Mahrad Sharifvaghefi — University of Pittsburgh

Lecture Note 4

- Selecting the number of lags for autoregressive process

In this note, we will discuss the following two ways of estimating the number of lags for autoregressive process:

1. Partial autocorrelation function (PACF)

2. Information criteria

Partial autocorrelation function (PACF)

Definition 1. *Consider a sequence of weakly stationary random variables* $x_1, x_2, \cdots$. *The partial covariance between* x_t *and* x_{t-k} *is defined as*

$$cov(x_t, x_{t-k}|x_{t-1}, x_{t-2}, \cdots, x_{t-(k-1)}).$$

The correponding partial correlation between x_t *and* x_{t-k}, $\alpha(k)$, *is defined as*

$$\alpha(k) = corr(x_t, x_{t-k}|x_{t-1}, x_{t-2}, \cdots, x_{t-(k-1)}) = \frac{cov(x_t, x_{t-k}|x_{t-1}, x_{t-2}, \cdots, x_{t-(k-1)})}{var(x_t|x_{t-1}, x_{t-2}, \cdots, x_{t-(k-1)})}.$$

Q: What is partial autocorrelation function for AR(1) process.

A: Consider the following AR(1) process:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \varepsilon_t.$$

Since there is no random variables between x_t and x_{t-1}, the partial covariance between x_t and x_{t-1} is equivalent to covariance between x_t and x_{t-1}. that is

$$\text{cov}(x_t, x_{t-1}) = \text{cov}(\phi_0 + \phi_1 x_{t-1} + \varepsilon_t, x_{t-1}) = \phi_1 \text{cov}(x_{t-1}, x_{t-1}) = \phi_1 \text{var}(x_{t-1}).$$

Hence, the partial correlation between x_t and x_{t-1}, $\alpha(1)$, is

$$\alpha(1) = \phi_1.$$

The partial covariance between x_t and x_{t-2} is

$$\text{cov}(x_t, x_{t-2}|x_{t-1}) = \text{cov}(\phi_0 + \phi_1 x_{t-1} + \varepsilon_t, x_{t-2}|x_{t-1}) = \phi_1 \text{cov}(x_{t-1}, x_{t-2}|x_{t-1}) = 0.$$

and hence $\alpha(2) = 0$. We can similarly show that $\alpha(k) = 0$ for all $k = 2, 3, \cdots$.

Q: What is partial autocorrelation function for AR(2) process.

A: Consider the following AR(2) process:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t.$$

The partial covariance between x_t and x_{t-1} is

$$\text{cov}(x_t, x_{t-1}) = \text{cov}(\phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t, x_{t-1}) = \phi_1 \text{var}(x_{t-1}) + \phi_2 \text{cov}(x_{t-1}, x_{t-2}).$$

and hence

$$\alpha(1) = \phi_1 + \phi_2 \rho(1).$$

Now consider the partial covariance between x_t and x_{t-2}. It is

$$\begin{aligned}\text{cov}(x_t, x_{t-2}|x_{t-1}) =& \text{cov}(\phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t, x_{t-2}|x_{t-1}) \\ =& \phi_1 \text{cov}(x_{t-1}, x_{t-2}|x_{t-1}) + \phi_2 \text{cov}(x_{t-2}, x_{t-2}|x_{t-1}) \\ =& \phi_2 \text{var}(x_{t-2}|x_{t-1}) = \phi_2 \text{var}(x_t|x_{t-1}).\end{aligned}$$

Hence, $\alpha(2) = \phi_2$. The partial covariance between x_t and x_{t-3} is

$$\begin{aligned}\text{cov}(x_t, x_{t-3}|x_{t-1}, x_{t-2}) =& \text{cov}(\phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t, x_{t-3}|x_{t-1}, x_{t-2}) \\ =& \phi_1 \text{cov}(x_{t-1}, x_{t-3}|x_{t-1}, x_{t-2}) + \phi_2 \text{cov}(x_{t-2}, x_{t-3}|x_{t-1}, x_{t-2}) = 0.\end{aligned}$$

Theorem 1. *Let x_t be generated by following autoregressive process of order p, $AR(p)$,*

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t.$$

*Then its partial autocorrelation function (PACF) is equal ϕ_p for $k = p$ and **zero** for all $k > p$, $\alpha(k) = 0$ for $k = p + 1, p + 2, \cdots$.*

Q: How can we estimated the partial autocorrelation function in practice?

A: Given observations x_t, $t = 1, 2, \cdots, T$, we can estimate the partial autocorrelation function by following the steps below.

(i) Regress of x_t on x_{t-1} as

$$x_t = \phi_{01} + \phi_{11} x_{t-1} + e_{1t}$$

and consider least square (LS) estimator of ϕ_{11}, $\hat{\phi}_{11}$. It can be shown that $\hat{\phi}_{11}$ is a consistent estimator of partial autocorrelation between x_t and x_{t-1}, $\alpha(1)$.

