25 Spring 439/639 TSA: Lecture 11

Dr Sergey Kushnarev

Table of contents

More transformations of time series

The goal of transformation is to make the time series "more" stationary, and/or make the time series a normal process (if possible). Last time we already introduced **Variance Stabilizing Transformation** (under certain settings).

Another option: **Box–Cox transformations**, transforming y to $g(y)$ as follows

$$g(y) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log y, & \lambda = 0 \end{cases}$$

Exercise: Show $\lim_{\lambda\to 0} \frac{y^\lambda - 1}{\lambda} = \log y$.

The λ in Box–Cox transformation above is chosen via an MLE approach. See R notebook later.

Another common way is to **take the difference of logarithm**, i.e., log-differences. This can be practically useful in specific applications.

In finance, suppose the time series (Y_t) can be written as follows

$$Y_t = Y_{t-1} + X_t \cdot Y_{t-1} = Y_{t-1}\left(1 + X_t\right).$$

So (X_t) is the percentage change of (Y_t), and we have

$$\log Y_t = \log Y_{t-1} + \log(1 + X_t).$$

Then the log-difference, (or the log-returns, the returns) of (Y_t) is

$$\nabla \log Y_t = \log Y_t - \log Y_{t-1} = \log(1 + X_t) \approx X_t$$

where the last step is because the percentage change X_t is usually small in finance. (*Note:* X_t can be positive or negative, but close to 0.) And in practice, the time series (X_t) is usually stationary. So taking the transformation $\nabla \log Y_t$ gives a more stationary time series.

Summary: to make a time series more stationary, we can consider differencing, variance stabilizing transformation, taking logarithm, Box-Cox transformations, and combination of these.

Roadmap ahead

So far in this course, we already studied models for time series (we mainly studied ARIMA(p, d, q)).

In the next module of this course, we will look at the following topics: Suppose we observe $(y_1, y_2, \dots, y_n)$,

1. **Model Specification:** determine the best/ the most appropriate (p, d, q) in the model, based on sample ACF or sample PACF.
2. **Fit the selected model or Parameter Estimation:** using MLE, LS, MoM, etc.
3. **Diagnostic:** analyze the residuals.
4. **Forecasting**

Bartlett's Theorem

First, recall the sample ACF for observed $(Y_1, ..., Y_n)$: (see lecture 9)

$$\hat{\rho}_k = r_k = \frac{\sum_{t=k+1}^{n} \left(Y_t - \overline{Y}\right)\left(Y_{t-k} - \overline{Y}\right)}{\sum_{t=1}^{n} \left(Y_t - \overline{Y}\right)^2}.$$

Remark: The theoretical ACF ρ_k of a given time series model is a non-random number (but it can be unknown). The sample ACF r_k, defined above, is a random variable if we think $(Y_1, ..., Y_n)$ as a random realization of a time series model. So the sample ACF r_k follows a certain sampling distribution (which depends on the underlying time series model).

The **Bartlett's Theorem** (see equation (6.1.2) in Cryan and Chan) says, for a fixed m, the sampling (joint) distribution of $(r_1, ..., r_m)$ approaches a multivariate normal distribution as $n \to \infty$, i.e.,

$$\vec{r} \sim \text{MVN}(\vec{\rho}, \frac{1}{n}C), \quad \text{as } n \to \infty,$$

$$\text{where} \quad \vec{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}, \quad \vec{\rho} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_m \end{bmatrix},$$

n is the sample size of observations (of the time series), and the matrix C is an $m \times m$ matrix. The full detailed formula for C can be found in equation (6.1.2) in Cryan and Chan. (*Note:* In general, C is not diagonal.)

In particular, the diagonal entries of C are

$$c_{ii} = \sum_{k=-\infty}^{+\infty} (\rho_{k+i}^2 + \rho_{k-i}\rho_{k+i} - 4\rho_i\rho_k\rho_{k+i} + 2\rho_i^2\rho_k^2).$$

Then the Bartlett's Theorem implies that, as $n \to \infty$,

$$r_i \sim \mathcal{N}\left(\rho_i, \frac{1}{n}c_{ii}\right).$$

Example: white noise

Suppose $Y_t \sim \text{WN}(0, \sigma^2)$, then

$$\rho_0 = 1, \text{ and } \rho_i = 0 \text{ for } i \geq 1.$$

Let's look at

$$c_{ii} = \sum_{k=-\infty}^{+\infty} \left(\rho_{k+i}^2 + \rho_{k-i}\rho_{k+i} - 4\rho_i\rho_k\rho_{k+i} + 2\rho_i^2\rho_k^2\right).$$

If $i \geq 1$, then $\rho_i = 0$, $c_{ii} = \sum_{k=-\infty}^{+\infty} (\rho_{k+i}^2 + \rho_{k-i}\rho_{k+i})$. Also note that $\rho_{k-i}\rho_{k+i} = 0$ for any k, and $\sum_{k=-\infty}^{+\infty} \rho_{k+i}^2 = \rho_0^2 = 1$, so $c_{ii} = 1$.

