Capital University of Economics and Business

ISEM

Financial Econometrics

AR Process in Practice

Instructor:
CHEUNG Ying Lun

Term:
Fall Semester, 2019

October 29, 2019

AR(p) process

We say X_t is an AR(p) process if it takes the form

$$X_t = a_0 + a_1 X_{t-1} + \cdots + a_p X_{t-p} + \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} \left(0, \sigma^2\right). \tag{1}$$

Estimation

In practice, we only observe $\{X_t\}$, but we do not know what the data generating process is. Therefore, we need to estimate Eq.(1) with data. Suppose we observe $\{X_1, \ldots, X_T\}$, we treat X_t as the dependent variable, and $\{X_{t-1}, \ldots, X_{t-p}\}$ as the independent variables, and regress X_t on its own lag.

	dependent variable	independent variables			
t	X_t	X_{t-1}	$\ldots$	X_{t-p+1}	X_{t-p}
1	X_1	-		-	-
2	X_2	X_1		-	-
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
p	X_p	X_{p-1}	$\ldots$	X_1	-
$p+1$	X_{p+1}	X_p	$\ldots$	X_2	X_1
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
T	X_T	X_{T-1}	$\ldots$	X_{T-p+1}	X_{T-p}

Notice that for $t = 1, \ldots, p$, we do not have any observations for X_{t-p}. Therefore, we can only start from $t = p+1$. In total, we now have $(T-p)$ pairs of observations. Writing Eq.(1) in the matrix form,

$$\begin{aligned} X_t &= \begin{pmatrix} a_0 & a_1 & \ldots & a_p \end{pmatrix} \begin{pmatrix} 1 \\ X_{t-1} \\ \vdots \\ X_{t-p} \end{pmatrix} + \varepsilon_t, \qquad t = p+1, \ldots, T \\ &= \mathbf{a}'\mathbf{Z}_t + \varepsilon_t. \end{aligned}$$

We can estimate $\mathbf{a}$ by regressing X_t on $\mathbf{Z}_t$ by the least square method. The solution is

$$\hat{\mathbf{a}} = \left(\sum_{t=p+1}^{T} \mathbf{Z}_t \mathbf{Z}_t' \right)^{-1} \left(\sum_{t=p+1}^{T} \mathbf{Z}_t X_t \right).$$

Finally, to estimate the variance of the innovation terms, we first obtain the residual series as

$$\begin{aligned} \hat{\varepsilon}_t &= X_t - \hat{X}_t \\ &= X_t - \hat{\mathbf{a}}'\mathbf{Z}_t \\ &= X_t - (\hat{a}_0 + \hat{a}_1 X_{t-1} + \cdots + \hat{a}_p X_{t-p}) \end{aligned}$$

We estimate σ^2 by

$$\hat{\sigma}^2 = \frac{1}{T - 2p - 1} \sum_{t=p+1}^{T} \hat{\varepsilon}_t^2.$$

Here note that we are dividing the summation by $T - 2p - 1$, since we only have $T - p$ observations, and we lose $p + 1$ degree of freedom as we are estimating $p + 1$ parameters $a_0, \ldots, a_p$.

Order selection

In practice, we usually do not know what order of the AR process we should use. There are several methods that can help us with that.

Model adequacy By assumption, the innovation term ε_t is a white noise sequence, i.e., there is no serial correlation in ε_t. If our model is adequate, then the estimation residuals $\hat{\varepsilon}_t$ should also inherit this property. Therefore, one way to check the model adequacy is by examining the autocorrelation of the residual sequence and check whether it is significantly different from zero. If the test results show serial correlation in the residual sequence, then we can increase the number of lags in the AR process.

Partial autocorrelation function Notice that an AR(p) process can also be written as an AR($p + 1$) process, with $a_{p+1} = 0$. Therefore, we can estimate an AR(j) model with increasing order,

$$\begin{aligned}
X_t &= a_{0,1} + a_{1,1} X_{t-1} + e_{1t} \\
X_t &= a_{0,2} + a_{1,2} X_{t-1} + a_{2,2} X_{t-2} + e_{2t} \\
X_t &= a_{0,3} + a_{1,3} X_{t-1} + a_{2,3} X_{t-2} + a_{3,3} X_{t-3} + e_{3t} \\
&\vdots
\end{aligned}$$

If the true model is an AR(p) process, then $\hat{a}_{p+1,p+1}$ should be close to zero, while $\hat{a}_{j,j}$ should be non-zero for $j \leq p$. We can look at the sequence $\{\hat{a}_{1,1}, \hat{a}_{2,2}, \ldots\}$ and find the cut-off point when the sequence drops to zero. This sequence is called the partial autocorrelation function (PACF).

