25 Spring 439/639 TSA: Lecture 19

Dr Sergey Kushnarev

Table of contents

Forecasting (continued)

Example 4: random walk with drift

Consider the following model

$$Y_t = Y_{t-1} + \theta_0 + e_t,$$

where θ_0 is a constant. (If $\theta_0 = 0$, then it reduces to random walk.)

The prediction $\widehat{Y}_t(1)$ is

$$\widehat{Y}_t(1) = \mathbb{E}\left[Y_{t+1} \mid Y_{1,\dots,t}\right] = \mathbb{E}\left[Y_t + \theta_0 + e_{t+1} \mid Y_{1,\dots,t}\right] = Y_t + \theta_0.$$

In general, $\widehat{Y}_t(h) = Y_t + h\theta_0$ for any $h \geq 1$.

Exercise: show that $\widehat{Y}_t(h) = Y_t + h\theta_0$.

The forecast error $e_t(h)$ is

$$
\begin{aligned}
e_t(h) &= Y_{t+h} - \widehat{Y}_t(h) \\
&= Y_{t+h-1} + \theta_0 + e_{t+h} - (Y_t + h\theta_0) \\
&= Y_{t+h-2} + 2\theta_0 + e_{t+h-1} + e_{t+h} - (Y_t + h\theta_0) \\
&= \cdots = Y_t + h\theta_0 + e_{t+1} + \cdots + e_{t+h} - (Y_t + h\theta_0) \\
&= e_{t+1} + \cdots + e_{t+h} = \sum_{j=1}^{h} e_{t+j}.
\end{aligned}
$$

So $\mathbb{E}[e_t(h)] = 0$ and $\text{Var}(e_t(h)) = h\sigma_e^2$. As $h \to \infty$, $\text{Var}(e_t(h)) \to \infty$. *Note:* this is a typical behavior of nonstaionary processes.

Example 5: ARMA$(1,1)$ with mean

Consider an invertible ARMA$(1,1)$ (i.e., $|\theta| < 1$) with nonzero mean. It can be parametrized as follows:

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1} + \theta_0.$$

Remark: θ_0 is not the mean of (Y_t). If using the mean parameter $\mu = \mathbb{E}[Y_t]$, then this model should be $Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t - \theta e_{t-1}$. From this we can see $\theta_0 = (1 - \phi)\mu$, so $\mu = \frac{\theta_0}{1-\phi}$ if $\phi \neq 1$.

The prediction $\widehat{Y}_t(1)$ is

$$\begin{aligned}
\widehat{Y}_t(1) = \mathbb{E}\left[Y_{t+1} \mid Y_{1,\ldots,t}\right] &= \mathbb{E}\left[\phi Y_t + e_{t+1} - \theta e_t + \theta_0 \mid Y_{1,\ldots,t}\right] \\
&= \phi Y_t - \theta\, \mathbb{E}\left[e_t \mid Y_{1,\ldots,t}\right] + \theta_0 \\
&= \phi Y_t - \theta\, e_t + \theta_0,
\end{aligned}$$

where the last step is of the same reasoning we had before (see the MA(1) example in lecture 18), e_t is a function of $Y_{1,\ldots,t}$ via the invertible representation.

For $h \geq 2$, we can show that

$$\widehat{Y}_t(h) = \phi \widehat{Y}_t(h-1) + \theta_0,$$

and ultimately get

$$\widehat{Y}_t(h) = \phi^h Y_t - \phi^{h-1}\theta e_t + \frac{1 - \phi^h}{1 - \phi}\theta_0.$$

Suppose this model is also causal (so $|\phi| < 1$): As $h \to \infty$, $\widehat{Y}_t(h) \to \frac{1}{1-\phi}\theta_0 = \mu$.

Exercise: derive these results above using the mean parameter of ARMA$(1,1)$, i.e. start from $Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t - \theta e_{t-1}$ instead of $Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1} + \theta_0$. (See the remark above for the connection between these two forms.)

Some observations from the examples

So far we have seen some examples (this lecture and the previous lecture) of forecasting.

Example 4 is non-stationary: $\mathrm{Var}(e_t(h)) \to \infty$ as $h \to \infty$.

In examples 2,3,5 (when stationarity is assumed), we have $\widehat{Y}_t(h) \to \mu$ as $h \to \infty$.

In examples 2,3, we also showed $\mathrm{Var}(e_t(h)) \to \gamma_0$ as $h \to \infty$.

In general, for invertible MA(q) models, when $h > q$, $\mathrm{Var}(e_t(h)) = \gamma_0$ exactly; for causal AR(p) and causal+invertible ARMA(p,q), $\mathrm{Var}(e_t(h))$ is increasing in h and it converges to γ_0.

We will show some of these results in a unified way in the rest of this lecture.

Prediction for ARMA(p, q)

Assume the ARMA(p, q) is **causal and invertible**.

