Exercise 1

Suppose that you have data on variable y_t for t = 1, 2, …, 50, estimate an AR(2) model and get the following sample regression equation[1]:

$$\hat{y}_t = \hat{\varphi}_0 + \hat{\varphi}_1 y_{t-1} + \hat{\varphi}_2 y_{t-2} = 0.045 + 1.182 y_{t-1} - 0.219 y_{t-2}$$

Obtain forecasts for one, two and three periods ahead (i.e., for t = 51, 52, 53), given that the last two available observations are y_{49} = 1.63 and y_{50} = 1.72.

To forecast y in t = 50 for one period ahead, we project the population regression equation,

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t$$

one period ahead,

$$y_{t+1} = \varphi_0 + \varphi_1 y_t + \varphi_2 y_{t-1} + \varepsilon_{t+1} \quad \rightarrow \quad y_{51} = \varphi_0 + \varphi_1 y_{50} + \varphi_2 y_{49} + \varepsilon_{51}$$

The corresponding optimal forecast is the conditional expected value of y_{51},

$$E(y_{t+1} \mid \Omega_t) = \varphi_0 + \varphi_1 E(y_t \mid \Omega_t) + \varphi_2 E(y_{t-1} \mid \Omega_t) + E(\varepsilon_{t+1} \mid \Omega_t)$$
$$\rightarrow \quad E(y_{51} \mid \Omega_{50}) = \varphi_0 + \varphi_1 E(y_{50} \mid \Omega_{50}) + \varphi_2 E(y_{49} \mid \Omega_{50}) + E(\varepsilon_{51} \mid \Omega_{50})$$

where the information set, Ω_t, is supposed to contain the realizations of y_t and ε_t up to and including time t,

$$\Omega_t = \{y_1, y_2, ..., y_t; \varepsilon_1, \varepsilon_2, ..., \varepsilon_t\} \quad \rightarrow \quad \Omega_{50} = \{y_1, y_2, ..., y_{50}; \varepsilon_1, \varepsilon_2, ..., \varepsilon_{50}\}$$

Since y_{49} and y_{50} are in the information set, i.e., they are known, their conditional expected values are themselves,

$$E(y_{49} \mid \Omega_{50}) = y_{49} \quad , \quad E(y_{50} \mid \Omega_{50}) = y_{50}$$

ε_{51}, however, is not in the information set, so its conditional expected value is the same as its unconditional expected value, i.e.,

$$E(\varepsilon_{51} \mid \Omega_{50}) = E(\varepsilon_{51}) = 0$$

Accordingly, our forecast for one period ahead is

$$E(y_{51} \mid \Omega_{50}) = \varphi_0 + \varphi_1 y_{50} + \varphi_2 y_{49}$$

So far, we have tacitly assumed that the φ_0, φ_1, φ_2 population parameters are known. In reality, however, they are unknown, so to make this forecasting procedure operational, we replace φ_0, φ_1, φ_2 with their estimates. Hence,

[1] It can be checked by calculating the characteristic roots that the process represented by this AR(2) model is stationary.

$$
\begin{aligned}
E(y_{51} \mid \Omega_{50}) &= \varphi_0 + \varphi_1 y_{50} + \varphi_2 y_{49} \\
&\approx \hat{\varphi}_0 + \hat{\varphi}_1 y_{50} + \hat{\varphi}_2 y_{49} = 0.045 + 1.182 y_{50} - 0.219 y_{49} \\
&= 0.045 + 1.182 \times 1.72 - 0.219 \times 1.63 = 1.721
\end{aligned}
$$

Likewise, to forecast y in t = 50 for two periods ahead, we project the population regression equation to t = 52,

$$
y_{52} = \varphi_0 + \varphi_1 y_{51} + \varphi_2 y_{50} + \varepsilon_{52} \quad \rightarrow \quad E(y_{52} \mid \Omega_{50}) = \varphi_0 + \varphi_1 E(y_{51} \mid \Omega_{50}) + \varphi_2 E(y_{50} \mid \Omega_{50}) + E(\varepsilon_{52} \mid \Omega_{50})
$$

y_{50} is in the information set but y_{51} is not. Hence, we use the observed value for the former and the previous forecast for the latter, i.e.,

$$
E(y_{50} \mid \Omega_{50}) = y_{50} \quad , \quad E(y_{51} \mid \Omega_{50}) = 1.721
$$