Information criterion Another popular method of choosing the lag order is by information criterion (IC), e.g.,

$$\begin{aligned}
AIC(j) &= \ln(\tilde{\sigma}_j^2) + \frac{2j}{T} \\
BIC(j) &= \ln(\tilde{\sigma}_j^2) + \frac{j \ln T}{T}
\end{aligned}$$

An information criterion can be separated into two parts: the first part $\ln(\tilde{\sigma}_j^2)$ measures the fitness of the model; while the second part is a penalty term that increases with the number of lags chosen in the AR model. The intuition behind is that:

- If $j < p$, then increasing j can significantly improve the fitness of the model, thereby reducing $\ln(\tilde{\sigma}_j^2)$. Therefore, the IC become smaller when we increase j.
- If $j \geq p$, then increasing j does not improve the fitness of the model. In this case, the penalty term will dominate and the IC become larger when we increase j.

Therefore, the optimal lag order is the one that *minimize* the IC.

Forecasting

The optimal forecast of an AR process is the conditional expectation. At time t, we already observe X_t and its past values. Therefore, our information set includes $\mathcal{I}_t = \{X_t, X_{t-1}, \ldots\}$. Our h-step-ahead forecast is given by

$$\hat{X}_t(h) = a_0 + a_1 \hat{X}_t(h-1) + \cdots + a_p \hat{X}_t(h-p),$$

where $\hat{X}_t(i) = X_{t+i}$ if $i \leq 0$.

Example: Forecasting an AR(1) process

Consider the AR(1) process

$$X_t = 1 + 0.5X_{t-1} + e_t, \qquad e_t \overset{\text{iid}}{\sim} \mathcal{N}(0, 1).$$

Find the h-step ahead forecast and the respective variances of the forecast errors for $h = 1, 2$. What if $h \to \infty$?

The one-step-ahead forecast is given by

$$\hat{X}_t(1) = \mathbb{E}[X_{t+1}|X_t] = 1 + 0.5X_t.$$

The estimation error and its variance are respectively

$$\hat{\varepsilon}_t(1) = \varepsilon_{t+1}, \qquad \text{var}(\hat{\varepsilon}_t(t)) = \sigma^2.$$

Note that X_{t+2} is

$$\begin{aligned} X_{t+2} &= 1 + 0.5X_{t+1} + \varepsilon_{t+2} \\ &= 1 + 0.5(1 + 0.5X_t + \varepsilon_{t+1}) + \varepsilon_{t+2} \\ &= 1.5 + 0.25X_t + \varepsilon_{t+2} + 0.5\varepsilon_{t+1}. \end{aligned}$$

Therefore, the two-step-ahead forecast is

$$\hat{X}_t(2) = 1.5 + 0.25X_t.$$

The respective estimation error and its variance are

$$\hat{\varepsilon}_t(2) = \varepsilon_{t+2} + 0.5\varepsilon_{t+1}, \qquad \text{var}(\hat{\varepsilon}_t(2)) = (1 + 0.5^2)\sigma^2 = 1.25\sigma^2.$$

Here $\text{var}(\hat{\varepsilon}_t(2)) > \text{var}(\hat{\varepsilon}_t(1))$, meaning that we have a more accurate one-step-ahead forecast than a two-step-ahead forecast. In other words, X_t contains more information on X_{t+1} than X_{t+2}.

