Lecture 15

Yule-Walker Equations
Textbook Sections: 2.5

A good portion of the lecture was spent going over the steps in fitting an $ARMA(p, q)$ model, and continuing the PACF discussion from last lecture.

Below is an excerpt from Dr. Burman's handout on the Yule-Walker equations.

Note: This is a more technical topic, and won't be covered heavily in this course. I expect you to know what the Yule-Walker equations are, and how they are used, but you won't be heavily tested on the details.

Estimation of parameters in AR.

Let us begin the discussion with the AR(1) case when the observations $X_1, \ldots, X_n$ are available. The AR(1) series can be expressed as $X_t = \phi_0 + \phi_1 X_{t-1} + \varepsilon_t$ with $\phi_0 = (1 - \phi_1)\mu$. Note that X_t is the dependent variable, and X_{t-1} is the independent variable. The method of least squares minimizes the sum of squares $\sum(X_t - f_0 - f_1 X_{t-1})^2$ with respect to f_0 and f_1 leading to the usual normal equations. Estimates of ϕ_0 and ϕ_1 are obtained by solving the normal equations.

Let us now examine the range of t in the summation. We cannot have $t = 1$ in the summation since X_0 is not available. Similarly, we cannot use the independent variable X_n (ie, when $t - 1 = n$) since X_{n+1} is not available. Thus in order to follow the least squares method, we need to minimize $\sum_{t=2}^{n}(X_t - f_0 - f_1 X_{t-1})^2$ with respect to f_0 and f_1. However, in order to use the entire data, we would like to minimize

$$(X_1 - f_0 - f_1 X_0)^2 + (X_2 - f_0 - f_1 X_1)^2 + \cdots + (X_n - f_0 - f_1 X_{n-1})^2 + (X_{n+1} - f_0 - f_1 X_n)^2$$

with respect to f_0 and f_1 since, in this case, we would use each observation both as independent and dependent variables.

The preceding discussion leads to two methods: least squares and modified least squares.

The Least squares.
Minimizing $\sum_{t=2}^{n}(X_t - f_0 - f_1 X_{t-1})^2$ with respect to f_0 and f_1 leads to the normal equations

$$\begin{aligned}(n-1)f_0 + \left(\sum_{t=2}^{n} X_{t-1}\right) f_1 &= \sum_{t=2}^{n} X_t, \\ \left(\sum_{t=2}^{n} X_{t-1}\right) f_0 + \left(\sum_{t=2}^{n} X_{t-1}^2\right) f_1 &= \sum_{t=2}^{n} X_{t-1}X_t.\end{aligned}$$

The solutions $\hat{\phi}_0$ and $\hat{\phi}_1$ of these equations are the least squares estimates of ϕ_0 and ϕ_1. It can be verified that in general $\hat{\phi}_0 \neq (1-\hat{\phi}_1)\bar{X}$ and $\hat{\mu} \neq \bar{X}$. We know that we must have $|\phi_1| < 1$ for an AR(1) series to be stationary. Unfortunately, the least squares estimate $\hat{\phi}_1$ is not guaranteed to be smaller than 1 in magnitude. The modified least squares estimation method remedies this defect.

In the AR(p) case, the series can be expressed as

$$\begin{aligned}X_t &= \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t \text{ with} \\ \phi_0 &= (1 - \phi_1 - \cdots - \phi_p)\mu.\end{aligned}$$

The method of least squares minimizes $\sum_{t=p+1}^{n}(X_t - f_0 - f_1 X_{t-1} - \cdots - f_p X_{t-p})^2$ with respect to $f_0, f_1, \ldots, f_p$ leading to the normal equations. Solutions of these equations $\hat{\phi}_0, \hat{\phi}_1, \ldots, \hat{\phi}_p$ are estimates of the AR(p) parameters $\phi_0, \phi_1, \ldots, \phi_p$. Once again, in general $\hat{\phi}_0 \neq (1-\hat{\phi}_1)\bar{X}$ and $\hat{\mu} \neq \bar{X}$. Similar to the AR(1) case, the estimated parameters $\hat{\phi}_1, \ldots, \hat{\phi}_p$ may not satisfy the conditions for stationarity.

The Modified least squares (Yule-Walker).

Once again, let us begin with the AR(1) case. If we approximate X_0 and X_{n+1} by the sample mean $\bar{X} = (X_1 + \cdots + X_n)/n$, then we can run the least squares method for the padded data $\{\bar{X}, X_1, \ldots, X_n, \bar{X}\}$, i.e., minimize $\sum_{t=1}^{n+1}(\tilde{X}_t - f_0 - f_1\tilde{X}_{t-1})^2$ with respect to f_0 and f_1, where $\tilde{X}_0 = \bar{X}, \tilde{X}_1 = X_1, \ldots, \tilde{X}_n = X_n, \tilde{X}_{n+1} = \bar{X}$. The normal equations can be simplified to

$$\begin{aligned} f_0 &= (1 - f_1)\bar{X}, \\ \hat{\gamma}(0)f_1 &= \hat{\gamma}(1). \end{aligned}$$

The second equation $\hat{\gamma}(0)f_1 = \hat{\gamma}(1)$ is known as the Yule-Walker equation. Note that the solutions are:

$$\hat{\phi}_1 = \hat{\gamma}(1)/\hat{\gamma}(0) \text{ and } \hat{\phi}_0 = (1 - \hat{\phi}_1)\bar{X},$$

and the estimate of μ is $\bar{X}$.