For any h, since $Y_{t+h} = \phi_1 Y_{t+h-1} + \cdots + \phi_p Y_{t+h-p} + e_{t+h} - \theta_1 e_{t+h-1} - \cdots - \theta_q e_{t+h-q}$, we have

$$\begin{aligned}
\widehat{Y}_t(h) &= \mathbb{E}\left[Y_{t+h} \mid Y_{1,\ldots,t}\right] \\
&= \phi_1 \widehat{Y}_t(h-1) + \cdots + \phi_p \widehat{Y}_t(h-p) + \mathbb{E}\left[e_{t+h} - \theta_1 e_{t+h-1} - \cdots - \theta_q e_{t+h-q} \mid Y_{1,\ldots,t}\right].
\end{aligned}$$

If $h > q$, then the last term is zero, i.e., $\mathbb{E}\left[e_{t+h} - \theta_1 e_{t+h-1} - \cdots - \theta_q e_{t+h-q} \mid Y_{1,\ldots,t}\right] = 0$, so

$$\widehat{Y}_t(h) = \phi_1 \widehat{Y}_t(h-1) + \cdots + \phi_p \widehat{Y}_t(h-p), \quad \text{for any } h > q,$$

which looks like the YW equations. So for lead time $h > q$, the predictions $\widehat{Y}_t(h)$ satisfy the recursive YW equations. By the earlier results from the course (see lectures 5,6), if $z_1, \ldots, z_p$ are distinct roots of the AR polynomial, (we already have $|z_i| > 1$ from causality), then there exist some complex numbers $C_1, \ldots, C_p$ such that

$$\widehat{Y}_t(h) = C_1 \left(\frac{1}{z_1}\right)^h + \cdots + C_p \left(\frac{1}{z_p}\right)^h, \quad \text{for any } h > q.$$

This implies
$$\widehat{Y}_t(h) \to 0, \text{ as } h \to \infty, \text{(and it decays exponentially)}.$$

Note that we considered mean zero ARMA(p, q) here, so the prediction converges to the mean of the ARMA(p, q). If the ARMA(p, q) has mean μ, then $\widehat{Y}_t(h) \to \mu$ as $h \to \infty$.

Remark: it seems we only used causality of the process in this part. But we still implicitly used the invertibility: When applying the recursion result for YW equations, we omitted the explanation that the initial conditions can be satisfied (i.e., $\widehat{Y}_t(h)$ are well defined for $h \le q$.) We need invertibility to make the term $\mathbb{E}\left[e_{t+h} - \theta_1 e_{t+h-1} - \cdots - \theta_q e_{t+h-q} \mid Y_{1,\ldots,t}\right]$ well defined for $h \le q$.

Forecast error for ARMA(p, q)

We can use the GLP representation to get the forecast error $e_t(h)$.

We still assume the ARMA(p, q) is **causal and invertible**. By causality, there exists a GLP representation for the ARMA(p, q) process, $Y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}$. (For ARMA$(p, q)$, we have $\psi_0 = 1$.) Then we can express Y_{t+h} in the form of GLP:

$$Y_{t+h} = \underbrace{\psi_0 e_{t+h} + \psi_1 e_{t+h-1} + \cdots + \psi_{h-1} e_{t+1}}_{\text{future}} + \underbrace{\psi_h e_t + \psi_{h+1} e_{t-1} + \cdots}_{\text{past}}$$
$$= I_t(h) + C_t(h),$$

where the first part $I_t(h) = e_{t+h} + \psi_1 e_{t+h-1} + \cdots + \psi_{h-1} e_{t+1}$ can be seen as the "future", and the second part $C_t(h) = \psi_h e_t + \psi_{h+1} e_{t-1} + \cdots$ involves the "past".

- The future part $I_t(h)$ is independent of the observed $Y_{1,\ldots,t}$ since the e_k terms in $I_t(h)$ are all ahead of time t and the process is causal.
- The past part $C_t(h)$ **is a function of** $Y_{1,\ldots,t}$ **by invertibility**, since each e_{t-j} (for any $j \ge 0$) can be expressed as a function of $Y_{1,\ldots,t}$.

Then if we take the conditional expectation $\mathbb{E}[\cdot | Y_{1,\ldots,t}]$ in the equation above, we get

$$\widehat{Y}_t(h) = \mathbb{E}[Y_{t+h} \mid Y_{1,\ldots,t}] = \mathbb{E}[I_t(h) \mid Y_{1,\ldots,t}] + \mathbb{E}[C_t(h) \mid Y_{1,\ldots,t}] = C_t(h).$$

So the prediction and the error are

$$\widehat{Y}_t(h) = C_t(h) = \psi_h e_t + \psi_{h+1} e_{t-1} + \cdots$$
$$e_t(h) = Y_{t+h} - C_t(h) = I_t(h) = \psi_0 e_{t+h} + \psi_1 e_{t+h-1} + \cdots + \psi_{h-1} e_{t+1}.$$

