Prediction for Box-Jenkins models

Filip Elvander

***Abstract*—Here are some notes elaborating on the prediction of Box-Jenkins models with M input signals. The case of prediction of ARMA is obtained for $M = 0$.**

***Index Terms*—**

I. Notes on prediction for Box-Jenkins models

Consider a system of the form

$$y_t = \sum_{m=1}^{M} \frac{B_m(z)}{A_m(z)} u_t^{(m)} + \frac{C(z)}{A(z)} e_t \tag{1}$$

where the polynomials A and A_m, for $m = 1, \ldots, M$ are assumed stable, C is assumed invertible, and e_t is a white Gaussian noise. Furthermore, $u_t^{(m)}$, for $m = 1, \ldots, M$, are M input signals, assumed uncorrelated with each other, as well as with e_t. We are here interested in predicting y_{t+k} for some $k \geq 1$ given the information available at time t, denoted $\mathcal{F}_t$. First of all, note that we may represent the system in (1) as

$$\tilde{A}(z) y_t = \sum_{m=1}^{M} \tilde{B}_m(z) u_t^{(m)} + \tilde{C}(z) e_t \tag{2}$$

where

$$\tilde{A}(z) = A(z) \prod_{n=1}^{M} A_n(z) \tag{3}$$

$$\tilde{B}_m(z) = B_m(z) A(z) \prod_{n=1, n\neq m}^{M} A_n(z) \tag{4}$$

$$\tilde{C}_m(z) = C(z) \prod_{n=1}^{M} A_n(z), \tag{5}$$

i.e., as an ARMAX model. It may be noted that the resulting polynomials inherit stability from the constituting polynomials. Furthermore, note that it is always possible to find a polynomial F of order $k-1$ and a polynomial G such that

$$\frac{\tilde{C}(z)}{\tilde{A}(z)} = F(z) + \frac{z^{-k} G(z)}{\tilde{A}(z)} \iff \tilde{C}(z) = F(z)\tilde{A}(z) + z^{-k} G(z). \tag{6}$$

Then, we may write

$$y_{t+k} = \frac{\tilde{C}(z)}{\tilde{C}(z)} y_{t+k} = \frac{F(z)\tilde{A}(z) + z^{-k} G(z)}{\tilde{C}(z)} y_{t+k} \tag{7}$$

$$= \frac{F(z)\tilde{A}(z)}{\tilde{C}(z)} y_{t+k} + \frac{G(z)}{\tilde{C}(z)} y_t. \tag{8}$$

Further, using (2), we have

$$\frac{F(z)\tilde{A}(z)}{\tilde{C}(z)} y_{t+k} = \frac{F(z)}{\tilde{C}(z)} \left(\sum_{m=1}^{M} \tilde{B}_m(z) u_{t+k}^{(m)} + \tilde{C}(z) e_{t+k} \right) \tag{9}$$

$$= \sum_{m=1}^{M} \frac{F(z)\tilde{B}_m(z)}{\tilde{C}(z)} u_{t+k}^{(m)} + F(z) e_{t+k}. \tag{10}$$

Thus, we have that

$$y_{t+k} = \sum_{m=1}^{M} \frac{F(z)\tilde{B}_m(z)}{\tilde{C}(z)} u_{t+k}^{(m)} + F(z) e_{t+k} + \frac{G(z)}{\tilde{C}(z)} y_t, \tag{11}$$

i.e., the future value may be split into three components: one corresponding to the influence of the input, one corresponding to future noise values (recall that the order of F is $k-1$), and one corresponding to the observed process history. Then, the optimal prediction is given by

$$\hat{y}_{t+k} = \mathbb{E}\left(y_{t+k} \mid \mathcal{F}_t \right) \tag{12}$$

$$= \sum_{m=1}^{M} \mathbb{E}\left(\frac{F(z)\tilde{B}_m(z)}{\tilde{C}(z)} u_{t+k}^{(m)} \mid \mathcal{F}_t \right) \tag{13}$$

$$+ \mathbb{E}\left(F(z) e_{t+k} \mid \mathcal{F}_t \right) + \mathbb{E}\left(\frac{G(z)}{\tilde{C}(z)} y_t \mid \mathcal{F}_t \right) \tag{14}$$

$$= \sum_{m=1}^{M} \mathbb{E}\left(\frac{F(z)\tilde{B}_m(z)}{\tilde{C}(z)} u_{t+k}^{(m)} \mid \mathcal{F}_t \right) + \frac{G(z)}{\tilde{C}(z)} y_t, \tag{15}$$

where the last equality follows from the fact that y_{t-n}, for $n \geq 0$, are known at time t, and $\mathbb{E}\left(e_{t+n} \mid \mathcal{F}_t \right) = 0$ for $n \geq 1$.

