5.1 Preliminary Estimation

5.1.2 Burg’s Algorithm
It is a modification of the sample ACF based Durbin-Levinson algorithm for $AR(p)$ model. The difference between the two is that in Burg’s algorithm, only the $\phi_{pj}$, $1 \leq j \leq p - 1$ coefficients are calculated through the D-L formula while all the $\phi_{ii}$ are calculated using forward and backward prediction errors. By definition, using observations $x_{n+1-t}, x_{n+2-t}, \cdots, x_{n+i+1-t}, t \geq i + 1, i \geq 0$

$$u_i(t) = x_{n+1+i-t} - P(x_{n+1+i-t}|x_{n+1-t}, \cdots, x_{n+i-t}), \text{ the forward prediction error}$$

$$v_i(t) = x_{n+1-t} - P(x_{n+1-t}|x_{n+2-t}, \cdots, x_{n+i+1-t}), \text{ the backward prediction error}$$

These seemingly complicated sequences of numbers satisfy recursive equations

$$u_0(t) = v_0(t) = x_{n+1-t}$$

$$u_i(t) = u_{i-1}(t - 1) - \phi_{ii}v_{i-1}(t), i \geq 1$$

$$v_i(t) = v_{i-1}(t) - \phi_{ii}u_{i-1}(t - 1), i \geq 1$$

So, they can be calculated recursively and then minimizing $\sum_{t=i+1}^n u_i^2(t) + v_i^2(t)$ for $i = 1, 2, \cdots, p$ to solve for $\phi_{11}, \cdots, \phi_{pp}$ one at a time and the corresponding minimum values are $\sigma_i^{(B)}$. The Burg estimation of the white noise variance is $\sigma_p^{(B)}$. For large samples, this modification causes no change in the distribution of estimators. See the two data examples 5.1.3 and 5.1.4 where Yule-Walker and Bug procedures are compared.
5.1.3 The Innovation Algorithm

Let \( \{X_t\} \) be a mean zero stationary process and assume that \( \hat{\gamma}(0) > 0 \), can one impose an \( MA(m) \) structure

\[
X_t = Z_t + \hat{\theta}_{m1}Z_{t-1} + \cdots + \hat{\theta}_{mm}Z_{t-m}, \quad Z_t \sim WN(0, \hat{v}_m)
\]

via the innovation algorithm?

**Definition** If \( \hat{\gamma}(0) > 0 \), then for \( m = 1, \cdots, n-1 \) define

\[
\hat{v}_0 = \hat{\gamma}(0)
\]

\[
\hat{\theta}_{m,m-k} = \hat{v}_k^{-1} \left[ \hat{\gamma}(m-k) - \sum_{j=0}^{k-1} \hat{\theta}_{m,m-j}\hat{\theta}_{k,k-j}\hat{v}_j \right], \quad k = 0, \cdots, m-1
\]

\[
\hat{v}_m = \hat{\gamma}(0) - \sum_{j=0}^{m-1} \hat{\theta}_{m,m-j}\hat{v}_j
\]

contrast this with

\[
v_0 = \kappa(1,1)
\]

\[
\theta_{m,m-k} = v_k^{-1} \left[ \kappa(m+1,k+1) - \sum_{j=0}^{k-1} \theta_{m,m-j}\theta_{k,k-j}v_j \right], \quad k = 0, \cdots, m-1
\]

\[
v_m = \kappa(m+1,m+1) - \sum_{j=0}^{m-1} \theta_{m,m-j}^2v_j
\]

**Theorem** If \( \{X_t\} \) is a causal and invertible \( ARMA(p,q) \) process \( \phi(B)X_t = \theta(B)Z_t \) with \( Z_t \sim IID(0,\sigma^2) \), \( E(Z_t^4) < \infty \), and write \( \psi(z) = \theta(z)/\phi(z) = \sum_{j=0}^{\infty} \psi_jz^j \), \( |z| \leq 1 \), then for any sequence \( m(n) \) such that \( m < n, m = o(n^{1/3}) \), \( m \to \infty \) as \( n \to \infty \), we have for each fixed \( k \)

\[
\sqrt{n} \left( \hat{\theta}_{m1} - \theta_1, \cdots, \hat{\theta}_{mk} - \theta_k \right) \Rightarrow N(0, A), \quad \hat{v}_m \to \sigma^2
\]

where \( A = [a_{ij}]_{i,j=1}^{k} \)

\[
a_{ij} = \min(i,j) \sum_{r=1}^{\min(i,j)} \psi_i^{-r}\psi_j^{-r}
\]

