Forecasting stationary time series.

- Given a set of observations from a time series $x_1, x_2, \cdots, x_T$, a basic problem is to predict values of x_{T+h} for some ~~h~~ integer $h \geq 1$.
- We discussed a bit about forecasting before. Recall the Holt-Winters algorithm.
- The main idea was smoothing the observations, ~~and~~ considering if ~~it ha~~ the series shows a trend and seasonality and forecast accordingly.
- In the absense of trend and seasonality ~~then~~ some other procedures can be postulated.
- Also recall that in Holt-Winter's algorithms we were required to specify the smoothing parameters, trend parameters etc.

Linear predictors.

- Let $x_1, x_2, \cdots, x_T$ is a time series of length T.
- We want to ~~estimate~~ find $\hat{x}_{T+h}$ for some integer $h \geqslant 1$.
- ~~Since the~~ Assume that $E[x_t] = \mu \quad \forall t$.
- We use a linear combination of $1, x_1, \cdots, x_T$ to predict x_{T+h}.
- So we want to find out constants $a_0, \cdots, a_T$ and declare

$$\hat{x}_{T+h} = a_0 + a_1 x_T + \cdots + a_T x_1 .$$

- To find the ~~best fit~~ values of $a_0, a_1, \cdots, a_T$ which produces the "best" predictor one minimises

$$L(a_0, \cdots, a_n) = E\left[(x_{T+h} - a_0 - a_1 x_1 - \cdots - a_T x_1)^2\right]$$

over $a_0, a_1, \cdots, a_T$.

Recall PACF, An approach that uses stationarity.

- We have seen these formulation before when we tried to find out the PACF function.

- Taking derivative ~~w.r.t. a~~ of $L(a_0, \cdots, a_T)$ w.r.t. $a_0, a_1, \cdots, a_T$ and equating them to zero we get,

$$E\left[X_{T+h} - a_0 - \sum_{i=1}^{T} a_i X_{T+h-i}\right] = 0 \quad \cdots\cdots \langle i \rangle$$

and

$$E\left[\left(X_{T+h} - a_0 - \sum_{i=1}^{T} a_i X_{T+h-i}\right) X_{T+h-j}\right] = 0, \quad j = 1, 2, \cdots, T \cdots \langle ii \rangle.$$

- Writing in matrix notation we can write.

$a_0 = \mu\left(1 - \sum_{i=1}^{T} a_i\right) \cdots\cdots$ from $\langle i \rangle$. From $\langle ii \rangle$ substitute a_0

~~and~~

$$\begin{bmatrix} \gamma_x(0) & \gamma_x(1) & \cdots & \gamma_x(T-1) \\ \gamma_x(1) & \gamma_x(0) & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_x(T-1) & \gamma_x(T-2) & \cdots & \gamma_x(0) \end{bmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_T \end{pmatrix} = \begin{bmatrix} \gamma_x(h) \\ \gamma_x(h+1) \\ \vdots \\ \gamma_x(h+T-1) \end{bmatrix} \equiv \Gamma_T a_T = \gamma_T(h).$$

Linear forecasting.

- The solution of ⟨i⟩ and ⟨ii⟩ ~~determine~~ specifies the $\hat{X}_{T+h}$ uniquely.
- ~~To see this, let~~ Once $a_0, a_1, \dots, a_T$ are determined,

$$\hat{X}_{T+h} = \mu + \sum_{i=1}^{T} a_h (X_{T+1-i} - \mu).$$

- The prediction error is given by,

  $X_{T+h} - \hat{X}_{T+h}$, which has expectation 0 and the mean squared ~~error~~ prediction error is given by,

$$E[(X_{T+h} - \hat{X}_{T+h})^2] = \gamma_x(0) - 2\sum_{i=1}^{T} a_i \gamma_x(h+1-i) + \sum_{i=1}^{T}\sum_{j=1}^{T} a_i \gamma(i-j) a_j$$

- From the equation $\Gamma_T a_T = \gamma_T(h)$ we get.

$$a_T^T \Gamma_T a_T = a_T^T \gamma_T(h) \quad \text{or} \quad \sum_{i=1}^{T}\sum_{j=1}^{T} a_i \gamma(i-j) a_j = \sum_{i=1}^{T} a_i \gamma_x(h+1-i).$$

- This implies $E[(X_{T+h} - \hat{X}_{T+h})^2] = \gamma_x(0) - \sum_{i=1}^{T} a_i \gamma_x(h+1-i)$.

