5.1 Preliminary Estimation

5.1.4 The Hannan-Rissanen Algorithm

The first two steps are implemented in ITSM while the last step stems from the Maximum Likelihood Estimator (MLE) that we will discuss next.

Step 1. Fit a higher order AR(m) model \((m > p + q)\) to obtain

\[ \hat{Z}_t = X_t - \hat{\phi}_{m1}X_{t-1} - \cdots - \hat{\phi}_{mm}X_{t-m}, \quad t = m + 1, \ldots, n. \]

Step 2. The model parameter \(\beta = (\phi', \theta')'\) is estimated by least square linear regression of \(X_t\) onto \((X_{t-1}, \ldots, X_{t-p}, \hat{Z}_{t-1}, \ldots, \hat{Z}_{t-q})\), \(t = m + 1 + q, \ldots, n\). And \(\sigma^2\) is estimated by

\[ \hat{\sigma}^2_{HR} = \frac{SSE(\hat{\beta})}{n - m - q}. \]

In ITSM, we eliminate Step 3 and use the model produced by the first two steps as the initial model for the calculation (by numerical maximization) of the maximum likelihood estimator itself.

Example 5.1.7 The lake data
5.2 Maximum Likelihood Estimator

let \{X_t\} be a mean zero Gaussian process and assume that \(\Gamma_n = E(\overline{X}_n \overline{X}'_n)\) is non-singular, where \(\overline{X}_n = (X_1, \ldots, X_n)'\). The likelihood of \(\overline{X}_n\) is

\[
L(\Gamma_n) = (2\pi)^{-n/2}(\det \Gamma_n)^{-1/2} \exp \left( -\frac{1}{2} \overline{X}_n' \Gamma_n^{-1} \overline{X}_n \right).
\]

Recall

\[
\hat{X}_{i+1} = \sum_{j=1}^{i} \theta_{ij} (X_{i+1-j} - \hat{X}_{i+1-j}) = \sum_{j=0}^{n-1} \theta_{i,i-j} (X_{j+1} - \hat{X}_{j+1}) - (X_{i+1} - \hat{X}_{i+1})
\]

where one defines \(\theta_{i,0} = 1\) and \(\theta_{ij} = 0\) for \(j < 0\). So one can write

\[
X_{i+1} = \sum_{j=1}^{n} \theta_{i,i+1-j} (X_{j} - \hat{X}_{j})
\]

or for matrix \(C_n = [\theta_{i,i-j}]_{i,j=0}^{n-1}\)

\[
\overline{X}_n = C_n (\overline{X}_n - \hat{X}_n)
\]

Hence \(\Gamma_n = C_n D_n C'_n\) where \(D_n = E(\overline{X}_n - \hat{X}_n) (\overline{X}_n - \hat{X}_n)' = \text{diag}(v_0, \ldots, v_{n-1})\) and

\[
\overline{X}'_n \Gamma_n^{-1} \overline{X}_n = (\overline{X}_n - \hat{X}_n)' D_n^{-1} (\overline{X}_n - \hat{X}_n) = \sum_{j=1}^{n} (X_j - \hat{X}_j)^2 / v_{j-1}
\]

and

\[
\det \Gamma_n = (\det C_n)^2 \det D = v_0 \cdots v_{n-1}.
\]

So,

\[
L(\Gamma_n) = (2\pi)^{-n/2}(v_0 \cdots v_{n-1})^{-1/2} \exp \left( -\frac{1}{2} \sum_{j=1}^{n} (X_j - \hat{X}_j)^2 / v_{j-1} \right)
\]

where all the ingredients are calculated according to

\[
v_0 = \gamma(0)
\]

\[
\theta_{n,n-k} = v_k^{-1} \left\{ \gamma(n-k) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \right\}
\]

\[
v_n = \gamma(0) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j
\]