In particular, for an invertible \( MA(q) \) process, \( X_t = \theta(B)Z_t \) with \( X_t \sim IID(0,\sigma^2) \), \( E(Z_t^4) < \infty \), one has

\[
\sqrt{n} \left( \hat{\theta}_{m1} - \theta_1, \cdots, \hat{\theta}_{mq} - \theta_q \right) \Rightarrow N(0, A), \quad \hat{v}_m \to \sigma^2
\]

where \( A = [a_{ij}]_{i,j=1}^{q} \)

\[
a_{ij} = \min(i,j) \sum_{r=1}^{\min(i,j)} \theta_i^{-r}\theta_j^{-r}
\]
For IID white noise invertible $MA(q)$ process, the $(1 - \alpha)$ asymptotic confidence interval for $\theta_j$ is

$$
\left\{ \theta \in R : \left| \theta - \hat{\theta}_{mj} \right| \leq z_{1-\alpha/2} n^{-1/2} \left( \sum_{k=0}^{j-1} \hat{\theta}_{mk}^2 \right)^{1/2} \right\}
$$

Order Selection:

- Use as a preliminary estimator of $q$ the smallest value $m$ such that $\hat{\rho}(k)$ is not significantly different from zero for all $k > m$, i.e. $|\hat{\rho}(k)| < 1.96 n^{-1/2} \sqrt{1 + 2\hat{\rho}^2(1) + \cdots + 2\hat{\rho}^2(m)}$ for $k > m$ by Bartlett’s formula or in practice, $|\hat{\rho}(k)| < 1.96 n^{-1/2}$ for $k > m$.

- Estimate the order $q$ of the model to be fitted as the largest lag $j$ for which the ratio of $\hat{\theta}_{mj}$ to $1.96$ times its approximate standard deviation is larger than $1$ in absolute value. A default value of $m$ is set by the program, but it may be altered manually.

- A more systematic approach to order selection is to find the values of $q$ and $\hat{\theta}_q = (\hat{\theta}_{m1}, \cdots, \hat{\theta}_{mq})'$ that minimize the AICC statistic.

Example 5.1.5 The Dow Jones Utilities Index
The innovation algorithm when $p > 0$ and $q > 0$

Let $\{X_t\}$ be a causal $ARMA(p,q)$ process $\phi(B)X_t = \theta(B)Z_t$ with $Z_t \sim WN(0,\sigma^2)$, then

$$\psi(z) = \theta(z)/\phi(z) = \sum_{j=0}^{\infty} \psi_j z^j, \quad |z| \leq 1,$$

and

$$\psi_0 = 1,$$

$$\psi_j = \theta_j + \sum_{i=1}^{\min\{p,j\}} \phi_i \psi_{j-i}, \quad j = 1, 2, \ldots$$

to estimate the total $p + q$ unknown parameters, we need $p + q$ equations to solve. The parameters $\psi_j, j = 1, \ldots, p + q$ can be estimated via innovation algorithm, by $\hat{\theta}_{mj}, j = 1, \ldots, p + q$. Then the equations are

$$\hat{\theta}_{mi} = \hat{\theta}_j + \sum_{i=1}^{\min\{p,j\}} \hat{\phi}_i \hat{\theta}_{m,j-i}, \quad j = 1, 2, \ldots, p + q, \quad m \geq p + q.$$ 

We estimate $\sigma^2$ by $\hat{v}_m$.

**Order selection** for $ARMA(p,q)$ model is done by fitting $ARMA(p,q)$ models with various $p,q$, and calculate the AICC values for each model. The pari $(p,q)$ with the smallest AICC is selected.

**Example 5.1.6** The lake data