Replacing the unknown population parameters with their estimates again,

$$
\begin{aligned}
E(y_{52} \mid \Omega_{50}) &= \varphi_0 + \varphi_1 E(y_{51} \mid \Omega_{50}) + \varphi_2 y_{50} \\
&\approx \hat{\varphi}_0 + \hat{\varphi}_1 E(y_{51} \mid \Omega_{50}) + \hat{\varphi}_2 y_{50} \\
&= 0.045 + 1.182 \times 1.721 - 0.219 \times 1.72 = 1.703
\end{aligned}
$$

Following the same logic, the forecast of y prepared in t = 50 for three periods ahead is:

$$
\begin{aligned}
& y_{53} = \varphi_0 + \varphi_1 y_{52} + \varphi_2 y_{51} + \varepsilon_{53} \\
& \rightarrow \quad E(y_{53} \mid \Omega_{50}) = \varphi_0 + \varphi_1 E(y_{52} \mid \Omega_{50}) + \varphi_2 E(y_{51} \mid \Omega_{50}) + E(\varepsilon_{53} \mid \Omega_{50}) \\
& \qquad\qquad\qquad\quad = \varphi_0 + \varphi_1 E(y_{52} \mid \Omega_{50}) + \varphi_2 E(y_{51} \mid \Omega_{50}) \\
& \qquad\qquad\qquad\quad \approx \hat{\varphi}_0 + \hat{\varphi}_1 E(y_{52} \mid \Omega_{50}) + \hat{\varphi}_2 E(y_{51} \mid \Omega_{50}) \\
& \qquad\qquad\qquad\quad = 0.045 + 1.182 \times 1.703 - 0.219 \times 1.721 = 1.681
\end{aligned}
$$

Exercise 2

Continuing the previous exercise, suppose that the test period is t = 51, 52, 53 and the true values of y_t in the test period are y_{51} = 1.70, y_{52} = 1.71 and y_{53} = 1.69. Given these true values and the forecasts in Exercise 1, calculate the *ME*, *MAE*, *RMSE*, *MPE* and *MAPE* measures of forecast accuracy for the test period with your calculator.

For the sake of simplicity, in the following solution some of the details were computed in Excel. Note, however, that on the exam you will have no access to a computer, so if you really want to benefit from this exercise, try to reproduce the results with your calculator only.

The observed values, forecasts, forecast errors, absolute errors, squared errors, percent errors and absolute percent errors over the test period, and their sums, are in the following table:

t	y_t	f_t	e_t	$\lvert e_t \rvert$	$e_t^{\,2}$	p_t	$\lvert p_t \rvert$
51	1.700	1.721	-0.021	0.021	0.000441	-1.235%	1.235%
52	1.710	1.703	0.007	0.007	0.000049	0.409%	0.409%
53	1.690	1.681	0.009	0.009	0.000081	0.533%	0.533%
Sum			-0.005	0.037	0.000571	-0.293%	2.177%

In this hypothetical example $T - t^* = 3$, so from the sums the forecast accuracy measures are

$$ME = \frac{1}{T - t^*} \sum_{t=t^*+1}^{T} e_t = \frac{-0.005}{3} = -0.0017$$

$$MAE = \frac{1}{T - t^*} \sum_{t=t^*+1}^{T} | e_t | = \frac{0.037}{3} = 0.0123$$

$$RMSE = \sqrt{\frac{1}{T - t^*} \sum_{t=t^*+1}^{T} e_t^2} = \sqrt{\frac{0.000571}{3}} = 0.0138$$

$$MPE = \frac{1}{T - t^*} \sum_{t=t^*+1}^{T} p_t = \frac{-0.293\%}{3} = -0.0977\%$$

$$MAPE = \frac{1}{T - t^*} \sum_{t=t^*+1}^{T} | p_t | = \frac{2.177\%}{3} = 0.7257\%$$

Let’s now see a ‘real’ forecast project.

Exercise 3

The *t7e3.xlsx* file contains observations on monthly unemployment rate (*URM*) in Australia from Feb 1978 to Jan 2023 (downloaded from ABS, 6/03/2023).

a) Plot *URM* and briefly describe the pattern of the data series.

Launch *RStudio*, create a new project and script, and name both *t7e3*. Import the data from the *URM* sheet and prepare a time series plot of *URM* by executing the following commands:

```
attach(t7e3)
URM = ts(URM, frequency = 12, start = c(1978,2), end = c(2023,1))
plot(URM, main = "Unemployment rate, monthly, percent, Australia", col = "red")
```