\beta_1 > 0 \Rightarrow \rho(s,t) = 1$ and $\beta_1 < 0 \Rightarrow \rho(s,t) = -1$.

Stationary Time Series

Definition: Strictly Stationary
A strictly stationary time series has the property $P(x_{t_1} \le c_1, \ldots, x_{t_k} \le c_k) = P(x_{t_1+h} \le c_{1+h}, \ldots, x_{t_k+h} \le c_{k+h})$ for all $k = 1, 2, \ldots$, all $t_1, t_2, \ldots$, all $c_1, \ldots, c_k$, and all shifts $h = 0, \pm 1, \pm 2, \ldots$.

Definition: Weakly Stationary
A weakly stationary time series $\{x_t\}$ is a process such that

1. $E(x_t) = \mu$.
2. $\gamma(s,t)$ depends only upon the difference $|s-t|$.

Note: Refer this as stationary.

Remark
- Strict stationary implies weak stationary, but the converse is not true in general.
- Special case: If $\{x_t\}$ is Gaussian, then weak stationary implies strict stationary,

Notation
In a stationary time series, write $\gamma(t+h, t) \stackrel{\text{def}}{=} \gamma(h)$.

Definition: Autocorrelation Function (ACF) for Stationary Time Series
The autocorrelation function of a stationary time series is defined by $\rho(h) = \frac{\gamma(t+h,t)}{\sqrt{\gamma(t+h,t+h)}\sqrt{\gamma(t,t)}} = \frac{\gamma(h)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}}$.

Definition: Linear Process
A linear process $\{x_t\}$ is given by $x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}$, where $w_t \sim wn(0, \sigma^2)$ and $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

Remark

One can show for x_t that $\gamma(h)=\sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}$.

Definition: Gaussian Process
A process $\{x_t\}$ is a Gaussian process if the vector $\vec{x}=\left(x_{t_1},\ldots,x_{t_k}\right)^T$ have a multivariate normal distribution for every collection $t_{1,}\ldots,t_k$ and all $k>0$.

Estimation of Autorrelation

Assumptions
$x_{1,}\ldots,x_n$ are data points (one realization) of a stationary process.

Definition: Sample Mean

The sample mean is defined by $\bar{x}=\frac{1}{n}\sum_{t=1}^{n} x_t$.

Definition: Sample Autocovariance Function

The sample autocovariance function is defined by $\hat{\gamma}(h)=\frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h}-\bar{x})(x_t-\bar{x}) \quad ,h=0,1,\ldots,n-1$.

Definition: Sample Autocorrelation Function

The sample autocorrelation function is defined by $\hat{\rho}(h)=\frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$.

Property: Large Sample Distribution of Sample ACF

If $x_t \sim wn\left(0,\sigma^2\right)$, then for large n , $\hat{\rho}(h) \quad ,h=1,2,\ldots,H$ is approximately normal with mean 0 and variance $\frac{1}{n}$, i.e. $\hat{\rho}(h)\,\tilde{A}N(0,\frac{1}{n})$.

Time Series Regression and Exploratory Data Analysis

Exploratory Data Analysis

Definition: Backshift Operator
The backshift operator B is defined by $Bx_t=x_{t-1}$ and extends to $B^k x_t=x_{t-k}$. Also, $\nabla x_t=(1-B)x_t=x_t-x_{t-1}$ which extends to $\nabla^k x_t=(1-B)^k x_t$.

Note
∇x_t is an example of a linear filter. Differencing plays a central role in ARIMA modelling.

Smoothing in Time Series

Used to detect possible trends and seasonality.

General Setup

$x_t = f_t + y_t$ where f_t is a smooth function of time and y_t is stationary.

Example: Kernel Smoothing

$\hat{f}_t = \sum_{i=1}^{n} w_t(i) x_t$ where $w_t(i) = \dfrac{k\left(\frac{t-i}{b}\right)}{\sum_{j=1}^{n} k\left(\frac{t-i}{b}\right)}$. $k(\cdot)$ is called a kernel function; frequently it is chosen to be $k(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$.

ARIMA Models

Autoregressive Moving Average (ARMA) Models

Previously examined a classical regression approach, which is not sufficient for all time series encountered in practice.

Definition: Autoregression Model

An autoregression model of order p denoted by AR(p) is given by $x_t = \mu + \phi_1(x_{t-1} - \mu) + \cdots + \phi_p(x_{t-p} - \mu) + w_t$, where x_t is stationary, $\phi_{1,} \ldots, \phi_p$ are constants ($\phi_q \neq 0$), $w_t \sim wn(0, \sigma^2)$.

Note

- $E(x_t) = \mu$.
- Also, $x_t = \alpha + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t$ where $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$.

Definition: Autoregressive Operator

$\phi(B) = 1 - B - \cdots - B^p$ is known as the autoregressive operator.

Note

The AR(p) model can be written as $(1 - \phi_1 B - \cdots - \phi_p B^p) x_t = w_t$ or $\phi(B) x_t = w_t$.

Example: AR(1) Process

$x_t = \phi x_{t-1} + w_t$ where $t = 0, \pm 1, \pm 2, \ldots$ and $w_t \sim wn(0, \sigma^2)$.

- Iterating k times yields $x_t = \phi^k x_{t-k} + \sum_{j=0}^{k-1} \phi^j w_{t-j}$.
- If $|\phi| < 1$ and x_t is stationary, then $\lim_{k \to \infty} E\left[\left(x_t - \sum_{j=0}^{k-1} \phi^j w_{t-j}\right)^2\right] = 0$ (convergence in mean-square), and so $x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j}$ and $E(x_t) = \sum_{j=0}^{\infty} \phi^j E(w_{t-j}) = 0$.

- $\gamma(h)=\text{cov}(x_{t+h},x_t)=\sigma^2\phi^h\frac{1}{1-\phi^2}\quad ,h\geq 0$, and $\rho(h)=\frac{\gamma(h)}{\gamma(0)}=\phi^h\quad ,h\geq 0$.
- If $\phi>0$, smooth path; if $\phi<0$, choppy path.