Uniqueness.

- To show that the predicted value ~~of~~ $\hat{X}_{T+h}$ is unique, ~~we~~ assume that $\{a_i^{(1)}, i=0,\cdots,T\}$ and $\{a_i^{(2)}, i=0,\cdots,T\}$ are two solutions and let ~~Z be the~~ $\hat{X}^{(1)}_{T+h}$ and $\hat{X}^{(2)}_{T+h}$ ~~are~~ be the corresponding predicted values.

- Suppose $Z = \hat{X}^{(1)}_{T+h} - \hat{X}^{(2)}_{T+h} = (a_0^{(1)} - a_0^{(2)}) + \sum_{j=1}^{T} (a_j^{(1)} - a_j^{(2)}) X_{T+1-j}$.

- Then $Z^2 = Z\left(a_0^{(1)} - a_0^{(2)} + \sum_{j=1}^{T} (a_j^{(1)} - a_j^{(2)}) X_{T+1-j}\right)$.

- From ⟨i⟩ and ⟨ii⟩ it follows that $E[Z]=0$ and $E[Z X_{T+1-j}]=0$ for $j=1,2,\cdots,T$.

- This implies $E[Z^2]=0$, since $Z \geq 0$, this implies that $Z=0$.

Example.

- One-step prediction of an AR(1) process: Consider the AR(1) process

$$X_t = \phi X_{t-1} + Z_t, \quad t = 0, \pm 1, \cdots,$$

where $|\phi| < 1$ and $\{Z_t\} \sim WN(0, \sigma^2)$.

- The best linear predictor of X_{T+1} in terms of $1, X_T, \cdots, X_1$ is given by.

$$\hat{X}_{T+1} = a_0 + \sum_{i=1}^{T} a_i X_{T+1-i}$$, where the coefficients $a_0, a_0, \cdots a_T$ is given by.

$$\begin{bmatrix} 1 & \phi & \cdots & \phi^{T-1} \\ \phi & 1 & & \\ \vdots & & \ddots & \\ \phi^{T-1} & & & 1 \end{bmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_T \end{pmatrix} = \begin{bmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^T \end{bmatrix}.$$

- Clearly $a = (\phi, 0, \cdots, 0)$ is a solution of the equation.
- So $\hat{X}_{T+1} = \phi X_T$. The mean squared error is given by, $\gamma_x(0) - \phi \gamma_x(1) = \sigma^2$.

Example: AR(1) process with non zero mean.

- If $\{Y_t\}$ is said to be an AR(1) process with mean μ if the process $X_t = Y_t - \mu$ is a zero mean AR(1) process.
- We see that $Y_t - \mu = \phi(Y_{t-1} - \mu) + Z_t$.
- The best linear predictor of Y_{T+1}, based on $1, Y_T, \cdots, Y_1$ is given by, $(\hat{Y}_{T+1} - \mu) = \phi(Y_T - \mu)$.
- So $\hat{Y}_{T+1} = \mu + \phi(Y_T - \mu)$.
- It can be shown that ~~for a~~ $\hat{Y}_{T+h} = \mu + \phi^h (Y_T - \mu)$.
- Also it can be shown that

$$E[(Y_{T+h} - \hat{Y}_{T+h})^2] = \gamma_x(0)\left[1 - \sum_{i=1}^{T} a_i \rho_T(h+1-i)\right]$$

$$= \frac{\sigma^2(1-\phi^{2h})}{1-\phi^2}.$$

Solving the equations to determine a_1.

- Note that we need to solve the equation,

$$\Gamma_T a_T = \gamma_T(h).$$

- ~~The matrix Γ_T~~ Clearly the solution is given by $a_T = \Gamma_T^{-1} \gamma_T(h)$.
- For this obviously one needs to ensure Γ_T is invertible.
- One sufficient condition for Γ_T to be ~~is~~ invertible is that $\gamma_x(0) > 0$ and $\gamma_x(h) \longrightarrow 0$ as $h \longrightarrow \infty$.