Note: Even if \(\{X_t\}\) is not Gaussian, We can still use \(L(\Gamma_n)\) above as a measurement of goodness of fit of the model to the data, its maximum points is referred as "Maximum likelihood" estimators. A Justification is that the large sample distribution of the estimators is the same for \(\{Z_t\} \sim IID(0, \sigma^2)\), regardless of whether or not \(\{Z_t\}\) is Gaussian.
If \( \{X_t\} \) is a causal ARMA\((p,q)\) process \( \phi(B)X_t = \theta(B)Z_t \) with \( Z_t \sim WN(0,\sigma^2) \). Through transferring \( \{X_t\} \) into \( \{W_t\} \), we have

\[
\hat{X}_{i+1} = \begin{cases} 
\sum_{j=1}^{i} \theta_{i,j} \left(X_{i+1-j} - \hat{X}_{i+1-j}\right) & 1 \leq i < m = \max(p,q) \\
\sum_{r=1}^{p} \phi_r X_{i+1-r} + \sum_{j=1}^{q} \theta_{ij} \left(X_{i+1-j} - \hat{X}_{i+1-j}\right) & i \geq m
\end{cases}
\]

and also

\[v_i = E \left( X_{i+1} - \hat{X}_{i+1} \right)^2 = \sigma^2 r_i, \quad i = 0, 1, 2, \ldots\]

One obtains

\[
L(\phi, \theta, \sigma^2) = \left(2\pi\sigma^2\right)^{-n/2}(r_0 \cdots r_{n-1})^{-1/2} \exp \left(-\frac{1}{2} \sigma^{-2} \sum_{j=1}^{n} \left(X_j - \hat{X}_j\right)^2 / r_{j-1}\right)
\]

If one fix the parameter \( \phi, \theta \) and maximizes over \( \sigma^2 \) one finds

\[\hat{\sigma}^2 = n^{-1} S(\hat{\phi}, \hat{\theta}), \quad S(\hat{\phi}, \hat{\theta}) = \sum_{j=1}^{n} (X_j - \hat{X}_j)^2 / r_{j-1}\]

and the estimators \( \hat{\phi}, \hat{\theta} \) are obtained by minimizing the reduced likelihood

\[l(\phi, \theta) = \ln \{n^{-1} S(\phi, \theta)\} + n^{-1} \sum_{j=1}^{n} \ln r_{j-1}\]

then one has the MLE \( \hat{\phi}, \hat{\theta}, \hat{\sigma}^2 \).

\textbf{Note:} Minimization of \( l(\phi, \theta) \) must be done numerically. Initial values for \( \phi \) and \( \theta \) can be obtained by using the preliminary estimators. Then it searches systematically for the minimum points.

**Least Square Estimation for Mixed Models**

Another estimation method is teh least square method, by minimizing \( S(\phi, \theta) \), which is equivalent to MLE if \( n^{-1} \sum_{j=1}^{n} \ln r_{j-1} \to 0 \) as \( n \to \infty \). This is the case when teh parameter \( \theta \) is required to be invertable. And then one estimates

\[\hat{\sigma}^2_{LS} = (n - p - q)^{-1} S(\bar{\phi}, \bar{\theta}).\]

Write \( \beta' = (\phi', \theta') \). If one restricts the MLE \( \hat{\beta} \) and the LSE \( \bar{\beta} \) to be causal and invertible, then they are equivalent.
Confidence Regions for the Coefficients

If vectors $\phi$ and $\theta$ have no common zeros, and $Z_t \sim IID(0, \sigma^2)$ then

$$n^{1/2}(\hat{\beta} - \beta) \Rightarrow N(0, V(\beta))$$

with

$$V(\beta) = \sigma^2 \begin{bmatrix} E(\vec{U}_t \vec{U}'_t) & E(\vec{U}_t \vec{V}'_t) \\ E(\vec{V}_t \vec{U}'_t) & E(\vec{V}_t \vec{V}'_t) \end{bmatrix}^{-1}$$

with $\phi(B)U_t = Z_t$ and $\theta(B)V_t = Z_t$.

Example 5.2.1-5.2.3
Order Selection

We use the AICC Criterion:
Choose $p$, $q$, $\phi_p$, and $\theta_q$ to minimize

$$\text{AICC} = -2 \ln L(\phi_p, \theta_q, S(\phi_p, \theta_q)/n) + 2(p + q + 1)n/(n - p - q - 2).$$

It can also be done through "Autofit" to obtain the order $p$ and $q$. Then the obtained model should be checked by preliminary estimation followed by maximum likelihood estimation.

Example 5.2.4, 5.2.5

Computer Project #3: Exercise 5.1 and 5.5 (Due April 6, Tuesday)