Definition: Moving Average Model
A moving average model of order q denoted by MA(q) is given by $x_t=w_t+\theta_1 w_{t-1}+\cdots+\theta_q x_{t-q}$, where $\theta_{1,}\ldots,\theta_q$ are constants ($\theta_q\neq 0$), $w_t\sim\text{iid } N(0,\sigma^2)$.

Definition: Moving Average Operator
$\theta(B)=1+\theta_1 B+\cdots+\theta_q B^q$ is known as the moving average operator.

Note
The MA(q) process can be written as $x_t=\theta(B)w_t$.

Example: MA(1) Process
$x_t=w_t+\theta w_{t-1}$, where $w_t\sim\text{iid } N(0,\sigma^2)$.
Then $\gamma(h)=\begin{cases}(1+\theta^2) & h=0\\ \theta\sigma^2 & h=1\\ 0 & h>1\end{cases}$ and $\rho(h)=\begin{cases}\frac{\theta}{1+\theta} & h=1\\ 0 & h>1\end{cases}$.
Note that for MA(1) process, x_t is only correlated with x_{t-1} .

Definition: ARMA(p, q)
A time series $\{x_t\}$ is ARMA(p, q) if it is stationary and $x_t=\alpha+\phi_1 x_{t-1}+\cdots\phi_p x_{t-p}+w_t+\theta_1 w_{t-1}+\cdots\theta_q w_{t-q}$, where $w_t\sim\text{iid } N(0,\sigma^2)$ and $\alpha=\mu(1-\phi_1-\cdots-\phi_p)$.
Note: We can write $\phi(B)x_t=\theta(B)w_t$.

Definition: Casual
An ARMA(p, q) model $\phi(B)x_t=\theta(B)w_t$ is said to be causal if $\{x_t\}$ can be written as a linear process
$x_t=\sum_{j=0}^{\infty}\psi_j w_{t-j}=\psi(B)w_t$, where $\psi(B)=\sum_{j=0}^{\infty}\psi_j B^j$ and $\sum_{j=0}^{\infty}|\psi_j|<\infty$.

Example: AR(1) Process
$x_t=\phi x_{t-1}+w_t$ where $|\phi|<1$.
We have $x_t=\sum_{j=0}^{\infty}\phi^j w_{t-j}$. So AR(1) with $|\phi|<1$ is causal. Equivalently, the process is causal only when the root of $\phi(z)=1-\phi z$ is greater than 1 in absolute valule.

Example: Stationary MA(q) Process
$x_t=\theta(B)w_t$ where $w_t\sim wn(0,\sigma^2)$ and $\theta(B)=1+\theta_1 B+\cdots+\theta_q B^q$.

- $E(x_t)=\sum_{j=0}^{q}\theta_j E(w_{t-j})=0$.
- $\gamma(h)=\text{cov}(x_{t+h},x_t)=E\left[\left(\sum_{j=0}^{q}\theta_j w_{t+h-j}\right)\left(\sum_{j=0}^{q}\theta_j w_{t-j}\right)\right]=\begin{cases}\sigma^2\sum_{j=0}^{q-h}\theta_j\theta_{j+h} & 0\leq h\leq q\\ 0 & h>q\end{cases}$.

- $\rho(h)=\dfrac{\gamma(h)}{\gamma(0)}=\begin{cases}\dfrac{\sum_{j=0}^{q-h}\theta_j\theta_{j+h}}{\sum_{j=0}^{q}\theta_j^2} & 0\le h\le q\\ 0 & h>q\end{cases}$

Example: Stationary AR(p) Process

$x_t=\phi_1 x_{t-1}+\cdots+\phi_p x_{t-p}+w_t$ where $w_t\sim wn(0,\sigma^2)$.

- $E(x_t)=0$.
- $\gamma(h)=E(x_t x_{t+h})$ by stationarity. Use Yule-Walker equations $\gamma(h)=\phi_1\gamma(h-1)+\cdots+\phi_p\gamma(h-p)$ and $\gamma(0)=\sigma^2$ to find $\gamma(h)$.

Example: ARMA(1, 1)

For an ARMA(1, 1) process, $\gamma(h)=\dfrac{\sigma^2(1+\theta\phi)(\phi+\theta)}{1-\phi^2}\phi^{h-1}\ ,h\ge 1$ and $\rho(h)=\dfrac{(1+\theta\phi)(\phi+\theta)}{1+2\theta\phi+\theta^2}\phi^{h-1}\ ,h\ge 1$.

Estimation AR(p) Parameters

Estimating AR(p) Parameters

- Time series data: $x_1,\dots,x_n$.
- Model: $x_t=\phi_1 x_{t-1}+\cdots\phi_p x_{t-p}+w_t\ ,w_t\sim wn(0,\sigma^2)$.
- Estimate parameters $\phi_1,\dots,\phi_p$ based on observed series $x_1,\dots,x_n$.
- Methods available: Yule-Walker estimation, minimum distance estimation, maximum likelihood.

Yule-Walker Estimation

- $\begin{bmatrix}\gamma(0) & \cdots & \gamma(p-1)\\ \vdots & \ddots & \vdots\\ \gamma(p-1) & \cdots & \gamma(0)\end{bmatrix}\begin{bmatrix}\phi_1\\ \vdots\\ \phi_p\end{bmatrix}=\begin{bmatrix}\gamma(1)\\ \vdots\\ \gamma(p)\end{bmatrix}$ or $\Gamma\vec{\phi}=\vec{\gamma}$, and $\sigma^2=\gamma(0)-\vec{\phi}^T\vec{\gamma}$.
- Equations used to determine $\gamma(0),\dots,\gamma(p)$ from σ^2 and $\vec{\phi}$.
- However, we can substitute $\hat{\gamma}(h), h=0,\dots,p$ into equations to estimate σ^2 and $\vec{\phi}$.

