Quantitative Methods III - AR models, ADL models and Forecasting

Exercise All of the following questions refer to the model estimates shown in **Table 1**, made on the time series of US inflation, denoted by y_t, and the US unemployment rate, denoted by x_t. Series are available yearly from 1960 to 2021 ($T = 62$), but models were estimated using observations up to 2017.

Table 1

	(1)	(2)	(3)	(4)	(5)	(6)
Indicate Model						
Constant	0.7129**	0.9430***	0.6558*	0.5783	1.4266**	1.0540
	(0.315)	(0.314)	(0.365)	(0.422)	(0.691)	(0.749)
y_{t-1}	0.8157***	1.0385***	1.0979***	1.0871***	1.0098***	0.9612***
	(0.104)	(0.183)	(0.193)	(0.187)	(0.187)	(0.179)
y_{t-2}		-0.2785*	-0.5532**	-0.5882**	-0.2363	-0.1388
		(0.160)	(0.237)	(0.246)	(0.169)	(0.185)
y_{t-3}			0.2287	0.3129		
			(0.239)	(0.233)		
y_{t-4}			0.0633	-0.1044		
			(0.171)	(0.184)		
y_{t-5}				0.1502		
				(0.164)		
x_{t-1}					-0.0881	-0.4810
					(0.118)	(0.393)
x_{t-2}						0.4224
						(0.353)
T	57	??	54	53	56	56
R^2	0.6695	0.6913	0.7116	0.7133	0.6931	0.7098
RSS	146.9	134.9	121.8	118.9	134.1	126.8
Ln(RSS/T)	0.947	0.879	0.813	0.808	0.873	0.817
AIC	1.017	0.986	0.999	1.034	1.016	0.996
BIC	1.089	1.095	1.183	1.257	1.161	???

The values of the two series in the last 5 years are shown below:

Year (t)	Inflation (y_t)	Unemployment (x_t)
2017	2.13	4.36
2018	2.44	3.9
2019	1.81	3.68
2020	1.23	8.09
2021	4.7	5.36

1. Indicate, in the first row of the first table, the type of model estimated in each column (use the conventional abbreviations AR(p) and ADL(p, q), specifying the value of p and q).
2. Indicate the number of observations used in the model (2).
3. Calculate the BIC of the model (6).
4. Which model would you choose based on Akaike's criterion?
5. Which model would you choose based on Bayes' criterion?
6. Obtain the RMSFE of the model shown in column (1) using the SER method.
7. Use the estimates of the model in column (1) to derive the Out-Of-Sample (OOS) predictions one step ahead of inflation from 2018 to 2021.
8. Obtain the RMSFE of the model (1) by exploiting the OOS predictions calculated in the previous point.
9. Provide an estimate of the forecast interval at 95% for 2022 (using the RMSFE).

Solution

1. In the model shown in column (1) y_t is regressed on a constant and on its first lag.

   The regression model is:

   $$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

   It is AutoRegressive model of order 1, AR(1). In a similar way, the models are identified in columns (2), (3) and (4).

   In column (5), y_t is regressed on a constant, its first two lags, and the first lag of x_t. In other words, the regression model is:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \delta_1 x_{t-1} + u_t$$

It is Autoregressive Distributed Lag model ADL(p, q), with $p = 2$ lags of the dependent variable y_t and $q = 1$ lag of the variable x_t, ADL(2,1). Similarly, the model in column (6) is identified as ADL(2,2).

Model names are listed in the first row of the table below.

	(1)	(2)	(3)	(4)	(5)	(6)
Model	AR(1)	AR(2)	AR(4)	AR(5)	ADL(2,1)	ADL(2,2)
Constant	0.7129**	0.9430***	0.6558*	0.5783	1.4266**	1.0540
	(0.315)	(0.314)	(0.365)	(0.422)	(0.691)	(0.749)
y_{t-1}	0.8157***	1.0385***	1.0979***	1.0871***	1.0098***	0.9612***
	(0.104)	(0.183)	(0.193)	(0.187)	(0.187)	(0.179)
y_{t-2}		-0.2785*	-0.5532**	-0.5882**	-0.2363	-0.1388
		(0.160)	(0.237)	(0.246)	(0.169)	(0.185)
y_{t-3}			0.2287	0.3129		
			(0.239)	(0.233)		
y_{t-4}			0.0633	-0.1044		
			(0.171)	(0.184)		
y_{t-5}				0.1502		
				(0.164)		
x_{t-1}					-0.0881	-0.4810
					(0.118)	(0.393)
x_{t-2}						0.4224
						(0.353)
T	57	56	54	53	56	56
R^2	0.6695	0.6913	0.7116	0.7133	0.6931	0.7098
RSS	146.9	134.9	121.8	118.9	134.1	126.8
Ln(RSS/T)	0.947	0.879	0.813	0.808	0.873	0.817
AIC	1.017	0.986	0.999	1.034	1.016	0.996
BIC	1.089	1.095	1.183	1.257	1.161	1.177

