TS Box Jenkins

March 2021

Time Series Analysis: Forecasting and Control, 5th Edition, 2015, George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung.

Grenander, U. and Rosenblatt, M. (1953). Statistical spectral analysis of time series arising from stationary stochastic processes. *Annals of Mathematical Statistics*, 24, 537-558

Time series z_t, $t = 0, \pm 1, \pm 2, ...$

Define the backward shift operator $Bz_t = z_{t-1}$, and the linear filter

$$\Psi(B) = 1 + \Psi_1 B + \Psi_2 B^2 + \cdots$$

Basic idea: "Random shocks" i.i.d u_t, $Eu_t = 0, Var(u_t) = \sigma^2$, drive a linear system:

$$z_t = \Psi(B)u_t = \sum_{j=0}^{\infty} \Psi_j u_{t-j}$$

where the sum converges in mean square.

Under certain conditions we have invertibility:

$$u_t = \Psi^{-1}(B)z_t \equiv \Pi(B)z_t$$

Hence

$$(z_t, z_{t-1}, z_{t-2}, \cdots) \iff (u_t, u_{t-1}, u_{t-2}, \cdots)$$

Fact: If $\Psi(B)$ converges for $|B| \leq 1$ then z_t is weekly stationary. In which case

$$R_k = \sigma^2 \sum_{j=0}^{\infty} \Psi_j \Psi_{j+k}, \quad R_0 = \sigma^2 \sum_{j=0}^{\infty} \Psi_j^2$$

Fact: If $\Psi^{-1}(B) = \Pi(B)$ converges for $B \leq 1$ then z_t is invertible

ARMA(p,q)

$$\begin{aligned}
\Phi(B) &= 1 - \phi_1 B - \cdots - \phi_p B^p \\
\Theta(B) &= 1 - \theta_1 B - \cdots - \theta_q B^q \\
\Phi(B) z_t &= \Theta(B) u_t
\end{aligned}$$

If the roots of $\Phi(B) = 0$ are outside the unit circle then z_t is stationary
If the roots of $\Theta(B) = 0$ are outside the unit circle than z_t is invertible

$$u_t = \frac{\Phi(B)}{\Theta(B)} z_t$$

Yule 1920's AR(p): $\Phi(B) z_t = u_t$
Slutski 1920's MA(q): $z_t = \Theta(B) u_t$
Wold 1930's: $\Phi(B) z_t = \Theta(B) u_t$

ARIMA(p,d,q)

Main idea: The $d'th$ difference $\bigtriangledown^d z_t = (1-B)^d z_t$ is stationary.

Autoregressive Integrated Moving Average:

$$\Phi(B)[\bigtriangledown^d z_t] = \Theta(B)u_t$$

Assumption: $E[\bigtriangledown^d z_t] = 0$

How do we find p, d, q? There is a whole list of diagnostic tools. Most important, once we fit and ARIMA model, we check for WN RESIDUALS.

Identification: For example by model selection criteria:

$$AIC = -2\log L(\hat{\boldsymbol{\beta}}) + 2p$$

$$BIC = -2\log L(\hat{\boldsymbol{\beta}}) + p\log n$$

.
Estimation: For example by ML.
Diagnostic Checking: acf(residuals), cpgram(residuals), Ljung-Box test, partial autcorrelation.

1. (0,1,1): $\bigtriangledown z_t = (1-\theta_1 B)u_t$
2. (0,2,2): $\bigtriangledown^2 z_t = (1-\theta_1 B - \theta_2 B^2)u_t$
3. (1,1,1): $(1-\phi_1 B)\bigtriangledown z_t = (1-\theta_1 B)u_t$
4. (1,1,0): $(1-\phi_1 B)\bigtriangledown z_t = u_t$

Minimum MSE Forecasts

Observed time series:

$$\cdots z_{t-2}, z_{t-1}, z_t$$

We wish to predict l-steps ahead:

$$z_{t+l}$$

By a predictor which is a linear function of the observed time series:

$$\hat{z}_t(l)$$

Since $\hat{z}_t(l)$ is a linear function of $\cdots z_{t-2}, z_{t-1}, z_t$, by invertibility, it is also a linear function of $\cdots u_{t-2}, u_{t-1}, u_t$. Therefore,

$$\hat{z}_t(l) = \Psi_l^* u_t + \Psi_{l+1}^* u_{t-1} + \Psi_{l+2}^* u_{t-2} + \cdots$$

Now, with $\Psi_0 \equiv 1$,

$$z_t = \sum_{j=0}^{\infty} \Psi_j u_{t-j} = \sum_{j=-\infty}^{t} \Psi_{t-j} u_j$$

Therefore,

$$\begin{aligned}
z_{t+l} &= \sum_{j=-\infty}^{t+l} \Psi_{t+l-j} u_j \\
&= u_{t+l} + \Psi_1 u_{t+l-1} + \Psi_2 u_{t+l-2} + \cdots + \Psi_{l-1} u_{t+1} \quad (\equiv e_t(l),\ \textbf{future u's}) \\
&+ \Psi_l u_t + \Psi_{l+1} u_{t-1} + \Psi_{l+2} u_{t-2} + \cdots \quad (\textbf{past u's}) \\
&\equiv e_t(l) + \Psi_l u_t + \Psi_{l+1} u_{t-1} + \Psi_{l+2} u_{t-2} + \cdots
\end{aligned}$$