By Bartlett's Theorem, for any fixed $i \geq 1$, $r_i \sim N(0, \frac{1}{n})$ (when sample size n is large). Using this result, we can construct 95% CI for ρ_i (for this example):

$$\left[r_i - \frac{2}{\sqrt{n}}, r_i + \frac{2}{\sqrt{n}}\right].$$

We can also compute c_{00}. If $i = 0$, then $c_{ii} = \sum_{k=-\infty}^{+\infty} (\rho_{k+i}^2 + \rho_k^2 - 4\rho_k^2 + 2\rho_k^2) = 0$. By a similar statement from Bartlett's Theorem, $Var(r_0) = 0$. This is not a surprise since the sample ACF at lag 0, i.e. r_0, is always 1.

Example: AR(1)

Suppose (Y_t) follows AR(1), then

$$\rho_k = \phi^k \text{ for } k \geq 0, \text{ and } \rho_k = \phi^{|k|} \text{ for } k < 0.$$

Using the c_{ii} formula from Bartlett's Theorem, we can derive that: (for large n)

$$\text{Var}(r_i) = \frac{c_{ii}}{n} = \frac{1}{n}\left[\frac{(1+\phi^2)(1-\phi^{2i})}{1-\phi^2} - 2i\,\phi^{2i}\right].$$

In particular, when $i = 1$, we have $\text{Var}(r_1) = \frac{1-\phi^2}{n}$. So if ϕ is close to ± 1, then r_1 is a very precise estimate of ρ_1.

If the lag i is very large, then $\text{Var}(r_i) \approx \frac{1}{n} \cdot \frac{1+\phi^2}{1-\phi^2}$. So if ϕ is close to ± 1, then for large i, r_i is not a precise estimate of ρ_i (in the sense that the variance is very large).

Example: MA(1)

Suppose (Y_t) follows MA(1), then

$$\rho_0 = 1, \quad \rho_1 = \rho_{-1} \neq 0, \quad \text{and } \rho_k = 0 \text{ for } |k| \geq 2.$$

For c_{11}, we have

$$\begin{aligned}
c_{11} &= \sum_{k=-\infty}^{+\infty} (\rho_{k+1}^2 + \rho_{k-1}\rho_{k+1} - 4\rho_1\rho_k\rho_{k+1} + 2\rho_1^2\rho_k^2) \\
&= (\rho_0^2 + 2\rho_1^2) + (\rho_{-1}\rho_1) - 4\rho_1(\rho_{-1}\rho_0 + \rho_0\rho_1) + 2\rho_1^2(\rho_0^2 + 2\rho_1^2) \\
&= 1 + 2\rho_1^2 + \rho_1^2 - 4\rho_1^2 - 4\rho_1^2 + 2\rho_1^2 + 4\rho_1^4 \\
&= 1 - 3\rho_1^2 + 4\rho_1^4.
\end{aligned}$$

By Bartlett's Theorem, we have the following (for large n)

$$r_1 \sim \mathcal{N}\left(\rho_1,\ \frac{1 - 3\rho_1^2 + 4\rho_1^4}{n}\right).$$

For $i \geq 2$, we can also derive that

$$c_{ii} = 1 + 2\rho_1^2, \quad r_i \sim \mathcal{N}\left(0,\ \frac{1+2\rho_1^2}{n}\right).$$

Exercise: verify that $c_{ii} = 1 + 2\rho_1^2$ for any $i \geq 2$ (under the MA(1) setting).

Example: MA(q)

For MA(q), we can show that: (for large n)

$$r_i \sim \mathcal{N}\left(0,\ \frac{1+2\sum_{j=1}^{q}\rho_j^2}{n}\right),\ \text{for any lag } i \geq q+1,$$

which is similar to the $i \geq 2$ case in MA(1).