Minimum Distance Estimations

- Minimize $\sum_{t=p+1}^{n} f(x_t-\phi_1 x_{t-1}-\cdots-\phi_p x_{t-p})$ where f is some function with $f(x)\to\infty$ as $|x|\to\infty$.
- Examples of f : $f(x)=x^2$ (least squares), $f(x)=|x|$.

Maximum Likelihood

- Assume $w_t\sim$ iid $N(0,\sigma^2)$.
- Use $x_1,\dots,x_n$ to obtain likelihood function – joint density of $x_1,\dots,x_n$ which is a function of unknown parameters.

Diagnostics for AR(p) Models

- How do we determine if AR(p) model is appropriate for data $x_1,\dots,x_n$?
- Residuals $e_t=x_t-\hat{\phi}_1 x_{t-1}-\cdots-\hat{\phi}_1 x_{t-1}\ ,t=1,\dots,p$ should look like white noise if the AR(p) model fits.

Portmanteau (Box-Pierce) Test

$T = n\sum_{j=1}^{n}\hat{\rho}^2(j)$, where $\hat{\rho}(h) \sim AN(0, \frac{1}{n})$ or $\sqrt{n}\,\hat{\rho}(h) \xrightarrow{d} N(0,1)$. Then T is approximately $\chi^2(h)$.

Definition: Partial Autocorrelation Function (PACF)

Define $\alpha(0)=1$, $\alpha(h)=\phi_{hh}, h \ge 1$ where $\begin{bmatrix}\phi_{1h}\\ \vdots \\ \phi_{hh}\end{bmatrix} = \begin{bmatrix}\gamma(0) & \cdots & \gamma(h-1)\\ \vdots & \ddots & \vdots \\ \gamma(h-1) & \cdots & \gamma(0)\end{bmatrix}^{-1}\begin{bmatrix}\gamma(1)\\ \vdots \\ \gamma(h)\end{bmatrix}$ or $\vec{\phi}_h = \Gamma_h^{-1}\vec{\gamma}_h$ (Yule-Walker).

Note: This is the correlation between x_t and x_{t+h} with dependence on $x_{t+1}, \ldots, x_{t+h-1}$ removed.

Sample Partial Autocorrelation Function (PACF)

$\hat{\alpha}(0)=1$, $\hat{\alpha}(h)=\hat{\phi}_{hh}$ where $\hat{\phi}_{hh}$ is determined from $\hat{\vec{\phi}}_h = \hat{\Gamma}_h^{-1}\hat{\vec{\gamma}}_h$. Also $\hat{\phi}_{hh} \sim AN(0, \frac{1}{n}) \Leftrightarrow \sqrt{n}\,\phi_{hh} \xrightarrow{d} N(0,1)$, $h > p$.

Lemma

If x_t is an AR(p) process, then $\alpha(h)=0$ for all $h > p$.

Procedure

Choose p such that $\hat{\alpha}(p+1), \hat{\alpha}(p+2), \ldots$ are close to zero (statistically zero).

Akaike Information Criterion (AIC)

- Compare different AR models.
- Based on likelihood function.
- How to compare? Look at estimates $\hat{\sigma}^2$ of $\operatorname{var}(w_t)$ for each model.
  - Estimate $\hat{\sigma}_p^2$ for each AR(p) model by Yule-Walker: $\hat{\sigma}_p^2 = \hat{\gamma}(0) - \hat{\phi}_1\hat{\gamma}(1) - \cdots - \hat{\phi}_p\hat{\gamma}(p)$.
  - Define $\mathrm{AIC}(p) = n\ln(\hat{\sigma}_p^2) + 2p$.
  - Choose model order that minimized $\mathrm{AIC}(p)$.
- Should use AIC together with graphical methods.
- AIC tends to choose higher order models. If wo models have similar AIC values, choose lower order model (parsimony).

Estimation of ARMA(p, q) Parameters

- $\phi(B)x_t = \theta(B)w_t$ where $w_t \sim \text{iid}\, N(0, \sigma^2)$. Then x_t is a Gaussian process.
- Given parameters $\phi_1, \ldots, \phi_p$, $\theta_1, \ldots, \theta_p$, and σ^2, one can determine the joint density $f(x_1, \ldots, x_n)$.
- Given parameters data , $x_1, \ldots, x_n$ one can determine the parameter values that maximizes $f(x_1, \ldots, x_n)$.

Linear Prediction for Stationary Processes

- Let x_t be a stationary process. So $E(x_t)=\mu$ and $\gamma(t, t+h)=\gamma(h)$.
- Predict x_{n+1} as a linear function of past values: $\hat{x}_{n+1} = \beta_0 + \beta_1 x_n + \cdots + \beta_n x_1$.
- Choose $\beta_0, \ldots, \beta_n$ to minimize the prediction MSE $P_n = E(x_{n+1} - \hat{x}_{n+1})$.
- We have $\begin{bmatrix}\gamma(0) & \cdots & \gamma(n-1)\\ \vdots & \ddots & \vdots \\ \gamma(n-1) & \cdots & \gamma(0)\end{bmatrix}\begin{bmatrix}\beta_1\\ \vdots \\ \beta_n\end{bmatrix} = \begin{bmatrix}\gamma(1)\\ \vdots \\ \gamma(n)\end{bmatrix}$ or $\Gamma\vec{\beta} = \vec{\gamma}$.

- Therefore the best linear predictor (BLP) is given by $\hat{x}_{n+1}=E(x_{n+1}|x_{1,}\ldots,x_n)=\beta_0+\beta_1 x_n+\cdots+\beta_n x_1=\mu+\beta_1(x_n-\mu)+\cdots+\beta_n(x_1-\mu)$ where $\beta_{0,}\ldots,\beta_n$ satisfies $\Gamma\vec{\beta}=\vec{\gamma}$.

ARMA Modelling

When do we fit an ARMA(p, q) model as opposed to a pure AR(p) or MA(q) model?

- Sample correlations are the key.
- Data appears to be stationary with no obvious trends (polynomial or seasonal) and the sample ACF $\hat{\rho}(h)$ decays fairly rapidly (geometrically).
- However, the sample ACF and PACF do not "cut-off" after some small lag, i.e. $\hat{\rho}(h)$ and $\hat{\alpha}(h)$ are not "small" for moderate h .