Therefore

$$(\star) \quad z_{t+l} - \hat{z}_t(l) = e_t(l) + u_t(\Psi_l - \Psi_l^*) + u_{t-1}(\Psi_{l+1} - \Psi_{l+1}^*) + \cdots$$

Therefore, since the u_t are i.i.d with mean 0 and variance σ^2,

$$E[z_{t+l} - \hat{z}_t(l)]^2 = \sigma^2(1 + \Psi_1^2 + \cdots + \Psi_{l-1}^2) + \sigma^2 \sum_{j=0}^{\infty} (\Psi_{l+j} - \Psi_{l+j}^*)^2$$

which is **minimized** for $\Psi_{l+j} = \Psi_{l+j}^*$. Therefore from $(\star)$ we have,

$$\boxed{z_{t+l} = \hat{z}_t(l) + e_t(l)}$$

where $e_t(l)$ is the prediction error.

Fact: If $\phi(t)$ is any function of $z_t, z_{t-1}, \cdots$, then

$$E[z_{t+l} - \phi(t)]^2 \geq E[z_{t+l} - E(z_{t+l}|z_t, z_{t-1}, \cdots)]^2$$

But

$$\begin{aligned}
E(z_{t+l}|z_t, z_{t-1}, \cdots) &= E(e_t(l) + \Psi_l u_t + \Psi_{l+1} u_{t-1} + \Psi_{l+2} u_{t-2} + \cdots |z_t, z_{t-1}, \cdots) \\
&= E(e_t(l)|u_t, u_{t-1}, \cdots) \\
&+ E(\Psi_l u_t + \Psi_{l+1} u_{t-1} + \Psi_{l+2} u_{t-2} + \cdots |u_t, u_{t-1}, \cdots) \\
&= 0 + E(\Psi_l u_t + \Psi_{l+1} u_{t-1} + \Psi_{l+2} u_{t-2} + \cdots |u_t, u_{t-1}, \cdots) \\
&= E(\Psi_l u_t + \Psi_{l+1} u_{t-1} + \Psi_{l+2} u_{t-2} + \cdots |u_t, u_{t-1}, \cdots) \\
&= E(\Psi_l^* u_t + \Psi_{l+1}^* u_{t-1} + \Psi_{l+2}^* u_{t-2} + \cdots |u_t, u_{t-1}, \cdots) = \hat{z}_t(l)
\end{aligned}$$

Therefore, linearity implies $\hat{z}_t(l)$ is the BEST MSE PREDICTOR since:

$$\boxed{\hat{z}_t(l) = E(z_{t+l}|z_t, z_{t-1}, \cdots) = E(z_{t+l}|\overset{\mapsto}{z}_t)}$$

Prediction Rules

1. $E(z_{t-j}|\overset{\mapsto}{z}_t) = z_{t-j}, \quad j = 0, 1, 2, ...$
2. $E(z_{t+j}|\overset{\mapsto}{z}_t) = \hat{z}_t(j), \quad j = 1, 2, ...$
3. $E(u_{t-j}|\overset{\mapsto}{z}_t) = u_{t-j}, \quad j = 0, 1, 2, ...$
4. $E(u_{t+j}|\overset{\mapsto}{z}_t) = 0, \quad j = 1, 2, ...$

One Step Ahead Forecast Error

From above

$$e_t(l) = u_{t+l} + \Psi_1 u_{t+l-1} + \Psi_2 u_{t+l-2} + \cdots + \Psi_{l-1} u_{t+1}$$

Hence

$$E[e_t(l)] = 0, \quad Var[e_t(l)] = \sigma^2(1 + \Psi_1^2 + \cdots + \Psi_{l-1}^2)$$

Recall,

$$z_{t+l} = \hat{z}_t(l) + e_t(l)$$

Therefore

$$e_t(1) = z_{t+1} - \hat{z}_t(1) = u_{t+1}$$

Therefore we obtain an important formula:

$$\boxed{u_{t+1} = z_{t+1} - \hat{z}_t(1)}$$

or

$$\boxed{u_t = z_t - \hat{z}_{t-1}(1)}$$

Example: Prediction of ARIMA(1,1,0)

$$(1-0.8B)(1-B)z_t = u_t$$

Or

$$z_t = 1.8z_{t-1} - 0.8z_{t-2} + u_t$$

Hence

$$E(z_{t+1}|\overset{\mapsto}{z}_t) = E(1.8z_t - 0.8z_{t-1} + u_{t+1}|\overset{\mapsto}{z}_t)$$

and

$$\boxed{\hat{z}_t(1) = 1.8z_t - 0.8z_{t-1}}$$

Similarly

$$E(z_{t+2}|\overset{\mapsto}{z}_t) = E(1.8z_{t+1} - 0.8z_t + u_{t+2}|\overset{\mapsto}{z}_t)$$