2. The observations of the entire time series, available up to 2021, are $T = 62$. However, we know from the text that the models use the observations only up to 2017, thus ignoring the last 4 observations. Also, since model (2) is an AR(2), 2 more observations are lost to compute the first two lags of y_t.

   Therefore, the number of observations actually used to estimate the model is $T = 62-4-2 = 56$. Alternatively, it could be observed that the AR(1) model uses 57 observations, therefore the AR(2) model uses one fewer observations for construction.

3. The BIC is:

$$BIC = ln\left(\frac{RSS}{T}\right) + \frac{k \times ln(T)}{T}$$

The table already provides the value of $ln\left(\frac{RSS}{T}\right)$ and of T for the estimated models, therefore it is simply necessary to identify the value of k in the model of interest. Since model (6) is an ADL(2,2), the number of parameters estimated is $k = 1 + 2 + 2 = 5$. It follows that

$$BIC = 0.817 + \frac{5 \times ln(56)}{56} = 1.177$$

4. The model chosen according to the Akaike's criterion (AIC) would be the AR(2), because it is the one with the smallest AIC.

5. If it is used the Bayes' criterion (BIC), the model chosen would be the AR(1), because it is the one with the smallest BIC.

6. A way to estimate the RMSFE is:

$$RM\hat{S}FE_{SER} = \sqrt{\frac{RSS}{T-k}}$$

For the model (1) we will have:

$$RM\hat{S}FE_{SER} = \sqrt{\frac{146.9}{57-2}} = 1.634$$

7. The AR(1) model is:

$$y_t = 0.7129 + 0.8157 y_{t-1}$$

It is therefore possible to use this model to get Out-Of-Sample (OOS) forecasts one step ahead, as:

$$\hat{y}_{t+1|t} = 0.7129 + 0.8157 y_t$$

For example, the expected value of inflation for $t + 1 = 2018$, based on the information available in $t = 2017$ would be:

$$\hat{y}_{t+1|t} = 0.7129 + 0.8157 \times 2.13 = 2.450$$

This forecast deviates from the observed value of:

$$\tilde{u}_{t+1} = y_{t+1} - \hat{y}_{t+1|t} = 2.44 - 2.45 = -0.01.$$

This value represents the prediction error.

The observed data, the predictions and the forecast errors (even squared) are shown in the following table:

t	Inflation (y_t)	OOS Prediction ($\hat{y}_{t+1\|t}$)	Forecast Error ($\tilde{u}_t$)	$\tilde{u}_t^2$
2017	2.13			
2018	2.44	2.450	-0.010	0.0001
2019	1.81	2.703	-0.893	0.797
2020	1.23	2.189	-0.959	0.920
2021	4.7	1.716	2.984	8.904

8. Another method to estimate RMSFE is based on OOS predictions:

$$RM\hat{S}FE_{OOP} = \sqrt{\frac{1}{P}\sum_{i=1}^{P}\tilde{u}_i^2}$$

In the previous point, the predictions were $P = 4$ and the squared forecast errors add up to 10.621. It follows that:

$$RM\hat{S}FE_{OOP} = \sqrt{\frac{1}{4}\sum_{i=1}^{4}\tilde{u}_i^2} = \sqrt{\frac{1}{4}10.621} = 1.630$$

9. The one step ahead prediction for 2022 is:

$$\hat{y}_{t+1|t} = 0.7129 + 0.8157 \times 4.7 = 4.547$$

The 95% prediction interval for this prediction, using the RMSFE obtained in the previous point, is:

$$[\hat{y}_{t+1|t} \pm z_{1-\alpha/2} \times RM\hat{S}FE_{OOP}] = [4.547 \pm 1.96 \times 1.63] = [1.353; 7.741]$$

With a 95% probability, inflation for 2022 will be between 1.353% and 7.741%.