ETC2410 Assignment 2

Alex Wong, Chelaka Paranahewa, Harjot Channa, Jonas Tiong

Question 2 (31 Marks)

2(a) (4 marks)

$$\begin{aligned}\widehat{HOUSTNSA} = \underset{(4.196)}{92.871} \;-\; \underset{(5.911)}{4.592}\; Jan \;-\; \underset{(5.911)}{1.935}\; Feb \;+\; \underset{(5.934)}{26.184}\; Mar \\ +\; \underset{(5.934)}{41.452}\; Apr \;+\; \underset{(5.934)}{46.786}\; May \;+\; \underset{(5.934)}{46.263}\; Jun \;+\; \underset{(5.934)}{40.937}\; Jul \\ +\; \underset{(5.934)}{38.714}\; Aug \;+\; \underset{(5.934)}{32.252}\; Sep \;+\; \underset{(5.934)}{36.170}\; Oct \;+\; \underset{(5.934)}{15.600}\; Nov\end{aligned} \tag{1}$$

The above linear regression estimates the US monthly 'housing starts' based on the month that is being modelled. The intercept on its own implies that estimated 'housing starts' for the month of December, which means the other values are relative to the 'housing starts' of december. The variables in the linear regression are seasonal dummies which mean they only take a biary value (0 or 1). The β values for the seasonal dummies are the average change in the 'housing starts' relative to the month December.

2(b) (4 marks)

Steps

In order to formulate the linear regression, first we need to determine the intercept: From equation 1 we can determine the values of each month because of the dummy variables. $92.871 - 4.592 = c \rightarrow c = 88.280$, where the LHS is the month of Jan from calculated from equation 1.

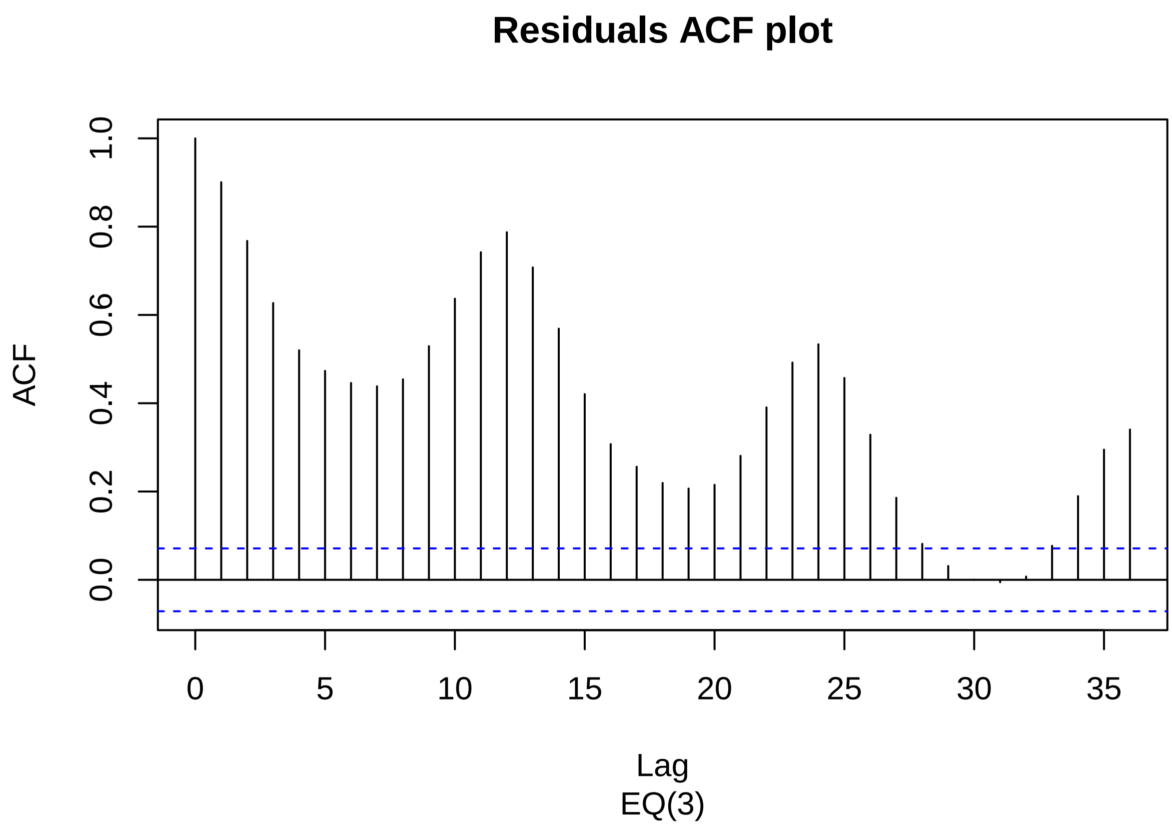
Next we need to determine the β values for Feb - Dec. Since we know the intercept for the new equation, we can substitute it in.