and

$$\boxed{\hat{z}_t(2) = 1.8\hat{z}_t(1) - 0.8z_t}$$

Similarly

$$E(z_{t+3}|\overset{\mapsto}{z}_t) = E(1.8z_{t+2} - 0.8z_{t+1} + u_{t+3}|\overset{\mapsto}{z}_t)$$

and

$$\boxed{\hat{z}_t(3) = 1.8\hat{z}_t(2) - 0.8\hat{z}_t(1)}$$

Similarly

$$E(z_{t+4}|\overset{\mapsto}{z}_t) = E(1.8z_{t+3} - 0.8z_{t+2} + u_{t+4}|\overset{\mapsto}{z}_t)$$

and

$$\boxed{\hat{z}_t(4) = 1.8\hat{z}_t(3) - 0.8\hat{z}_t(2)}$$

In general

$$\boxed{\hat{z}_t(l) = 1.8\hat{z}_t(l-1) - 0.8\hat{z}_t(l-2)}, \quad l = 3, 4, ...$$

Example: Prediction of ARIMA(0,2,2)

$$(1-B)^2 z_t = (1-0.9B+0.5B^2)u_t$$

or

$$(1-2B+B^2)z_t = (1-0.9B+0.5B^2)u_t$$

or

$$z_t = 2z_{t-1} - z_{t-2} + u_t - 0.9u_{t-1} + 0.5u_{t-2}$$

Therefore

$$E(z_{t+1}|\overset{\mapsto}{z}_t) = E(2z_t - z_{t-1} + u_{t+1} - 0.9u_t + 0.5u_{t-1}|\overset{\mapsto}{z}_t)$$

or

$$\hat{z}_t(1) = 2z_t - z_{t-1} - 0.9u_t + 0.5u_{t-1}$$

Likewise

$$\hat{z}_t(2) = 2\hat{z}_t(1) - z_t + 0.5u_t$$

$$\hat{z}_t(3) = 2\hat{z}_t(2) - \hat{z}_t(1)$$

where

$$\begin{aligned} u_t &= z_t - \hat{z}_{t-1}(1) \\ u_{t-1} &= z_{t-1} - \hat{z}_{t-2}(1) \end{aligned}$$

Example: Prediction of MA(1)=ARIMA(0,0,1)

$$z_t = u_t - \theta u_{t-1}, \quad |\theta| < 1, \ \textbf{invertible}$$

$$\hat{z}_t(1) = E(z_{t+1} | \overset{\mapsto}{z}_t) = E(u_{t+1} - \theta u_t | \overset{\mapsto}{z}_t) = -\theta u_t$$

But

$$u_t = z_t - \hat{z}_{t-1}(1)$$

Therefore,

$$\begin{aligned}
\hat{z}_t(1) &= -\theta[z_t - \hat{z}_{t-1}(1)] \\
&= -\theta z_t + \theta \hat{z}_{t-1}(1) = -\theta z_t + \theta[-\theta u_{t-1}] \\
&= -\theta z_t - \theta^2[z_{t-1} - \hat{z}_{t-2}(1)] \\
&= -\theta z_t - \theta^2 z_{t-1} + \theta^2 \hat{z}_{t-2}(1) \\
&= -\theta z_t - \theta^2 z_{t-1} + \theta^2[-\theta u_{t-2}] \\
&= -\theta z_t - \theta^2 z_{t-1} - \theta^3 u_{t-2} \\
&= -\theta z_t - \theta^2 z_{t-1} - \theta^3 z_{t-2} - \cdots - \theta^n z_1 - \theta^{n+1} u_0
\end{aligned}$$

Example: Prediction Intervals

Recall

$$\begin{aligned} e_t(l) &= u_{t+l} + \Psi_1 u_{t+l-1} + \Psi_2 u_{t+l-2} + \cdots + \Psi_{l-1} u_{t+1} \\ &= z_{t+l} - \hat{z}_t(l) \end{aligned}$$

Consider the conditional distribution

$$P(z_{t+l} | z_t, z_{t-1}, \cdots)$$

with mean

$$\hat{z}_t(l) = E(z_{t+l} | z_t, z_{t-1}, \cdots)$$

and variance

$$\begin{aligned} E[(z_{t+l} - \hat{z}_t(l))^2 | z_t, z_{t-1}, \cdots] &= E[e_t^2(l)] \\ &= (1 + \Psi_1^2 + \cdots + \Psi_{l-1}^2)\sigma^2 \end{aligned}$$

Assumption: $\{u_t\}$ are i.i.d **Gaussian**. That is, $\{u_t\}$ are i.i.d $N(0, \sigma^2)$.

Therefore $100(1-\alpha)\%$ prediction limits are

$$\boxed{z_{t+l} | \overset{\mapsto}{z}_t = \hat{z}_t(l) \pm z_{\alpha/2} \left[1 + \sum_{j=1}^{l-1} \Psi_j^2 \right]^{\frac{1}{2}} \times \sigma}$$

Therefore, the prediction intervals increase as l increases.