- One important problem ~~in~~ is to invert Γ_T.
- This may be difficult, because Γ_T is a ~~large matrix~~ $T \times T$ matrix, which may be quite big.
- However, note that Γ_T has a structure which can be exploited.

Γ_T matrix.

- Note that the Γ_T matrix is given by:

$$\Gamma_T = \begin{bmatrix} \gamma_x(0) & \gamma_x(1) & \gamma_x(2) & \cdots & \gamma_x(T-1) \\ \gamma_x(1) & \gamma_x(0) & \gamma_x(1) & & \vdots \\ \gamma_x(2) & \gamma_x(1) & \ddots & \ddots & \\ \vdots & & \ddots & \ddots & \\ \gamma_x(T-2) & & & \ddots & \gamma_x(1) \\ \gamma_x(T-1) & \gamma_x(T-2) & \cdots & \gamma_x(1) & \gamma_x(0) \end{bmatrix}.$$

- These sort of matrices are called Toeplitz matrices.
- There are recursive fast algorithms to find inverses of such matrices and solve ~~equation~~ system of equations $\Gamma_T a = \gamma_T$.
- For more details see Matrix computation by Golub and Van Loan.

The Durbin – Levinson algorithm.

- If $\{x_t\}$ is a zero-mean stationary series with autocovariance function γ, then the equations ⟨i⟩ and ⟨ii⟩ in principle completely solve the problem of determining the best linear predictor $\hat{x}_{T+h}$ in terms of $1, x_1, \cdots, x_T$.

- Consider the linear prediction problem to predict $\hat{x}_{T+1}$ based on $1, x_T, \cdots, x_1$.

- We know

$$\hat{x}_{T+1} = \phi_{T1} x_T + \cdots + \phi_{TT} x_1 .$$

where $\phi_T = (\phi_{T1}, \cdots, \phi_{TT})^T = \Gamma_T^{-1} \gamma_T$ and

$$\gamma_T = (\gamma_x(1), \cdots, \gamma_x(T))^T .$$

The algorithm.

- The algorithm proceeds as follows.

- step 0 : $i=0$, $v_0 = \gamma_x(0)$.
- step 1 : $i=1$, $\phi_{11} = \gamma_x(1)/\gamma_x(0)$, $v_1 = v_0(1-\phi_{11}^2)$.
- step 2 : $i=2$, $\phi_{22} = [\gamma_x(2) - \phi_{11}\gamma_x(1)]/v_1$.

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11}, \quad v_2 = v_1(1-\phi_{22}^2).$$

step 3 : $i=3$, $\phi_{33} = [\gamma_x(3) - \phi_{21}\gamma_x(2) - \phi_{22}\gamma_x(1)]/v_2$.

$$\begin{bmatrix}\phi_{31}\\ \phi_{32}\end{bmatrix} = \begin{bmatrix}\phi_{21}\\ \phi_{22}\end{bmatrix} - \phi_{33}\begin{bmatrix}\phi_{22}\\ \phi_{21}\end{bmatrix}, \quad v_3 = v_2(1-\phi_{33}^2).$$

step i :

$$\phi_{ii} = \left[\gamma_x(i) - \sum_{j=1}^{i-1}\phi_{(i-1)j}\,\gamma_x(i-j)\right]/v_{i-1}.$$

$$\begin{bmatrix}\phi_{i1}\\ \vdots\\ \phi_{i(i-1)}\end{bmatrix} = \begin{bmatrix}\phi_{(i-1)1}\\ \vdots\\ \phi_{(i-1)(i-1)}\end{bmatrix} - \phi_{ii}\begin{bmatrix}\phi_{(i-1)(i-1)}\\ \vdots\\ \phi_{(i-1)1}\end{bmatrix}, \quad v_i = v_{i-1}(1-\phi_{ii}^2).$$

Recursion.

- Note that we can write.

$$\hat{X}_2 = \phi_{11} X_1 .$$
$$\hat{X}_3 = \phi_{21} X_2 + \phi_{22} X_1 .$$
$$\hat{X}_4 = \phi_{31} X_3 + \phi_{32} X_2 + \phi_{33} X_1 .$$

and so on.