Furthermore, we may simplify the remaining expectation by noting that there always exist polynomials $\hat{F}_m$, for $m = 1, \ldots, M$, all of order $k-1$, and polynomials $\hat{G}_m$, for $m = 1, \ldots, M$, so that

$$\frac{F(z)\tilde{B}_m(z)}{\tilde{C}(z)} = \hat{F}_m(z) + \frac{z^{-k}\hat{G}_m(z)}{\tilde{C}(z)} \tag{16}$$

$$\iff F(z)\tilde{B}_m(z) = \hat{F}_m(z)\tilde{C}(z) + z^{-k}\hat{G}_m(z), \tag{17}$$

implying

$$\frac{F(z)\tilde{B}_m(z)}{\tilde{C}(z)} u_{t+k}^{(m)} = \hat{F}_m(z) u_{t+k}^{(m)} + \frac{\hat{G}_m(z)}{\tilde{C}(z)} u_t^{(m)}, \tag{18}$$

yielding

$$\mathbb{E}\left(\frac{F(z)\tilde{B}_m(z)}{\tilde{C}(z)} u_{t+k}^{(m)} \mid \mathcal{F}_t \right) = \mathbb{E}\left(\hat{F}_m(z) u_{t+k}^{(m)} \mid \mathcal{F}_t \right) + \frac{\hat{G}_m(z)}{\tilde{C}(z)} u_t^{(m)}, \tag{19}$$

as $u_{t-n}^{(m)}$, for $n \geq 0$, are known at time t. We may summarize this in the following proposition.

Proposition 1 (Prediction)**.** *Consider the ARMAX model*

$$\tilde{A}(z)y_t = \sum_{m=1}^{M} \tilde{B}_m(z)u_t^{(m)} + \tilde{C}(z)e_t, \tag{20}$$

with the polynomial $\tilde{A}$ being stable and $\tilde{C}$ invertible. Then, the optimal linear k-step predictor is given by

$$\hat{y}_{t+k} = \mathbb{E}\left(y_{t+k} \mid \mathcal{F}_t\right) \tag{21}$$

$$= \sum_{m=1}^{M} \mathbb{E}\left(\hat{F}_m(z)u_{t+k}^{(m)} \mid \mathcal{F}_t\right) + \frac{\hat{G}_m(z)}{\tilde{C}(z)}u_t^{(m)} + \frac{G(z)}{\tilde{C}(z)}y_t \tag{22}$$

where the polynomials F and $\hat{F}_m$, for $m = 1, \ldots, M$, all are of order $k-1$ and (F, G) and $(\hat{F}_m, \hat{G}_m)$ satisfy

$$\tilde{C}(z) = F(z)\tilde{A}(z) + z^{-k}G(z) \tag{23}$$

$$F(z)\tilde{B}_m(z) = \hat{F}_m(z)\tilde{C}(z) + z^{-k}\hat{G}_m(z). \tag{24}$$

Remark 1 (Deterministic and known inputs)**.** *Note that for the case of deterministic and known input signals $u^{(m)}$, the expectation collapses to*

$$\mathbb{E}\left(\hat{F}_m(z)u_{t+k}^{(m)} \mid \mathcal{F}_t\right) = \hat{F}_m(z)u_{t+k}^{(m)} = \sum_{n=0}^{k-1} \hat{f}_n^{(m)} u_{t+k-n}^{(m)}. \tag{25}$$

Also, for this case, we do not need to compute $(\hat{F}_m, \hat{G}_m)$, but may include the contribution from the input signals using the polynomials $F(z)\tilde{B}_m(z)$ and $\tilde{C}(z)$ directly, allowing us to compute the prediction as

$$\hat{y}_{t+k} = \sum_{m=1}^{M} \frac{F(z)\tilde{B}_m(z)}{\tilde{C}(z)}u_{t+k}^{(m)} + \frac{G(z)}{\tilde{C}(z)}y_t. \tag{26}$$

Remark 2 (Stochastic inputs)**.** *If the inputs are stochastic, but predictable, e.g., using an ARMA structure, the expectation is instead given by*

$$\mathbb{E}\left(\hat{F}_m(z)u_{t+k}^{(m)} \mid \mathcal{F}_t\right) = \sum_{n=0}^{k-1} \hat{f}_n^{(m)} \hat{u}_{t+k-n}^{(m)}. \tag{27}$$

where $\hat{u}_{t+k-n}^{(m)} = \mathbb{E}\left(u_{t+k-n}^{(m)} \mid \mathcal{F}_t\right)$, for $n = 0, \ldots, k-1$ are predicted input values.

Remark 3. *It may be noted that for deterministic and known input signals, the prediction error is given by*

$$y_{t+k} - \hat{y}_{t+k} = F(z)e_{t+k}, \tag{28}$$

which is an MA$(k-1)$ process. If, instead, the input has to be predicted, each predicted input signal will contribute with a prediction error with MA$(k-1)$ structure. Thus, for this case, the prediction error is a sum of $M+1$ independent MA$(k-1)$ processes.