Diagnostics

- Observations: $x_{1,}\ldots,x_n$.
- Fit ARMA(p, q) model, i.e choose model order and obtain the MLEs of parameters $\hat{\vec{\phi}}$, $\hat{\vec{\theta}}$, and $\hat{\sigma}^2$.
- Then obtain the empirical predicted values $\hat{x}_t(\hat{\vec{\phi}},\hat{\vec{\theta}}),t=1,\ldots,n$ which depends upon the MLEs.
- Define the residues $e_t=\dfrac{x_t-\hat{x}_t(\hat{\vec{\phi}},\hat{\vec{\theta}})}{r_{t-1}(\hat{\vec{\phi}},\hat{\vec{\theta}})},t=1,\ldots,n$ where $r_{t-1}(\ )=\dfrac{E\left((x_t-\hat{x}_t(\cdot))^2\right)}{\sigma^2}=\dfrac{P_{t-1}}{\sigma^2}$.

Model Building Non-Stationary Data

How do we detect non-stationarity?

- Time series plot: data contains a trend and/or seasonal component.
- Data changes slowly over time, no obvious equilibrium value.
- Sample ACF: $\hat{\rho}(h)$ decays to 0 very slowly or oscillate around 0; does not exhibit geometric decay of ARMA(p, q).

ARIMA Models

Definition

Let $d=1,2,\ldots$ be nonnegative integers. $\{x_t\}$ is an ARIMA(p, q, d) process if $y_t=(1-B)^d x_t$ is an ARMA(p, q) process.

Basic Steps for ARIMA Modelling

1. Time series plot. Detect non-stationarity (sample ACF $\hat{\rho}(h)$); possible transformations (eg: Box-Cox transformations, $\log x_t$ for changing variance).
2. If data is non-stationary, try differencing. Look at plot and sample ACF of differenced data; $d=1$ or 2 should be sufficient – don't over difference.
3. Fit ARMA(p, q) to differenced data. Use AIC, sample PACF (for AR(p)), sample ACF (for MA(q)).
4. Test goodness-of-fit of ARMA(p, q). Residuals close to white noise.
5. Use model for prediction or description.

Note

Don't over difference. If x_t is a stationary ARMA(p, q) process, then ∇x_t is a ARMA(p, q + 1).

Note

Recall conditions for ARMA(p, q) model to be stationary and causal. $\phi(B)x_t=\theta(B)w_t\ \ ,w_t\sim WN(0,\sigma^2)$. x_t is

stationary if and only if $\phi(z) \neq 0 \, \forall |z|=1$, i.e. $\phi(z)$ has no unit root. If $\phi(z) \neq 0 \, \forall |z| \leq 1$, then x_t is causal and $x_t = \sum_{j=1}^{\infty} \psi_j w_{t-j}$ (MA(∞)).

Note
ARIMA(p, d, q): $\phi(B)(1-B)^d x_t = \theta(B) w_t$, or equivalently, $\phi^*(B) x_t = \theta(B) w_t$. $\phi^*(z)$ has a root (of order d) at $z=1$ since $\phi^*(z) = \phi(z)(1-z)^d$.

- Very specific type of non-stationarity.
- Useful for slowly decaying sample ACF's where the data has a polynomial trend.

Unit Root Test
- Can test the hypothesis $H_0 : \phi = 1$ in AR(1) process (non-stationary random walk) vs $H_1 : |\phi| < 1$ (stationary).
- Unit roots in $\phi(z)$ are very common in economic and financial time series.
- If we reject H_0 , then differencing is not necessary.
- Tests can be extended to AR(p) processes: $H_0 : \phi_1 + \cdots + \phi_p = 1$.
- In summary, if unit root in $\phi(z)$, data should be differenced; if unit root in $\theta(z)$, data over differenced.

Seasonal ARIMA Models (SARIMA)

Special Case
- Suppose series x_t has a period s (deterministic or random).
- Differenced series: $y_t = (1-B^s) x_t = x_t - x_{t-s}$.
- Fit ARMA(p, q) model to y_t : $\phi(B) y_t = \theta(B) w_t \ , w_t \sim WN(0, \sigma^2)$, i.e. $\phi(B)(1-B^s) = \theta(B) w_t$.

Definition: SARIMA Process
x_t is SARIMA $(p, d, q) \times (P, D, Q)_s$ process with period s if $y_t = (1-B)^d (1-B^s)^D x_t$ is an ARMA model. That is, $\phi(B) \Phi(B^s) y_t = \theta(B) \Theta(B^s) w_t$ where $\Phi(z) = 1 - \Phi_1 z - \cdots - \Phi_P z^P$ and $\Theta(z) = 1 + \Theta_1 z + \cdots + \Theta_Q z^Q$.

Note
- d and D are nonnegative integers.
- Usually $d=1$ and $D=1$ is sufficient.

Note
x_t is expressed as $\phi(B) \Phi(B^s)(1-B)^d (1-B^s)^D x_t = \theta(B) \Theta(B^s) w_t$.

- $(1-B)^d$ is trend or low frequency.
- $(1-B^s)^D$ is seasonal component.
- SARIMA models both polynomial trend and seasonality (periodicity) at the same time.

Note
- Models become complicated very quickly, making interpretation difficult.
- Potentially many models to assess for a given set of data.
- Usually low-order models suffice: $d, D = 0, 1$, $p, q \leq 2$.

SARIMA Model Selection Strategies

Strategy A

1. Difference data so it appears stationary.
2. Assume $P, Q=0$; choose p and q using AIC like in ARMA case.
3. Carry out diagnostics, i.e. white noise checks.

Strategy B

1. Difference data so it appears stationary.
2. Assume $p=P=0$; choose q and Q by looking at sample ACF $\hat{\rho}(1), \hat{\rho}(2), \ldots$ and $\hat{\rho}(s), \hat{\rho}(2s), \ldots$, and choose q and Q to be "cut off" points.
3. Carry out diagnostics, i.e. white noise checks.