- Notice that in calculating ϕ_{21}, ϕ_{22} we use ϕ_{11}, ~~ϕ_{31}~~ while calculating $\phi_{31}, \phi_{32}, \phi_{33}$ use ϕ_{21}, ϕ_{22}. In general to calculate $\phi_{i1}, \cdots, \phi_{ii}$ we use $\phi_{(i-1)1}, \cdots, \phi_{(i-1)(i-1)}$ and so on.
- This is what is ~~recurs~~ meant by recursion.
- The Durbin–Levinson algorithm uses this recursion, which allows the inverse to be calculated in ~~roughly~~ $2T^2$ to $4T^2$ flops.

The Innovations Algorithm

- The recursive Innovations algrithm is applicable to all series with finite second moments, whether they are stationary or not.
  - For this algorithm we assume that $\{X_t\}$ has zero mean with $E[X_t^2] < \infty$ for each t and
$$E[X_i X_j] = \kappa(i,j).$$
  - Also we define the best one step predictors and their mean squared errors as:
$$\hat{X}_n = \begin{cases} 0 & \text{if } n = 1. \\ P_{n-1} X_n & \text{if } n = 2, 3, \cdots . \end{cases}$$

Here $P_{n-1} X_n$ is the prediction of X_n based on previous $n-1$ values.
ie $P_{n-1} X_n = a_1 X_1 + \cdots + a_{n-1} X_{n-1}$.

Also $\quad v_n = E\left[(X_{n+1} - P_n X_{n+1})^2 \right].$

Innovations.

- Define the innovations as

$$u_n = x_n - P_{n-1} x_n = x_n - \hat{x}_n.$$

- Note that we can write

$$\begin{bmatrix} u_1 \\ \vdots \\ \vdots \\ u_T \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & & 0 \\ a_{11} & 1 & & & 0 \\ a_{22} & a_{21} & 1 & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ a_{T-1,T-1} & a_{T-1,T-2} & \cdots & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ x_T \end{bmatrix} \Rightarrow u_T = A_T \underline{x}_T.$$

- If $\{X_t\}$ is stationary $a_{ij} = -a_j$.

- Clearly A_T is non singular and suppose $C_T = A_T^{-1}$ ie

$$\underline{x}_T = C_T u_T.$$

- Now by construction

$$\hat{\underline{x}}_T = \underline{x}_T - u_T = C_T u_T - u_T = (C_T - I_T) u_T$$
$$= \Theta_T (\underline{x}_T - \hat{\underline{x}}_T) \quad \cdots\cdots \text{(iii)}$$

Innovations Algorithm.

- Here

$$\Theta_T = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ \theta_{11} & 0 & & & \\ \theta_{22} & \theta_{21} & \ddots & & \\ \vdots & & & \ddots & \\ \theta_{T-1,T-1} & \theta_{T-1,T-2} & \cdots & \cdots & 0 \end{bmatrix}.$$

- Note that $\hat{X}_n$ satisfies an recursive equation.

$$\hat{\underline{X}}_T = C_T(\underline{X}_T - \hat{\underline{X}}_T),$$

where $C_T = \Theta_T + I_T$.

- Clearly from ⟨iii⟩ we get.

$$\hat{X}_{n+1} = \begin{cases} 0 & \text{if } n=0 \\ \sum_{j=1}^{n} \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}) & \text{if } n = 1, 2, \cdots \end{cases}.$$

- Thus the one step predictors may be computed if we can get Θ_T.

Innovations Algorithm.

- The steps are as follows:
- Step 0: $v_0 = \kappa(1,1)$.
- Step 1: $n=1$, $k=0$,

$$\theta_{1,1} = \kappa(2,1)/v_0 . \qquad v_1 = \kappa(2,2) - \theta_{11}^2 v_0 .$$

- Step 2: $n=2$, $k=0$,

$$\theta_{2,2} = \kappa(3,1)/v_0$$

$n=2$, $k=1$

$$\theta_{21} = \left(\kappa(3,2) - \theta_{11}\theta_{21}v_0\right)/v_1 .$$

$$v_2 = \kappa(3,3) - \theta_{22}^2 v_0 - \theta_{21}^2 v_1 .$$

- Step n:

$$\theta_{n,n-k} = \left(\kappa(n+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j} v_j\right)/v_k , \quad 0 \le k < n .$$

$$v_n = \kappa(n+1,n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j .$$

- We successively calculate v_0; θ_{11}, v_1; θ_{22}, θ_{21}; v_2; $\theta_{33}, \theta_{32}, \theta_{31}, v_3$; ...