Hypothesis Testing for MA(q)

We want to do hypothesis testing in the following form

$$H_0 : \text{series is } \text{MA}(q) \quad \text{vs.} \quad H_a : \text{series is not } \text{MA}(q).$$

For example, let's first look at the hypothesis testing for MA(1):

$$H_0 : \text{series is } \text{MA}(1) \quad \text{vs.} \quad H_a : \text{series is not } \text{MA}(1).$$

By the earlier results from this lecture, under H_0, i.e. MA(1),

$$r_i \sim \mathcal{N}\left(0, \frac{1+2\rho_1^2}{n}\right), \quad \text{for any } i \geq 2.$$

So under H_0, with 95% probability,

$$r_i \in \left[-\frac{2}{\sqrt{n}}\sqrt{1+2\rho_1^2},\ \frac{2}{\sqrt{n}}\sqrt{1+2\rho_1^2}\right],$$

and approximately, with 95% probability,

$$r_i \in \left[-\frac{2}{\sqrt{n}}\sqrt{1+2r_1^2},\ \frac{2}{\sqrt{n}}\sqrt{1+2r_1^2}\right].$$

So we construct the rejection region

$$|r_i| > \frac{2}{\sqrt{n}}\sqrt{1+2r_1^2},\ \text{for } i \geq 2.$$

For example if $|r_2| > \frac{2}{\sqrt{n}}\sqrt{1+2r_1^2}$, then reject H_0. (*Note:* we may have access to multiple sample ACFs ($r_2, r_3, ...$), rather than just one single r_2. We can use them collectively to reject H_0.)

If the H_0 of the previous hypothesis testing, i.e. MA(1) is rejected, then we turn to the next hypothesis testing:

$$H_0 : \text{series is } \text{MA}(2) \quad \text{vs.} \quad H_a : \text{series is not } \text{MA}(2).$$

We can use a similar analysis to derive the rejection region (see the earlier example of MA(q) in this lecture) for this hypothesis testing. The rejection region for H_0 (series is MA(2)) is

$$|r_i| > \frac{2}{\sqrt{n}}\sqrt{1+2r_1^2+2r_2^2},\ \text{for } i \geq 3.$$

We can repeat this process until we fail to reject a hypothesis that the series is an MA(q) (for some q).

Example: Assume we observe a sample with $n = 100$, and the sample ACFs are $r_1 = 0.5$, $r_2 = 0.4$, $r_3 = 0.4$, $r_4 = 0.3$. Suppose we want to hypothesis testing to determine an MA(q) model for this sample.

We start with testing whether it is white noise (i.e. MA(0)). For white noise, $r_i \in [-\frac{2}{\sqrt{n}}, \frac{2}{\sqrt{n}}]$ for all $i \geq 1$ with 95% probability. (The rejection region for white noise is $|r_i| > \frac{2}{\sqrt{n}} = 0.2$ for $i \geq 1$.) We reject the white noise using the observed sample ACFs r_1, r_2, r_3, r_4.

For testing MA(1): the rejection region is

$$|r_i| > \frac{2}{\sqrt{n}}\sqrt{1+2r_1^2} = \frac{2}{10}\sqrt{1+2\cdot 0.5^2} = 0.2\sqrt{1.5} \approx 0.245, \quad \text{for } i \geq 2.$$

So we reject the MA(1) using the observed sample ACFs r_2, r_3, r_4. Next, test MA(2).

Exercise: finish this testing process for this example.

Brief introduction to Partial Autocorrelation Function (PACF)

Motivation: We just saw that one can test if a process is an MA(q) for a specific q using the sample ACFs. A natural question is, can we determine the order p of AR(p) from the sample ACFs? The answer is no. And we need to look at Partial ACF (PACF) to determine the order p of AR(p). The sample PACF can help us determine the order of AR(p).

The notation of **Partial Autocorrelation Function (PACF)** is ϕ_{kk}, which denotes the partial autocorrelation at lag k. e will see several different definitions in the next lecture. Here is one way to define it.

$$\phi_{kk} \stackrel{\text{def}}{=} \operatorname{corr}\left(Y_t,\ Y_{t-k} \mid Y_{t-1},\, Y_{t-2},\, \dots,\, Y_{t-k+1}\right).$$

This means, ϕ_{kk} is the **conditional correlation** between Y_t and Y_{t-k} **conditional on** all intermediate values $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$.

Example: consider the AR(1) process, $Y_t = \phi Y_{t-1} + e_t$ (with $\phi \neq 0$). We can show that $\phi_{11} = \rho_1 \neq 0$, and $\phi_{kk} = 0$ for any lag $k \geq 2$.

In general, for AR(p), the PACF at lags 1 through p can be nonzero, and PACF at lags $k \geq p+1$ are all zero.

Remark: for any time series (not necessarily special models like ARIMA), the definition itself contains two special cases. For $k = 0$, ϕ_{00} is always 1 by definition. For $k = 1$, the conditional correlation reduces to an unconditional correlation, so $\phi_{11} = \operatorname{corr}(Y_t, Y_{t-1})$, and $\phi_{11} = \rho_1$ assuming stationarity.