1st Assignment
(2nd Part)
Time Series Econometrics

Mohamed Ayadi (18-814-996)
Aram Jivan Chetirian (19-607-258)
Maximilian Droschl (21-616-412)

11th of March, 2025

Contents

1 Optimal MSE Forecasts

First, we need to show that the optimal MA(3) forecast is given by eq.(1):

$$\hat{X}^*_{t+h|t} = \begin{cases} \theta_1\varepsilon_t + \theta_2\varepsilon_{t-1} + \theta_3\varepsilon_{t-2}, & h = 1 \\ \theta_2\varepsilon_t + \theta_3\varepsilon_{t-1}, & h = 2 \\ \theta_3\varepsilon_t, & h = 3 \\ 0, & h \geq 4 \end{cases} \tag{1}$$

Under quadratic loss (equivalent to minimizing the MSE), the optimal forecast, given time-t information $\mathcal{F}_t$, is given by (slide 3-52)

$$\hat{X}^*_{t+h|t} = \mathbb{E}[X_{t+h}|\mathcal{F}_t],$$

which we will define as $\mathbb{E}_t[X_{t+h}]$ from now on. Taking conditional expectations of our MA(3) for each lag h leads to the following solution:

$$\begin{aligned} \mathbb{E}_t[X_{t+1}] &= \mathbb{E}_t[\varepsilon_{t+1} + \theta_1\varepsilon_t + \theta_2\varepsilon_{t-1} + \theta_3\varepsilon_{t-2}] && (2)\\ &= \mathbb{E}_t[\varepsilon_{t+1}] + \mathbb{E}_t[\theta_1\varepsilon_t] + \mathbb{E}_t[\theta_2\varepsilon_{t-1}] + \mathbb{E}_t[\theta_3\varepsilon_{t-2}] && (3)\\ &= 0 + \theta_1\varepsilon_t + \theta_2\varepsilon_{t-1} + \theta_3\varepsilon_{t-2} = \hat{X}^*_{t+1|t}, && (4) \end{aligned}$$

since $\mathbb{E}_t[\varepsilon_{t+h}] = 0$. Similarly we have that

$$\begin{aligned} \mathbb{E}_t[X_{t+2}] &= \mathbb{E}_t[\varepsilon_{t+2} + \theta_1\varepsilon_{t+1} + \theta_2\varepsilon_t + \theta_3\varepsilon_{t-1}] && (5)\\ &= \mathbb{E}_t[\varepsilon_{t+2}] + \mathbb{E}_t[\theta_1\varepsilon_{t+1}] + \mathbb{E}_t[\theta_2\varepsilon_t] + \mathbb{E}_t[\theta_3\varepsilon_{t-1}] && (6)\\ &= 0 + 0 + \theta_2\varepsilon_t + \theta_3\varepsilon_{t-1} = \hat{X}^*_{t+2|t} && (7) \end{aligned}$$

$$\begin{aligned} \mathbb{E}_t[X_{t+3}] &= \mathbb{E}_t[\varepsilon_{t+3} + \theta_1\varepsilon_{t+2} + \theta_2\varepsilon_{t+1} + \theta_3\varepsilon_t] && (8)\\ &= \mathbb{E}_t[\varepsilon_{t+3}] + \mathbb{E}_t[\theta_1\varepsilon_{t+2}] + \mathbb{E}_t[\theta_2\varepsilon_{t+1}] + \mathbb{E}_t[\theta_3\varepsilon_t] && (9)\\ &= 0 + 0 + 0 + \theta_3\varepsilon_t = \hat{X}^*_{t+3|t} && (10) \end{aligned}$$

$$\begin{aligned} \mathbb{E}_t[X_{t+4}] &= \mathbb{E}_t[\varepsilon_{t+4} + \theta_1\varepsilon_{t+3} + \theta_2\varepsilon_{t+2} + \theta_3\varepsilon_{t+1}] && (11)\\ &= \mathbb{E}_t[\varepsilon_{t+4}] + \mathbb{E}_t[\theta_1\varepsilon_{t+3}] + \mathbb{E}_t[\theta_2\varepsilon_{t+2}] + \mathbb{E}_t[\theta_3\varepsilon_{t+1}] && (12)\\ &= 0 + 0 + 0 + 0 = \hat{X}^*_{t+4|t} && (13) \end{aligned}$$