Spectral Analysis and Filtering

- Stationarity implies temporal regularity.
- Two approaches to modelling this regular behavior:
  - Time Domain. Essentially regression of the present on the past via linear difference equations, i.e. ARMA $\phi(B) x_t = \theta(B) w_t \;\; , w_t \sim WN(0, \sigma^2)$.
  - Frequency Domain. Regression of the present on linear combinations of periodic sines and cosines with random amplitudes: $x_t = \sum_{k=1}^{q} U_{k_1} \cos(2\pi\omega_k t) + U_{k_2} \sin(2\pi\omega_k t)$, where U_{k_1} and U_{k_2} are independent random variables with $E(U_{k_1}) = E(U_{k_2}) = 0$ and $\mathrm{var}(U_{k_1}) = \mathrm{var}(U_{k_2}) = \sigma_k^2$

Notation

- ω is frequency; cycles per time.
- T is period; time per cycle, $T = \frac{1}{\omega}$.

Example

$x_t = A\cos(2\pi\omega t + \phi)$ where A and ϕ are independent random variables. We can rewrite x_t in regression form as $x_t = U_1 \cos(2\pi\omega t) + U_2 \sin(2\pi\omega t)$ where $U_1 = A\cos\phi$ and $U_2 = A\sin\phi$.

PERIODOGRAM

Identify possible periodicities or cycles in the data.

Definition: Fundamental Fourier Frequencies

Let $w_k = \frac{k}{n}$ where $k \in \mathbb{Z}$ and $-\left\lfloor \frac{n-1}{2} \right\rfloor \le k \le \left\lfloor \frac{n}{2} \right\rfloor$. Let $F_n = \left\{ w_k = \frac{k}{n} \middle| k = -\left\lfloor \frac{n-1}{2} \right\rfloor, \ldots, \left\lfloor \frac{n}{2} \right\rfloor \right\}$ be the set of fundamental Fourier frequencies.

Note: $\lfloor \cdot \rfloor$ denotes the greatest integer function.

Note: $F_n \subseteq \left[-\frac{1}{2}, \frac{1}{2} \right]$.

Definition: Discrete Fourier Transform

Let $\vec{e}_k = \frac{1}{\sqrt{n}} \begin{pmatrix} e^{i2\pi\omega_k} \\ \vdots \\ e^{ni2\pi\omega_k} \end{pmatrix}$ $, k = -\left\lfloor \frac{n-1}{2} \right\rfloor, \ldots, \left\lfloor \frac{n}{2} \right\rfloor$. It can be shown that $\{\vec{e}_1, \ldots, \vec{e}_n\}$ is an orthonormal basis for $\mathbb{C}^n$. Any vector $\vec{x} \in \mathbb{C}^n$ can be written as $\vec{x} = \sum_{k=-\lfloor\frac{n-1}{2}\rfloor}^{\lfloor\frac{n}{2}\rfloor} a_k \vec{e}_k$ where $a_k = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{-2\pi i \omega_k t}$.

$\{a_k\}$ is called the discrete Fourier transform of the sequence $\{x_1, \ldots, x_n\}$ (think of $\{x_t\}$ as the sample).

Note: From $\vec{x} = \sum_{k=-\lfloor\frac{n-1}{2}\rfloor}^{\lfloor\frac{n}{2}\rfloor} a_k \vec{e}_k$, the t-th component is $x_t = \sum_{k=-\lfloor\frac{n-1}{2}\rfloor}^{\lfloor\frac{n}{2}\rfloor} a_k e^{2\pi i \omega_k t} = \sum_{k=-\lfloor\frac{n-1}{2}\rfloor}^{\lfloor\frac{n}{2}\rfloor} a_k (\cos(2\pi\omega_k t) + i\sin(2\pi\omega_k t))$, or equivalently, $x_t = \sum_{j=1}^{\frac{n-1}{2}} b_j \cos(2\pi \frac{j}{n} t) + c_j \sin(2\pi \frac{j}{n} t)$ since $\sin w = \frac{e^{iw} - e^{-iw}}{2i}$ and $\cos w = \frac{e^{iw} + e^{-iw}}{2}$. So x_t $, t = 1, \ldots, n$ can be expressed as an exact linear combination of sine waves with frequencies $\omega_k \in F_n$.

Definition: Periodogram

Define $I_n(\lambda) = \frac{1}{n} \left| \sum_{t=1}^{n} x_t e^{-2\pi i \lambda t} \right|^2$. If $\lambda = \omega_k$, then the periodogram is defined to be

$$I_n(\omega_k) = \frac{1}{n} \left| \sum_{t=1}^{n} x_t e^{-2\pi i \omega_k t} \right|^2 = |a_k|^2 = \frac{1}{n} \left(\sum_{t=1}^{n} x_t \cos(2\pi\omega_k t) \right)^2 + \frac{1}{n} \left(\sum_{t=1}^{n} x_t \sin(2\pi\omega_k t) \right)^2.$$

Note

- This measures the dependence (correlation) of the data $x_1, \ldots, x_n$ on sinusoides oscillating at a given frequency ω_k.
- If data contains strong periodic behavior corresponding to a frequency ω_k, then $I_n(\omega_k)$ will be large.

Note

It can be shown that if $x_1, \ldots, x_n \in \mathbb{R}$ and $\omega_k = \frac{k}{n}$ is a Fourier frequency, then $I_n(\omega_k) = \sum_{h=-n+1}^{n-1} \hat{\gamma}(h) e^{-2\pi i \omega_k h}$ where $\hat{\gamma}(h)$ is the sample ACV function.

Spectral Densities

Definition: Spectral Density

Suppose x_t is a zero-mean stationary time series with an absolutely summable ACV function $\gamma(\cdot)$, i.e. $\sum_{n=-\infty}^{\infty} |\gamma(h)| < \infty$. The spectral density of x_t is the function $f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h}$ $, \omega \in \left[-\frac{1}{2}, \frac{1}{2}\right]$.