In general,

$$X_{t+h} = (1 + 0.5 + \cdots + 0.5^{h-1}) + 0.5^h X_t + \sum_{j=0}^{h-1} 0.5^j \varepsilon_{t+h-j}.$$

Therefore, the h-step-ahead forecast is

$$\hat{X}_t(h) = (1 + 0.5 + \cdots + 0.5^{h-1}) + 0.5^h X_t.$$

The estimation error and its variance are

$$\hat{\varepsilon}_t(h) = \sum_{j=0}^{h-1} 0.5^j \varepsilon_{t+h-j}, \qquad \text{var}(\hat{\varepsilon}_t(h)) = \sum_{j=0}^{h-1} 0.5^{2j}\sigma^2.$$

As $h \to \infty$,

$$\lim_{h\to\infty} \hat{X}_t(h) = \sum_{j=0}^{\infty} 0.5^j = \frac{1}{1-0.5} = \mu_X.$$

Notice that the forecast of X_t will go back to its unconditional mean. We call this feature *mean-reverting*. Every weakly stationary process is mean-reverting. Recall that the autocovariance $\gamma_X(h) \to 0$ as $h \to \infty$. The information on X_{t+h} from X_t diminishes when h increases. Therefore, for the value of X_{t+h} in the infinite future, the best guess given current information is its unconditional mean, which is the same as if we do not have any information on X_t.

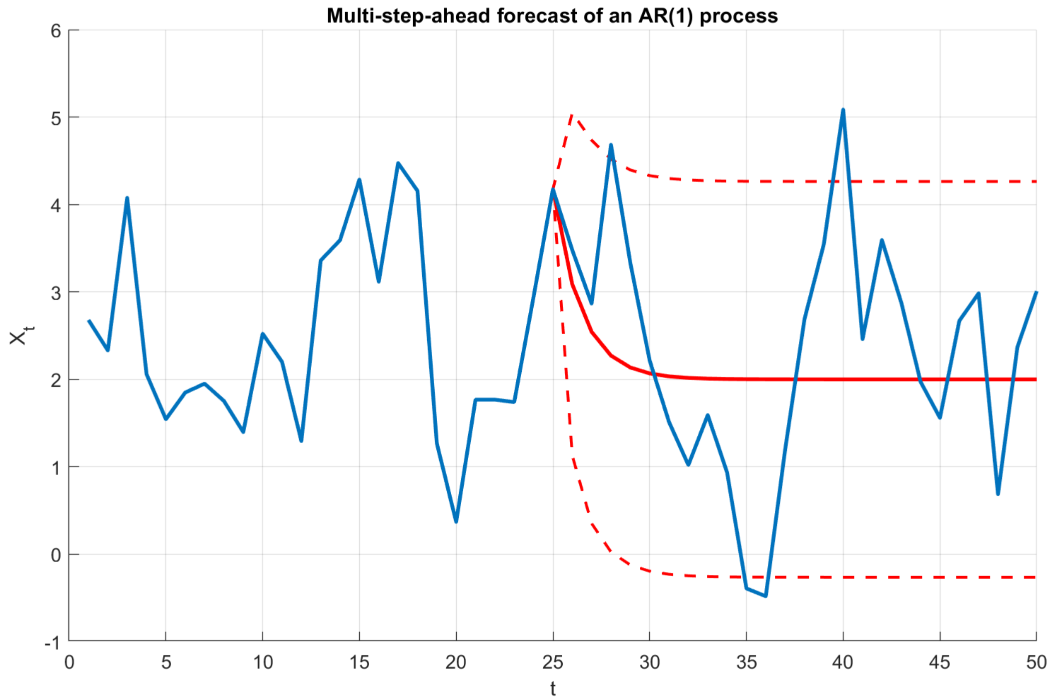
The estimation error and its variance are

$$\lim_{h\to\infty} \hat{\varepsilon}_t(h) = \lim_{h\to\infty} \sum_{j=0}^{h-1} 0.5^j \varepsilon_{t+h-j},$$

$$\lim_{h\to\infty} \text{var}(\hat{\varepsilon}_t(h)) = \sum_{j=0}^{\infty} 0.5^{2j}\sigma^2 = \frac{\sigma^2}{1-0.5^2} = \text{var}(X_{t+h}).$$

The variance of the estimation error is the same as the variance of X_{t+h} itself. This again implies that X_t does not contain any information on X_{t+h} as $h \to \infty$.

The following graph plots a simulated AR(1) process (blue line), together with the multi-step-ahead forecast (red line) and the 95% interval (red dotted line).



We observe that the forecast drops back to the unconditional mean very quickly.