2 Approach 2 - Projection Method

We start by deriving the autocovariance function of the MA(3) process:

$$\gamma(h) = \begin{cases} \sigma_w^2 \sum_{j=0}^{3-h} \theta_j \theta_{j+h}, & 0 \leq h \leq 3 \\ 0, & h > 3 \end{cases} \tag{14}$$

$$\begin{aligned}
\gamma(h) &= \mathrm{Cov}(X_{t+h}, X_t) = \mathbb{E}[(X_{t+h} - \mu_{t+h})(X_t - \mu_t)] \\
&= \mathbb{E}[X_{t+h} X_t], \quad \textit{as } \mathbb{E}[\theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1} + \theta_3 \varepsilon_{t-2}] = 0 \\
&= \mathbb{E}[(\theta_1 \varepsilon_{t-1+h} + \theta_2 \varepsilon_{t-2+h} + \theta_3 \varepsilon_{t-3+h} + \varepsilon_{t+h})(\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \varepsilon_t)] \\
&= \theta_1^2 \, \mathbb{E}[\varepsilon_{t+h-1} \varepsilon_{t-1}] + \theta_1 \theta_2 \, \mathbb{E}[\varepsilon_{t+h-1} \varepsilon_{t-2}] + \theta_1 \theta_3 \, \mathbb{E}[\varepsilon_{t+h-1} \varepsilon_{t-3}] + \theta_1 \, \mathbb{E}[\varepsilon_{t+h-1} \varepsilon_t] \\
&+ \theta_2 \theta_1 \, \mathbb{E}[\varepsilon_{t+h-2} \varepsilon_{t-1}] + \theta_2^2 \, \mathbb{E}[\varepsilon_{t+h-2} \varepsilon_{t-2}] + \theta_2 \theta_3 \, \mathbb{E}[\varepsilon_{t+h-2} \varepsilon_{t-3}] + \theta_2 \, \mathbb{E}[\varepsilon_{t+h-2} \varepsilon_t] \\
&+ \theta_3 \theta_1 \, \mathbb{E}[\varepsilon_{t+h-3} \varepsilon_{t-1}] + \theta_3 \theta_2 \, \mathbb{E}[\varepsilon_{t+h-3} \varepsilon_{t-2}] + \theta_3^2 \, \mathbb{E}[\varepsilon_{t+h-3} \varepsilon_{t-3}] + \theta_3 \, \mathbb{E}[\varepsilon_{t+h-3} \varepsilon_t] \\
&+ \theta_1 \, \mathbb{E}[\varepsilon_{t+h} \varepsilon_{t-1}] + \theta_2 \, \mathbb{E}[\varepsilon_{t+h} \varepsilon_{t-2}] + \theta_3 \, \mathbb{E}[\varepsilon_{t+h} \varepsilon_{t-3}] + \mathbb{E}[\varepsilon_{t+h} \varepsilon_t].
\end{aligned}$$

As $\mathbb{E}[\varepsilon_i \varepsilon_j] = \sigma_w^2$ if $i = j$ and $= 0$ otherwise, we can further simplify for each case of h:

$$= \begin{cases} \sigma^2(1 + \theta_1^2 + \theta_2^2 + \theta_3^2), & h = 0 \\ \sigma^2(\theta_1 + \theta_1\theta_2 + \theta_2\theta_3), & h = 1 \\ \sigma^2(\theta_2 + \theta_1\theta_3), & h = 2 \\ \sigma^2\theta_3, & h = 3 \\ 0, & h = 4, \end{cases} \tag{15}$$

which can then be generalised to equation (14).

3 Comparison of Forecasting Methods

Horizon	Optimal	Approach1	Approach2
1	1.0280785	1.028078	1.027005
2	1.0156999	1.318465	1.317595
3	0.9866151	1.291512	1.292853

Based on the table reporting the MSEs of the different approaches as well as the plots comparing either the different methods throughout the different forecast horizons or throughout different M levels, one can note the following insights:

1. Firstly, the two first approaches yields nearly the same level of accuracies. This is aligned with the fact that an $MA(q)$ process has an $AR(\infty)$ representation and vice-versa. Thus, using recursively reconstructed an error sequence to forecast h steps ahead should have an AR representation too and therefore one should also be able to rely on the projection method to build a linear forecast. Since the MSE is evaluated on a large sample ($K = 1000$) and since the two approach are equivalents, their MSE should converge to the same values for each given forecast horizon. This can be seen in Figure 2, where the MSE is plotted as a function of the forecasting horizon ($M = 1, ..., 25$ in our case). In Figure 1, it is clear that since the forecasts are not based on the infeasible series, both methods necessarily underperform the optimal MSE approach, which is the theoretical benchmark. Therefore, the insight from this exercise is that a $MA(3)$ process can be forecast as an AR process and vice versa.
2. Changing the M levels only initially leads to a noticeable improvement in forecast accuracy. Based on the simulation, it can be concluded that beyond the memory length of the process of 3 lags, there are limited practical gains from adding more lags. This is consistent with the course insight that while the $MA(3)$ process should theoretically be represented as an $AR(\infty)$, in practice most of the accuracy gains are already captured by the first $q = 3$ lags. This argument is clearly consistent with the noticeable convergence of the MSEs in the table showing the effect of M. By plotting the MSEs for each time horizon h as a function of M (which was not asked in the exercise, but was something we looked at), one could clearly see that lags greater than 3 did not significantly improve the MSEs anymore.
3. Lastly, commenting on the decreasing MSE of the projection method, one can bring forth the following interpretation. The MSE declines because the forecasting horizon is closer to the length of the process memory. If this interpretation is correct, one could conclude that there is a rule of thumb for determining the best forecast horizon.