Note

- Compare $f(\omega)$ with $I_n(\omega)$; periodogram is used to estimate the spectral density.
- f can be written as $f(\omega) = \gamma(0) + 2\sum_{h=1}^{\infty} \gamma(h) \cos(2\pi\omega h)$.
- Also, $f(\omega) = f(-\omega)$, i.e. f is an even function.
- We can recover the ACV function from f. It can be shown $\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i \omega h} f(\omega) \, d\omega = \int_{-1/2}^{1/2} \cos(2\pi\omega h) f(\omega) \, d\omega$.

Spectral Representation of ACV

Theorem

Let x_t be a stationary process with ACV γ. Then:

1. There exists a non-decreasing function F on $\left[-\frac{1}{2}, \frac{1}{2}\right]$ such that $\gamma(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i \omega h} dF(\omega) \ \forall h \in \mathbb{Z}$.
    - $F\left(-\frac{1}{2}\right) = 0$, $F\left(\frac{1}{2}\right) = \gamma(0) = \text{var}(x_t)$.
    - F is called the spectral distribution function. F is also called a generalized distribution function because $G(w) \overset{\text{def}}{=} \frac{F(\omega)}{F(1/2)}$ is a probability function.
2. If $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, then F is differentiable with $\frac{dF(\omega)}{d\omega} = f(\omega)$. In this case, $\gamma(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i \omega h} f(\omega) d\omega$ where f is the spectral density of x_t.

Spectral Representation of x_t

Theorem

Suppose x_t is a zero-mean stationary process with ACV γ and spectral distribution function F. Then there exist two continuous real-valued processes $C(\omega)$ and $S(\omega)$, $-\frac{1}{2} \le \omega \le \frac{1}{2}$, with

- $\text{cov}(C(\omega_1), S(\omega_2))$, $\forall \omega_1, \omega_2$
- $\text{var}(C(\omega_2) - C(\omega_1)) = \text{var}(S(\omega_2) - S(\omega_1)) = F(\omega_2) - F(\omega_1)$ for $\omega_1 \le \omega 2$,
- $\text{cov}(C(\omega+\Delta) - C(\omega), C(\omega)) = \text{cov}(S(\omega+\Delta) - S(\omega), S(\omega)) = 0$ for $\Delta \ne 0$,

such that

$$
\begin{aligned}
x_t &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos(2\pi \omega t) dC(\omega) + \int_{-\frac{1}{2}}^{\frac{1}{2}} \sin(2\pi \omega t) dS(\omega) \\
&= \lim_{n\to\infty} \sum_{j=-\left\lfloor \frac{n}{2} \right\rfloor}^{\left\lfloor \frac{n}{2} \right\rfloor} \cos\left(\frac{2\pi j t}{n}\right)\left[C\left(\frac{j}{n}\right) - C\left(\frac{j-1}{n}\right)\right] + \lim_{n\to\infty} \sum_{j=-\left\lfloor \frac{n}{2} \right\rfloor}^{\left\lfloor \frac{n}{2} \right\rfloor} \sin\left(\frac{2\pi j t}{n}\right)\left[S\left(\frac{j}{n}\right) - S\left(\frac{j-1}{n}\right)\right].
\end{aligned}
$$

This is equivalent to $x_t = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-2\pi i \omega t} dz(\omega)$ where $z(\omega)$ is a complex valued stochastic process on $\left[-\frac{1}{2}, \frac{1}{2}\right]$ having stationary uncorrelated increments and $\text{var}(z(\omega_2) - z(\omega_1)) = F(\omega_2) - F(\omega_1)$ for $\omega_1 \le \omega 2$.

Example: Application of the Spectral Representation Theorem

Suppose x_t has spectral density $f(\omega) = F'(\omega)$. Let n be an arbitrarily large integer. Define $A_k = C\left(\frac{k}{n}\right) - C\left(\frac{k-1}{n}\right)$ and $B_k = S\left(\frac{k}{n}\right) - S\left(\frac{k-1}{n}\right)$; they are mutually uncorrelated and that $\text{cov}(A_j, A_k) = \text{cov}(B_j, B_k) = 0 \ \forall j \ne k$ and $\text{cov}(A_j, B_k) = 0 \ \forall j, k$. Also, $\text{var}(A_k) = \text{var}(B_k) = F\left(\frac{k}{n}\right) - F\left(\frac{k-1}{n}\right) \approx f\left(\frac{k}{n}\right) \times \frac{1}{n}$ since $f(\omega) = \frac{dF(\omega)}{d\omega}$. Then we have

Reimann sums $x_t=\sum_{j=-\left\lfloor\frac{n}{2}\right\rfloor}^{\left\lfloor\frac{n}{2}\right\rfloor} A_k\cos\left(\frac{2\pi k t}{n}\right)+\sum_{j=-\left\lfloor\frac{n}{2}\right\rfloor}^{\left\lfloor\frac{n}{2}\right\rfloor} B_k\sin\left(\frac{2\pi k t}{n}\right)$, and hence $\operatorname{var}(x_t)=\sum_{j=-\left\lfloor\frac{n}{2}\right\rfloor}^{\left\lfloor\frac{n}{2}\right\rfloor}\cos^2\left(\frac{2\pi k t}{n}\right)\operatorname{var}(A_k)+\sum_{j=-\left\lfloor\frac{n}{2}\right\rfloor}^{\left\lfloor\frac{n}{2}\right\rfloor}\sin^2\left(\frac{2\pi k t}{n}\right)\operatorname{var}(B_k)\approx\int_{-\frac{1}{2}}^{\frac{1}{2}} f(\omega)d\omega$. So the spectral density f provides the proportion of the total variance $\operatorname{var}(x_t)$ for a given frequency.