From this we can see $\mathbb{E}[e_t(h)] = 0$, $\text{Var}(e_t(h)) = \sigma_e^2 \sum_{j=0}^{h-1} \psi_j^2$. So $\text{Var}(e_t(h))$ is increasing in h, and has the following convergence result:

$$\text{Var}(e_t(h)) = \sigma_e^2 \sum_{j=0}^{h-1} \psi_j^2 \to \sigma_e^2 \sum_{j=0}^{\infty} \psi_j^2 = \gamma_0, \quad \text{as } h \to \infty.$$

Forecasting ARIMA(p, d, q)

For ARIMA(p, d, q), rewrite it as

$$\underbrace{\Phi(B) \ (1 - B)^d \ Y_t = \Theta(B) \ e_t}_{\text{ARIMA}(p,d,q)} \implies \underbrace{\widetilde{\Phi}(B) \ Y_t = \Theta(B) \ e_t}_{\text{nonstationary ARMA}(p+d,q)}$$

where $\Phi(x)$ is the original AR polynomial of order p, and $\widetilde{\Phi}(x) = \Phi(x) \ (1-x)^d$ is the modified AR polynomial of order $p + d$ which has (at least) d unit roots.

Example 6. Consider an ARIMA$(1, 1, 1)$,

$$Y_t - Y_{t-1} = W_t, \quad W_t = \phi W_{t-1} + e_t - \theta e_{t-1}.$$

We can write it as a non-stationary ARMA$(2, 1)$:

$$Y_t = (1 + \phi)Y_{t-1} - \phi Y_{t-2} + e_t - \theta e_{t-1}.$$

By some simple derivations, (assume the invertibility still holds)

$$\widehat{Y}_t(1) = \mathbb{E}\left[(1 + \phi)Y_t - \phi Y_{t-1} + e_{t+1} - \theta e_t \mid Y_{1,\ldots,t}\right] = (1 + \phi)Y_t - \phi Y_{t-1} - \theta e_t,$$

$$\widehat{Y}_t(2) = \mathbb{E}\left[(1 + \phi)Y_{t+1} - \phi Y_t + e_{t+2} - \theta e_{t+1} \mid Y_{1,\ldots,t}\right] = (1 + \phi)\widehat{Y}_t(1) - \phi Y_t,$$

$$\widehat{Y}_t(h) = (1 + \phi)\widehat{Y}_t(h - 1) - \phi\widehat{Y}_t(h - 2), \quad \text{for } h \geq 3.$$

Results for general ARIMA(p, d, q): If $d \geq 1$, then ARIMA(p, d, q) is not stationary or causal. So strictly speaking, our previous results for causal and invertible ARMA(p, q) do not hold for ARIMA(p, d, q).

Idea: we can think of ARIMA(p, d, q) as a non-stationary ARMA$(p + d, q)$, and try the previous approach for ARMA(p, q). We can still get some similar results.

Suppose there exists a "causal GLP-like representation" for an ARIMA(p, d, q): $Y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}$, but $\sum |\psi_j| < \infty$ fails.

(Just like the causal and invertible ARMA(p, q) results,) We still have

$$e_t(h) = I_t(h) = \psi_0 e_{t+h} + \psi_1 e_{t+h-1} + \cdots + \psi_{h-1} e_{t+1} = \sum_{j=0}^{h-1} \psi_j e_{t-j},$$

where ψ_j are from the "GLP-like representation". But now we have a diffrent limiting behavior

$$\text{Var}(e_t(h)) = \sigma_e^2 \sum_{j=0}^{h-1} \psi_j^2 \to \sigma_e^2 \sum_{j=0}^{\infty} \psi_j^2 = \infty, \quad \text{as } h \to \infty.$$

So $\text{Var}(e_t(h))$ diverges to infinity as $h \to \infty$ for ARIMA(p, d, q) with $d \geq 1$.

Prediction interval of forecasting

In general (for either GLP or GLP-like representations), we have

$$e_t(h) = I_t(h) = \sum_{j=0}^{h-1} \psi_j e_{t-j}.$$

If $e_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_e^2)$, then

$$e_t(h) \sim \mathcal{N}\left(0, \sigma_e^2 \sum_{j=0}^{h-1} \psi_j^2\right).$$

Consider the quantile,

$$\Pr\left(-z_{1-\frac{\alpha}{2}} \leq \frac{e_t(h) - 0}{\sqrt{\sigma_e^2 \sum_{j=0}^{h-1} \psi_j^2}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\implies \Pr\left(Y_{t+h} \in \left[\widehat{Y}_t(h) \pm z_{1-\alpha/2}\sqrt{\text{Var}(e_t(h))}\right]\right) = 1 - \alpha,$$

so the $(1 - \alpha)100\%$ PI for Y_{t+h} is $\left[\widehat{Y}_t(h) \pm z_{1-\alpha/2}\sqrt{\text{Var}(e_t(h))}\right]$.

Note: The concept PI and CI are different. Prediction interval is for the random variable Y_{t+h}. Confidence interval is for a parameter θ, where θ is a fixed but unknown value.