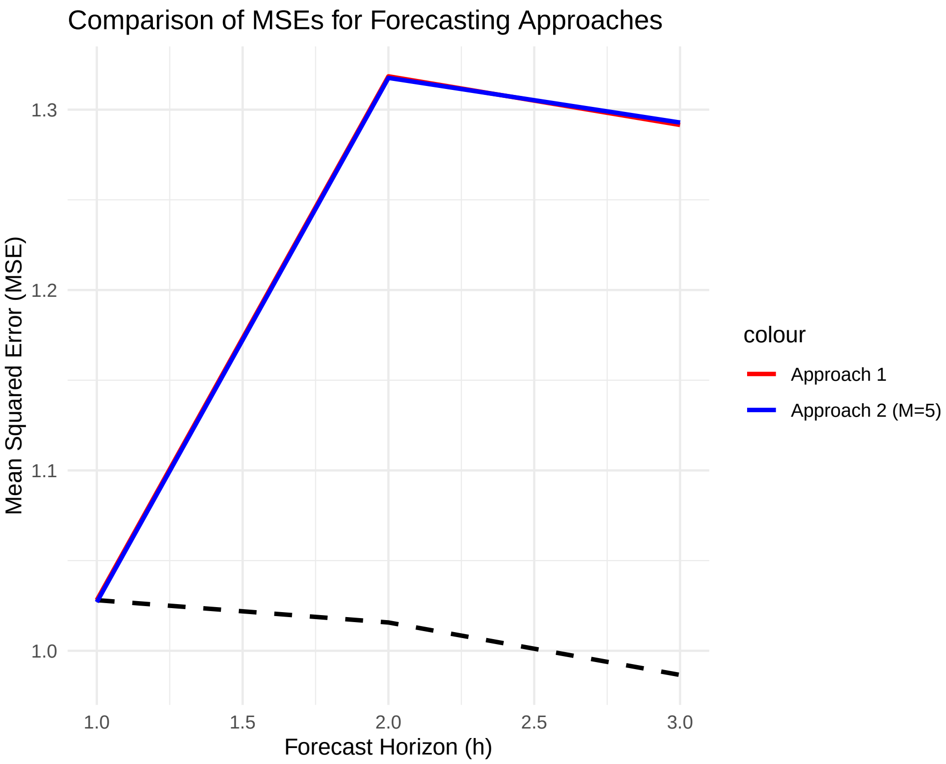

Figure 1: MSE Comparison

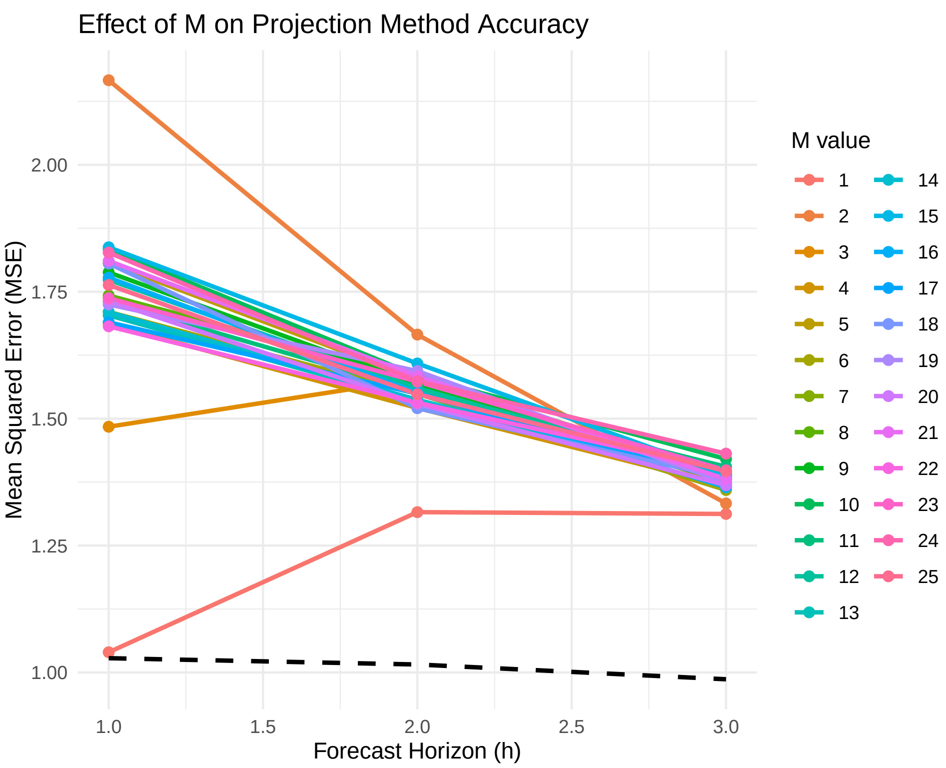

Figure 2: Effect of M