Filtering

Theorem: Filtering Theorem

Let x_t be a stationary time series with $E(x_t)=0$ and spectral density $f_x(\omega)$. Let $y_t=\sum_{j=-\infty}^{\infty}\psi_j x_{t-j}$ where $y_t=\sum_{j=-\infty}^{\infty}|\psi_j|<\infty$. Then y_t is stationary with $E(y_t)=0$ and $f_y(\omega)=\left|\Psi(e^{-2\pi i\omega})\right|^2 f_x(\omega)=\Psi(e^{-2\pi i\omega})\Psi(e^{2\pi i\omega})f_x(\omega)$ where $\Psi(e^{-2\pi i\omega})=\sum_{j=-\infty}^{\infty}\psi_j e^{-2\pi i\omega j}$. Ψ is called the transfer or frequency response function; $|\Psi|^2$ is called the squared frequency response.

Filtering Time Series

Goals:

- Estimate trends, "smooth" the data.
- Remove trends and seasonality.
- Make non-stationary series stationary.

Linear Filters

Given a time series $x_{1,}\ldots,x_n$, we want to replace x_t be a linear combination of past and/or future values, i.e. $x_t\rightarrow y_t=\sum_u c_u x_{t-u}$.

1. Moving average filter $y_t=\sum_{u=-r}^{r} c_u x_{t-u}$ where $c_u>0$ and $\sum_{u=-r}^{r} c_u=1$. Used to estimate underlying/hidden ternds.
2. Differencing $y_t=\nabla x_t=x_t-x_{t-1}=\sum_{u=-\infty}^{\infty} c_u x_{t-u}$ where $c_1=-1$, $c_0=1$, $c_u=0, u\neq 0, 1$.
3. Seasonal differencing $y_t=\nabla_k x_t=x_t-x_{t-k}$.

GARCH Modelling

ARCH Modelling

- Autoregression conditional heteroscedastic.
- Model time-varying second moments, i.e. variances.
- Very important for modelling financial time series.

Definition: ARCH(1) Model

Note that $y_t=\ln\left(\frac{P_t}{P_{t-1}}\right)$ has financial interpretation as the continuously compounded daily return, i.e. $P_t=P_{t-1}e^{y_t}$.

Model: y_t is a solution of $y_t=\sigma_t\varepsilon_t$ where $\varepsilon_t\sim\text{iid } N(0,1)$ and $\sigma_t=\alpha_0+\alpha_1 y_{t-1}^2$, $\alpha_0>0, \alpha_1\geq 0$.

Note

- If $\sigma_t=\sigma\ \forall t$ (constant), then $y_t=\sigma\varepsilon_t$. This is equivalent to $\ln P_t$ following a simple random walk $\ln P_t=\ln P_{t-1}+\sigma\varepsilon_t$.
- Also, in this case where $\sigma_t=\sigma\ \forall t$, $y_t\sim\text{iid } N(0,\sigma^2)$. This is not the case for ARCH(1) where $\sigma_t=\alpha_0+\alpha_1 y_{t-1}^2$.

Note

- $E(y_t|y_{t-1})=\sigma_t E(\varepsilon_t|y_{t-1})=0$ since $E(\varepsilon_t)=0$, and that $\text{var}(y_t|y_{t-1})=\sigma_t\text{var}(\varepsilon_t|y_{t-1})=\alpha_0+\alpha_1 y_{t-1}^2$ since $\text{var}(\varepsilon_t)=1$. Thus $y_t|y_{t-1}\sim N(0,\sigma_t^2)$ (conditional distribution).
- From $\begin{cases} y_t^2=\sigma_T^2\varepsilon_T^2 \\ \alpha_0+\alpha_1 y_{t-1}^2=\sigma_t^2 \end{cases}$, we get $y_t^2=\alpha_0+\alpha_1 y_{t-1}^2+\sigma_T^2(\varepsilon_t^2-1)$. Here $\varepsilon_t-1\sim\chi_{(1)}^2$ with $E(\varepsilon_t-1)=0$. Let $u_t=\sigma_T^2(\varepsilon_t^2-1)$. Then $E(u_t|y_s, s\leq t-1)=0$ and hence $E(u_t)=E(E(u_t|y_s, s\leq t-1))=0$. Therefore $y_t^2=\alpha_0+\alpha_1 y_{t-1}^2+u_t$ is a non-Gaussion AR(1) process.
- By recursively substituting, we can write $y_t^2=\alpha_0\sum_{j=0}^{\infty}\alpha_1^j\varepsilon_t^2\varepsilon_{t-1}^2\cdots\varepsilon_{t-j}^2$. Therefore $E(y_t^2)=\frac{\alpha_0}{1-\alpha_1}$ if $|\alpha_1|<1$. Also, $y_t=\varepsilon_t\sqrt{\alpha_0\left(1+\sum_{j=1}^{\infty}\alpha_1^j\varepsilon_{t-1}^2\cdots\varepsilon_{t-j}^2\right)}$.
- So now $E(y_t|\varepsilon_s, s\leq t-1)=0$, and therefore $E(y_t)=0$ and $\text{var}(y_t)=\frac{\alpha_0}{1-\alpha_1}$. Also, $E(y_{t+h}y_t|\varepsilon_s, s\leq t+h-1)=0$, $h>0$ and so $cov(y_{t+h}, y_t)=0$, $h>0$.
- It can be shown that $E(y_t^4)=\frac{3\alpha_0^2}{(1-\alpha_1)^2}\frac{1-\alpha_1^2}{1-3\alpha_1^2}$ if $3\alpha_1^2<1$. Therefore the kurtosis K of y_t is $K=\frac{E(y_t^4)}{(E(y_t^2))^2}=3\frac{1-\alpha_1^2}{1-3\alpha_1^2}>3$.
- Hence the solution y_t of the ARCH(1) process is strictly stationary white noise. However, y_t is not iid noise and is not Gaussian.

Remark

- With ARCH(1), y_t^2 is an AR(1) process with heteroscedastic non-Gaussian noise. Thus y_t^2 is serially correlated; this should show up on a time series plot and sample ACF for y_t^2 .
- Also note that $E(\sigma_t^2)=\frac{\alpha_0}{1-\alpha_1}=\text{var}(y_t)$. This is the “expected variance” or “long term average variance”.