$$\begin{aligned}
92.871 + 1.935 &= 88.280 + \beta_2 \ Feb \\
&\rightarrow \ \beta_2 = 2.656 \\[1em]
92.871 + 26.184 &= 88.280 + \beta_3 \ Mar \\
&\rightarrow \ \beta_3 = 30.776 \\[1em]
92.871 + 41.452 &= 88.280 + \beta_4 \ Apr \\
&\rightarrow \ \beta_4 = 46.044 \\[1em]
92.871 + 46.786 &= 88.280 + \beta_5 \ May \\
&\rightarrow \ \beta_5 = 51.377 \\[1em]
92.871 + 46.263 &= 88.280 + \beta_6 \ Jun \\
&\rightarrow \ \beta_6 = 50.855 \\[1em]
92.871 + 40.937 &= 88.280 + \beta_7 \ Jul \\
&\rightarrow \ \beta_7 = 45.528 \\[1em]
92.871 + 38.714 &= 88.280 + \beta_8 \ Aug \\
&\rightarrow \ \beta_8 = 43.306 \\[1em]
92.871 + 32.252 &= 88.280 + \beta_9 \ Sep \\
&\rightarrow \ \beta_9 = 36.844 \\[1em]
92.871 + 36.170 &= 88.280 + \beta_{10} \ Oct \\
&\rightarrow \ \beta_{10} = 40.762 \\[1em]
92.871 + 15.600 &= 88.280 + \beta_{11} \ Nov \\
&\rightarrow \ \beta_{11} = 20.192 \\[1em]
92.871 &= 88.280 + \beta_{12} \ Dec \\
&\rightarrow \ \beta_{12} = 4.592
\end{aligned}$$

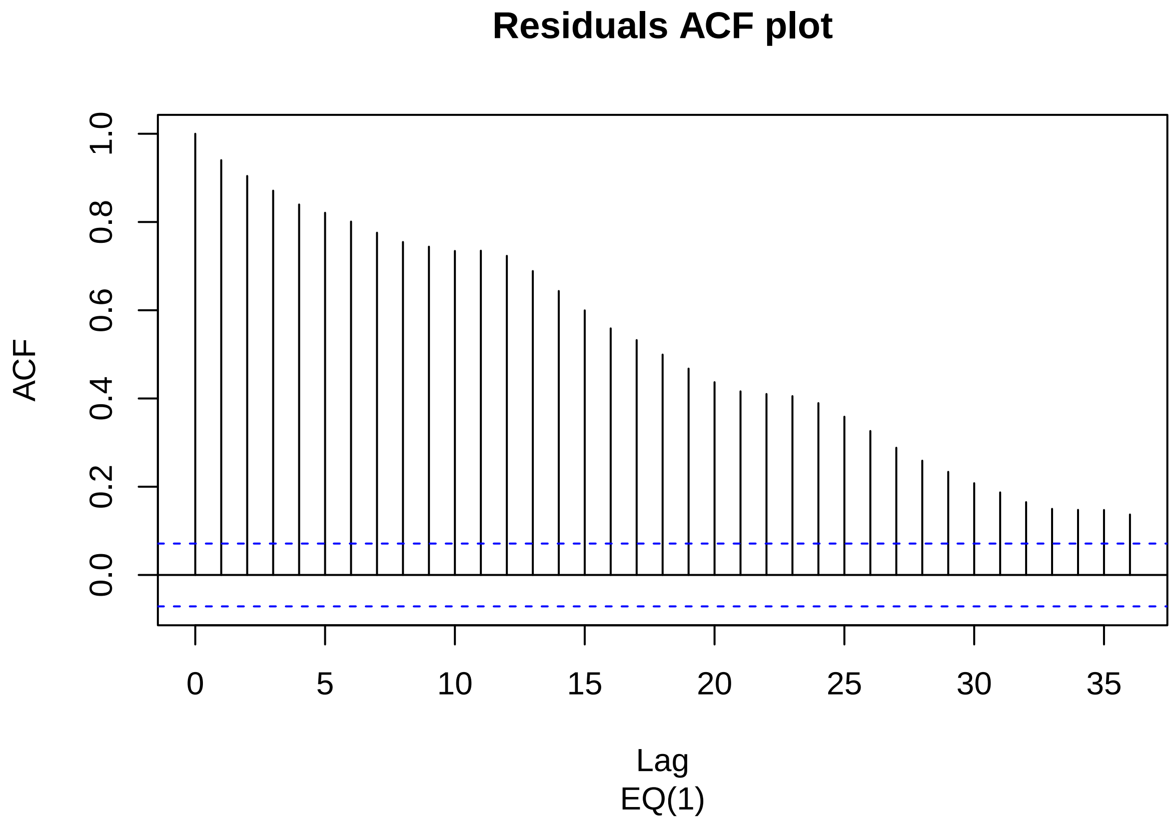
$$\widehat{HOUSTNSA} = 88.280 \ + \ 2.656 \ Feb \ + \ 30.776 \ Mar \ + \ 46.044 \ Apr \\ + \ 51.377 \ May \ + \ 50.855 \ Jun \ + \ 45.528 \ Jul \ + \ 43.306 \ Aug \\ + \ 36.844 \ Sep \ + \ 40.762 \ Oct \ + \ 20.192 \ Nov \ + \ 4.592 \ Dec \tag{2}$$