(ii) Regress of x_t on x_{t-1} and x_{t-2} as

$$x_t = \phi_{02} + \phi_{12}x_{t-1} + \phi_{22}x_{t-2} + e_{2t}$$

and estimate ϕ_{22} by least square (LS), $\hat{\phi}_{22}$. It can be shown that $\hat{\phi}_{22}$ is a consistent estimator of partial autocorrelation between x_t and x_{t-2}, $\alpha(2)$.

So in general we can consistently estimate the partial autocorrelation between x_t and x_{t-p}, $\alpha(p)$, by running a regression of x_t on x_{t-1}, x_{t-2}, $\cdots$, x_{t-p} as

$$x_t = \phi_{0p} + \phi_{1p}x_{t-1} + \phi_{2p}x_{t-2} + \cdots + \phi_{pp}x_{t-p} + e_{pt}.$$

The LS estimator of ϕ_{pp}, $\hat{\phi}_{pp}$, is a consistence estimator of $\alpha(p)$.

Q: Is it possible to use partial autocorrelation function to determine the number of lags for autoregressive process.

A: The answer is yes. Suppose the correct number of lags for an AR process, p, is between 0 and $p_{\max}$. Then, we can use the following steps to determine the number of lags by partial autocorrelation function.

(i) Estimate the autocorrelations between x_t and x_{t-k} for $k = 1, 2, \cdots, p_{\max}$, denoted by $\hat{\alpha}(k) = \hat{\phi}_{kk}$, as described above.

(ii) By Theorem 1, $\alpha(k) = 0$ for all $k > p$ and hence their corresponding estimator $\hat{\alpha}(k) = \hat{\phi}_{kk}$ should converge to 0 and the asymptotic variance of $\hat{\alpha}(k)$ is $\frac{1}{T}$.

(iii) So we can test the null hypothesis of $\alpha(k) = 0$ for $k = 1, 2, \cdots, p_{\max}$, iteratively, and set the number of lags equal to highest order partial autocorrelation between 0 and $p_{\max}$ that is significantly different from zero.

The above suggested procedure has some problems due to multiple testing nature of the problem. To aviod these issues, in what follows we propose two alternative procedures for variable selection.

Information criteria

Broadly speaking, each information criterion for selecting number of lags consistent of two terms. The **first term decreases** as the **number of lags increases** while the second term, known as a **penalty term**, **increases** usually at a constant rate as the **number of lags increases**. The linear combination of these two terms give us a **convex function** as a information criteria. The goal is to **select the number of lags** such that the **information criterion** is **minimized**.

Akaike information criterion (AIC)

Akaike information criterion (AIC) can be written as

$$\text{AIC}(k) = \ln(\hat{\sigma}_k^2) + \frac{2k}{T},$$

where k is the number of lags, T is the number of observations, and $\hat{\sigma}_k^2$ is the sum of square of regression residuals divided by T. Note that as the number of lags increases $\ln(\hat{\sigma}_k^2)$ decreases while $\frac{2k}{T}$ increases.

Bayesian information criterion (BIC)

Bayesian information criterion (BIC) can be written as

$$\text{BIC}(k) = \ln(\hat{\sigma}_k^2) + \frac{\ln(T)k}{T},$$

where k is the number of lags, T is the number of observations, and $\hat{\sigma}_k^2$ is the sum of square of regression residuals divided by T.

Q: How can we use one of these information criteria to select the number of lags?

A: We need to take the following steps in order to determined the number of lags using one of these information criteria:

(i) Set the maximum number of lags, $p_{\max}$ that you like to consider.

(ii) For a given number of lags $0 < k < p_{\max}$, run a regression of x_t on $x_{t-1}, x_{t-2}, \cdots, x_{t-k}$, as

$$x_t = \phi_{0k} + \phi_{1k} x_{t-1} + \phi_{2k} x_{t-2} + \cdots + \phi_{kk} x_{t-k} + e_{kt}.$$

(iii) Compute the residuals of estimated regression model, $\hat{e}_{kt}$, by

$$\hat{e}_{kt} = x_t - \hat{\phi}_{0k} - \hat{\phi}_{1k} x_{t-1} + \hat{\phi}_{2k} x_{t-2} - \cdots + \hat{\phi}_{kk} x_{t-k},$$

where $\hat{\phi}_{ik}$, $i = 0, 1, 2, \cdots, k$ are the LS estimator of the regression coefficients.

(iv) Compute $\hat{\sigma}_k^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{e}_{kt}^2$ and then compute AIC(k) or BIC(k).

(v) Select the number of lags, k^*, with corresponding smallest value of the computed information criterion.

• Statistical literature suggests to set the maximum number of lags, $p_{\max}$, equal to $T^{\frac{1}{3}}$ or $\ln(T)$.

• Compared to AIC, BIC tends to select lower number of lags when the sample size is moderate or large, since the penalty term for each additional number of lags increases by $\frac{\ln(T)}{T}$ instead of $\frac{2}{T}$.