For the AR(p) case, we can pad the data with with $2p$ $\bar{X}$ values: p values at the beginning and p values at the end, ie, the padded data are now of the form $\{\bar{X}, \ldots, \bar{X}, X_1, \ldots, X_n, \bar{X}, \ldots, \bar{X}\}$. Let

$$\tilde{X}_t = \begin{cases} \bar{X} & t = 1-p, \ldots, 0 \\ X_t & t = 1, \ldots, n \\ \bar{X} & t = n+1, \ldots, n+p. \end{cases}.$$

If we minimize $\sum_{t=1}^{n+p}(\tilde{X}_t - f_0 - f_1\tilde{X}_{t-1} - \cdots - f_p\tilde{X}_{t-p})^2$ with respect to $f_0, \ldots, f_p$, it can be shown that the last p of the normal equations (there are $p+1$ equations in total) are

$$\begin{aligned} \hat{\gamma}(0)f_1 + \cdots + \hat{\gamma}(p-1)f_p &= \hat{\gamma}(1) \\ \hat{\gamma}(1)f_1 + \cdots + \hat{\gamma}(p-2)f_p &= \hat{\gamma}(2). \\ &\vdots \quad \vdots \\ \hat{\gamma}(p-1)f_1 + \cdots + \hat{\gamma}(0)f_p &= \hat{\gamma}(p). \end{aligned}$$

These are called the Yule-Walker equations. The solutions $\hat{\phi}_1, \ldots, \hat{\phi}_p$ are the estimates of $\phi_1, \ldots, \phi_p$. The first equation in the normal equations turns out to be $f_0 = (1 - f_1 - \cdots - f_p)\bar{X}$. Estimates of ϕ_0 and μ are $\hat{\phi}_0 = (1 - \hat{\phi}_1 - \cdots - \hat{\phi}_p)\bar{X}$ and $\bar{X}$, respectively.

When $p = 2$, the Yule-Walker equations are

$$\begin{aligned} \hat{\gamma}(0)f_1 + \hat{\gamma}(1)f_2 &= \hat{\gamma}(1), \\ \hat{\gamma}(1)f_1 + \hat{\gamma}(0)f_2 &= \hat{\gamma}(2). \end{aligned}$$

Mathematical results show that the AR(p) model obtained via Yule-Walker estimates is stationary. Using matrix notations, the Yule-Waler equations can be written as $\hat{\mathbf{\Gamma}}_p \boldsymbol{f} = \hat{\boldsymbol{\gamma}}_p$, where $\hat{\mathbf{\Gamma}}_p$ is a $p \times p$ matrix whose element (i, j) is $\hat{\gamma}(i-j)$ (note that $\hat{\gamma}(-h) = \hat{\gamma}(h)$ for any integer h), $\boldsymbol{f}$ is the column vector of $f_1, \ldots, f_p$, and $\hat{\boldsymbol{\gamma}}_p$ is the column vector of $\hat{\gamma}(1), \ldots, \hat{\gamma}(p)$. The estimates are (denoting the $p \times 1$ vector with elements $\hat{\phi}_1, \ldots, \hat{\phi}_p$ by $\hat{\boldsymbol{\phi}}$)

$$\hat{\boldsymbol{\phi}} = \hat{\mathbf{\Gamma}}_p^{-1}\hat{\boldsymbol{\gamma}}_p \text{ and } \hat{\sigma}^2 = \hat{\gamma}(0)(1 - \hat{\boldsymbol{\phi}}^T\hat{\boldsymbol{\gamma}}_p) = \hat{\gamma}(0)(1 - \hat{\boldsymbol{\gamma}}_p^T\hat{\mathbf{\Gamma}}_p\hat{\boldsymbol{\gamma}}_p).$$

The following results are approximately true

$$E(\hat{\boldsymbol{\phi}}) \approx \boldsymbol{\phi}, Var(\hat{\boldsymbol{\phi}}) \approx (1/n)\sigma^2\mathbf{\Gamma}_p^{-1}, s^2(\hat{\boldsymbol{\phi}}) \approx (1/n)\hat{\sigma}^2\hat{\mathbf{\Gamma}}_p^{-1}.$$

The last result can be used to test (for any $1 \leq j \leq p$) $H_0 : \phi_j = 0$ against $H_1 : \phi_j \neq 0$. The test statistic is $z^* = \hat{\phi}_j/s(\hat{\phi}_j)$, where $s(\hat{\phi}_j) = \hat{\sigma}\sqrt{\hat{\gamma}^{jj}}/\sqrt{n}$, where $\hat{\gamma}^{jj}$ is the j^{th} diagonal element of $\hat{\mathbf{\Gamma}}_p^{-1}$. If the level of significance $\alpha = 0.05$, then reject the null hypothesis if $|z^*| > 1.96$.