2(c) (6 marks)

$$\widehat{HOUSTNSA} = \underset{(1.377)}{119.3} \tag{3}$$



The residual ACF plot for the linear model HOUSTNSA ~ 1 shows a few things. Firstly, each non seasonal lag (lag $\neq \{12, 24, 36\}$) have a large positive spike which is the tell tail sign of the existance of a trend. Secondly, the seasonal lags have large positive spikes which indicates that there is also seasonality in the data. So we can attribute the gradual decrease in the lag values is because of trend while the scallop pattern is due to the seasonality.



The residual ACF plot for the linear model EQ(1), shows a significant difference compared to the residual ACF plot for the linear model EQ(3). The linear model EQ(1), has no seasonality in the residual acf plot since the linear model has already captured it. What is remaining is the trend as depicted by decreasing values as the lags increase. Thus showing that the seasonal dummy variables have improved the model.

2(d) (9 marks)

$$\begin{aligned}\widehat{HOUSTNSA} = \underset{(2.165)}{-14.042} \; - \; \underset{(2.003)}{11.014}\; Jan \; - \; \underset{(2.091)}{20.084}\; Feb \; + \; \underset{(2.179)}{47.498}\; Mar \\ + \; \underset{(2.652)}{40.081}\; Apr \; + \; \underset{(2.360)}{28.510}\; May \; + \; \underset{(2.198)}{21.153}\; Jun \; + \; \underset{(2.123)}{15.297}\; Jul \\ + \; \underset{(2.067)}{17.297}\; Aug \; + \; \underset{(2.095)}{13.490}\; Sep \; + \; \underset{(2.049)}{22.810}\; Oct \; + \; \underset{(2.168)}{0.347}\; Nov \\ + \; \underset{(0.036)}{0.775}\; HOUSTNSA_{t-1} \; + \; \underset{(0.036)}{0.177}\; HOUSTNSA_{t-2}\end{aligned} \tag{4}$$

$$
\begin{aligned}
H_0 &: \underset{i\in\{1,2,3,4,5,6,7,8,9,10,11\}}{\forall} \beta_i = 0 \\
H_1 &: \underset{i\in\{1,2,3,4,5\}}{\exists} \beta_i \neq 0 \text{ at least one regressor coef is zero} \\
\text{Significance Level} &: \alpha = 0.05 \\
\text{Unresticted Model} &: \widehat{HOUSTNSA} = \underset{(2.165)}{-14.042} \;-\; \underset{(2.003)}{11.014}\; Jan \;-\; \underset{(2.091)}{20.084}\; Feb \\
&\quad + \underset{(2.179)}{47.498}\; Mar \;+\; \underset{(2.652)}{40.081}\; Apr \;+\; \underset{(2.360)}{28.510}\; May \\
&\quad + \underset{(2.198)}{21.153}\; Jun \;+\; \underset{(2.123)}{15.297}\; Jul \;+\; \underset{(2.067)}{17.297}\; Aug \\
&\quad + \underset{(2.095)}{13.490}\; Sep \;+\; \underset{(2.049)}{22.810}\; Oct \;+\; \underset{(2.168)}{0.347}\; Nov \\
&\quad + \underset{(0.036)}{0.775}\; HOUSTNSA_{t-1} \;+\; \underset{(0.036)}{0.177}\; HOUSTNSA_{t-2} \\
\text{Resticted Model} &: \widehat{HOUSTNSA} = \underset{(1.970)}{14.598} \;+\; \underset{(0.035)}{1.110}\; HOUSTNSA_{t-1} \;-\; \underset{(0.035)}{0.232}\; HOUSTNSA_{t-2} \\
\text{Test stat and null dist} &: \frac{(SSR_R - SSR_{UR})}{SSR_{UR}} \frac{(n-k-1)}{q} \sim F_{(q,n-k-1)} = F_{5,468} \\
F_{calc} &= 74.7468276 \\
F_{crit} &= 2.0101347 \\
\text{Decision rule} &: \text{reject } H_0 \text{ if } t_{calc} > F_{crit} \\
\text{Decision} &: \text{Since } 74.7468276 > 2.0101347, \text{ reject } H_0
\end{aligned}
$$

In conclusion, at 0.05 level of significance, we reject the null hypothesis that the seasonal dummies (Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct) are jointly insignificant in effecting the 'Housing Starts' in favour of the alternative hypothesis that at least one of these variables are significant.

2(e) (8 marks)