Remark

1. Regression with ARCH errors: $x_t=\vec{\beta}\,'\vec{z}_t+y_t$ where $y_t\sim\text{ARCH}(1)$.
2. AR(p) with ARCH errors: $x_t=\phi_1 x_{t-1}+\cdots+\phi_p x_{t-p}+y_t$ where $y_t\sim\text{ARCH}(1)$.
3. ARCH(1) extends to ARCH(p): $y_t=\sigma_t\varepsilon_t$ where $\varepsilon_t\sim\text{iid } N(0,1)$ and $\sigma_t=\alpha_0+\alpha_1 y_{t-1}^2+\cdots+\alpha_1 y_{t-p}^2$.

GARCH Modelling

Definition: GARCH(1, 1) Model
$y_t = \sigma_t \varepsilon_t$ where $\varepsilon_t \sim \text{iid } N(0,1)$ and $\sigma_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, $\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0, \alpha_1 + \beta_1 < 1$.

Note
$\sigma_t^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_{t-1}^2 + \alpha_1 \sigma_{t-1}^2 (\varepsilon_{t-1}^2 - 1)$. Here $\varepsilon_{t-1}^2 - 1 \sim \chi^2(1)$. Let $u_t = \sigma_t^2(\varepsilon_t^2 - 1)$. Then $y_t^2 = \alpha_0 + (\alpha_1 + \beta_1) y_{t-1}^2 + u_t - \beta_1 u_t$. Hence y_t^2 is a non-Gaussian ARMA(1, 1) process with heteroscedastic noise.

Note
One can forecast volatility. Let $w = \dfrac{\alpha_0}{1 - \alpha_1 - \beta_1}$. Then σ_t^2 can be expresses as $\sigma_t^2 = w + \alpha_1 (y_t^2 - w) + \beta_1 (\sigma_{t-1}^2 - w)$.
Forecast function: If $\alpha_1 + \beta_1 < 1$, then one can show $E(\sigma_{t+j}^2 | \sigma_t^2) = w + (\alpha_1 + \beta_1)^j (\sigma_t^2 - w), j \geq 1$. Hence volatilities are either above or below their long run average w . They mean-revert geometrically to this value; the speed of mean-reversion is determined by $\alpha_1 + \beta_1$.

Note
Maximum likelihood methods can be used to estimate ARCH/GARCH parameters.

Bivariate Time Series

Example
Two time series $x_{t,1}$ and $x_{t,2}$. Two basic objectives:

1. Study relationship between the two series.
2. Build a model for the bivariate series $\vec{x}_t$.

Example
Transfer function model. $x_{t,1}$ is an input and $x_{t,2}$ is an output. Model: $x_{t,2} = \beta_0 x_{t,1} + \beta_1 x_{t-1,1} + \cdots + \beta_p x_{t-p,1} + w_t$.

Example
Vector ARMA and ARIMA models.
Vector AR model: $\begin{pmatrix} x_{t,1} \\ x_{t,2} \end{pmatrix} = \Phi_1 \begin{pmatrix} x_{t-1,1} \\ x_{t-1,2} \end{pmatrix} + \cdots \Phi_p \begin{pmatrix} x_{t-p,1} \\ x_{t-p,2} \end{pmatrix} + \begin{pmatrix} w_{t,1} \\ w_{t,2} \end{pmatrix}$, where Φ_i 's are 2×2 possibly diagonal matrices.
Now we have $E(\vec{x}_t) = \begin{pmatrix} E(x_{t,1}) \\ E(x_{t,2}) \end{pmatrix}$ and $\text{cov}(\vec{x}_{t+h}, \vec{x}_t) = \begin{bmatrix} \text{cov}(x_{t+h,1}, x_{t,1}) & \text{cov}(x_{t+h,1}, x_{t,2}) \\ \text{cov}(x_{t+h,2}, x_{t,1}) & \text{cov}(x_{t+h,2}, x_{t,2}) \end{bmatrix}$. The series $\vec{x}_t$ is weakly stationary if $E(\vec{x}_t)$ and $\text{cov}(\vec{x}_{t+h}, \vec{x}_t)$ are independent of time t ; in this case, $\Gamma(h) = \begin{bmatrix} \gamma_{1,1}(h) & \gamma_{1,2}(h) \\ \gamma_{2,1}(h) & \gamma_{2,2}(h) \end{bmatrix}$.

Continuous-Time Models

Brownian Motion
1. B_t is continuous and $B_0 = 0$.

2. $B_t \sim N(0,t)$.
3. $B_{t+s} - B_t \sim N(0,s)$ $,s \geq 0$ and non-overlapping increments are independent.

Continuous-Time AR(1)

$dx_t = (b - a\,x_t)\,dt + \sigma\,dB_t$. This is a stochastic differential equation (SDE).

The solution is given by $x_t = \mathrm{e}^{-at} x_0 + b \int_0^t \mathrm{e}^{-a(t-u)}\,du + \sigma \int_0^t \mathrm{e}^{-a(t-u)}\,dB_u$, where x_0 is independent of B_t .

Let $I_t = \int_0^t \mathrm{e}^{-a(t-u)}\,dB_u$ (Ito stochastic integral). Then it can be shown that $E(I_t) = 0$ and

$\operatorname{cov}(I_{t+h}, I_t) = \int_0^t \mathrm{e}^{2au}\,du$ $,t, h \geq 0$.

Suppose now that $a > 0$, $E(x_0) = \frac{b}{a}$, and $\operatorname{var}(x_0) = \frac{\sigma^2}{2a}$. Then $E(x_t) = \frac{b}{a}$ and $\operatorname{cov}(x_{t+h}, x_t) = \frac{\sigma^2}{2a}\mathrm{e}^{-ah}$ $,t, h \geq 0$, and hence x_t is stationary.

If $x_0 \sim N\left(\frac{b}{a}, \frac{\sigma^2}{2a}\right)$, then x_t is a strictly stationary Gaussian process.