Comparison

- The Durbin-Levinson recursion gives the coefficients of $X_n, \cdots, X_1$ in the representation $\hat{X}_{n+1} = \sum_{j=1}^{n} \phi_{nj} X_{n+1-j}$.
- The innovations algorithm gives the coefficients of $(X_n - \hat{X}_n), \cdots, (X_1 - \hat{X}_1)$, in the expansion

$$\hat{X}_{n+1} = \sum_{j=1}^{n} \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}).$$

- The innovations algorithm has several advantages.
- The innovations are uncorrelated.
- For ARMA(p,q) it is sometimes better to use innovations algorithm. We can write.

$$X_{n+1} = X_{n+1} - \hat{X}_{n+1} + \hat{X}_{n+1} = \sum_{j=0}^{n} \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), \quad n = 0, 1, 2, \cdots$$

- Here we define $\theta_{n0} = 1$.

Example: MA(1)

- ~~Assum~~ Let $\{X_t\}$ is the time series defined by

$$X_t = Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim WN(0,\sigma^2),$$

then $\kappa(i,j) = 0$ for $|i-j| > 1$, $\kappa(i,i) = \sigma^2(1+\theta^2)$ and $\kappa(i,i+1) = \theta\sigma^2$.

- In this case the innovations lead to the recursion.

$$\theta_{nj} = 0, \quad 2 \le j \le n.$$

$$\theta_{n1} = \theta\sigma^2/v_{n-1}, \quad v_0 = (1+\theta^2)\sigma^2,$$

and $$v_n = \left(1 + \theta^2 - \frac{\theta^2\sigma^2}{v_{n-1}}\right)\sigma^2.$$

More Specific Example.

- Consider the case

$$X_t = Z_t - .9\, Z_{t-1}, \quad \{Z_t\} \sim WN(0,1).$$

- The innovations algorithm would give:

$v_0 = 1.8100$

$\theta_{11} = -.4972 \quad v_1 = 1.3625$

$\theta_{21} = -.6606 \quad \theta_{22} = 0 \quad v_2 = 1.2155$

$\theta_{31} = -.7404 \quad \theta_{32} = 0 \quad \theta_{33} = 0 \quad v_3 = 1.1436$

$\theta_{41} = -.7870 \quad \theta_{42} = 0 \quad \theta_{43} = 0 \quad \theta_{44} = 0 \quad v_4 = 1.1017.$

- On the other hand the Durbin-Levinson algorithm would produce

$v_0 = 1.8100$

$\phi_{11} = -.4972 \quad v_1 = 1.3625$

$\phi_{21} = -.6606 \quad \phi_{22} = -.3285 \quad v_2 = 1.2155$

$\phi_{31} = -.7404 \quad \phi_{32} = -.4892 \quad \phi_{33} = -.2433 \quad v_3 = 1.1436$

$\phi_{41} = -.7870 \quad \phi_{42} = -.5828 \quad \phi_{43} = -.3850 \quad \phi_{44} = -.1914 \quad v_4 = 1.1017.$

ARMA(1,1)

- Consider an ARMA(1,1) process with

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim WN(0, \sigma^2).$$

with $|\phi| < 1$

- It can be shown that the recursion equations are given by.

$$\hat{X}_{n+1} = \phi X_n + \theta_{n1} (X_n - \hat{X}_n), \qquad n \geq 1.$$

- Also note that $\gamma_X(0) = \sigma^2(1 + 2\theta\phi + \theta^2)/(1 - \phi^2)$.
- Further for $i, j \geq 1$

$$\kappa(i,j) = \begin{cases} (1+2\theta\phi+\theta^2)/(1-\phi^2), & i,j = 1, \\ 1+\theta^2, & i = j \geq 2, \\ \theta, & |i-j| = 1, i \geq 1, \\ 0 & \text{o.w.} \end{cases}$$

- It turns out that the innovations would look like.

$$v_0 = (1+2\theta\phi+\theta^2)/(1-\phi^2), \quad \theta_{n1} = \theta/r_{n-1}, \quad v_n = 1 + \theta - \theta^2/v_{n-1}.$$

- This can be solved explicitly.