# ISyE4031 Regression and Forecasting
## Practice Problems 3 Solutions
## Spring 2016

1. We begin by calculating the estimated residual for the last observation in the series.

$\hat{y}_{30} = 10 + 3.2(4.75) + 1.1(8.5) + 1.5(30) = 79.55$

$\hat{\varepsilon}_{30} = y_{30} - \hat{y}_{30} = 85 - 79.55 = 5.45$

$\hat{\varepsilon}_{31} = 0.74\,\hat{\varepsilon}_{30} = 0.74(5.45) = 4.033$

$\hat{y}_{31} = 10 + 3.2(6.9) + 1.1(4.75) + 1.5(31) + 4.033 = 87.838.$

2. a. $\hat{y}_{87}(84) = l_{84} + 3b_{84} = 12.66 + 3(0.029) = 12.747.$

b. $\hat{y}_{86}(84) = l_{84} + 2b_{84} + sn_{74} = 13.22 + 2(-0.006) - 2.38 = 10.828.$

c. $l_{85} = l_{84} + b_{84} + \alpha(y_{85} - (l_{84} + b_{84} + sn_{73})) = 13.22 - 0.006 + 0.2(13.05 - (13.22 - 0.006 - 1.93))$
$= 13.5672$

$b_{85} = -0.006 + (.2)(.3)(13.05 - (13.22 - 0.006 - 1.93)) = 0.09996$

$sn_{85} = -1.93 + (.8)(.4)(13.05 - (13.22 - 0.006 - 1.93)) = -1.36488.$

3. a. The growth rate smoothing constant, $\gamma$, can be negative, if the time series is decreasing over time. False

b. iii. $[10.5 \pm (1.96)(1.5)(\sqrt{1.02})]$.

c. The Holt's Trend Corrected (Double) exponential smoothing method should be selected, since it minimizes MAPE, MAD, and MSD. $\{y_t\}$ is probably a time series that has a linear trend, but doesn't have a seasonal variation.

d. The initial seasonal factor for Quarter 4: $sn_0 = \dfrac{-6.1 - 2.9 - 4.5}{3} = -4.5$

4. MA(2) model: $\hat{y}_t = 35 + \hat{a}_t - 0.52\,\hat{a}_{t-1} + 0.65\,\hat{a}_{t-2}$

a. $\hat{y}_1 = 35 + \hat{a}_1 - 0.52\,\hat{a}_0 + 0.65\,\hat{a}_{-1}$

Since $\hat{a}_1 = 0$ (future), $\hat{a}_0 = 0$ (no $y_0$), and $\hat{a}_{-1} = 0$ (no $y_{-1}$) => $\hat{y}_1 = 35 + 0 - 0 + 0 = 35$,

then $\hat{a}_1 = y_1 - \hat{y}_1 = 39.9 - 35 = 4.9$.

$\hat{y}_2 = 35 + \hat{a}_2 - 0.52\,\hat{a}_1 + 0.65\,\hat{a}_0$

Since $\hat{a}_2 = 0$ (future), $\hat{a}_0 = 0$ (no $y_0$), and $\hat{a}_1 = 4.9$ => $\hat{y}_2 = 35 + 0 - 0.5237*4.9 + 0 = 32.452$,

then $\hat{a}_2 = y_2 - \hat{y}_2 = 31.9 - 32.452 = -0.552$.

$\hat{y}_3 = 35 + \hat{a}_3 - 0.52\,\hat{a}_2 + 0.65\,\hat{a}_1$

$= 35 + 0 + 0.52*0.552 + 0.65*4.9 = 38.472.$

b. $\hat{y}_{151} = 35 + \hat{a}_{151} - 0.52\,\hat{a}_{150} + 0.65\,\hat{a}_{149} = 35 + 0 + 0.52*2.28 - 0.65*3.32 = 34.0276.$

5. a. ARIMA(1,1) model. Theoretical autocorrelation function, $\rho_1 = \dfrac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\theta_1\phi_1}$ and

$\rho_2 = \phi_1\rho_1$. The estimated parameters, $\hat{\phi}_1 = 0.8$ and $\hat{\theta}_1 = 0.2$.

$$r_1 = \hat{\rho}_1 = \frac{(1-0.8*0.2)(0.8-0.2)}{1+0.2^2-2(0.8)(0.2)} = 0.7.$$

$$r_2 = \hat{\rho}_2 = 0.8*0.7 = 0.56.$$

b. ARIMA(1,1) model: $\delta = \mu(1-\phi_1) \Rightarrow \hat{\mu} = 25/(1-0.8) = 125$.

c. AR(2) model, $\hat{\phi}_1 = 0.6$ and $\hat{\phi}_2 = 0.3$: No invertibility condition. All stationarity conditions are satisfied:

i. $\phi_1 + \phi_2 < 1 \Rightarrow 0.6 + 0.3 = 0.9 < 1$

ii. $\phi_2 - \phi_1 < 1 \Rightarrow 0.3 - 0.6 = -0.3 < 1$

iii. $|0.3| < 1$.

6. I. d. Nonseasonal ARIMA(1,1): $z_t = \delta + \phi_1 z_{t-1} + a_t - \theta_1 a_{t-1}$

II. b. Nonseasonal MA(1): $z_t = \delta + a_t - \theta_1 a_{t-1}$, Seasonal ARIMA(1,1): $z_t = \delta + \phi_{1,L}\, y_{t-12} + a_t - \theta_{1,L}\, a_{t-12}$

7. Estimation, diagnostics, and model assumptions.

a. Significance: The nonseasonal model parameters, $\phi_1$ and $\phi_2$ are significant (We reject $H_0$: $\phi_1 = 0$ and $H_0$: $\phi_2 = 0$), since corresponding $p$-values are $0 < 0.05$ and $0.005 < 0.05$, respectively. The seasonal model parameter, $\phi_{1,12}$ is strongly significant (We reject $H_0$: $\theta_{1,12} = 0$), since corresponding $p$-value $= 0 < 0.05$.

However, $\phi_3$ is not significant (We fail to reject $H_0$: $\phi_3 = 0$) since $p$-value $= 0.612 > 0.05$. Also, the constant, $\delta$, is not significant (We fail to reject $H_0$: $\delta = 0$) since $p$-value $= 0.832 > 0.05$. Nonsignificant $\phi_3$ implies that the tentative model, nonseasonal AR(3) is not exactly the right model, it should be modified (removed from the model and re-ran).

b. Model adequacy: We look at the Ljung-Box statistic for model inadequacy. The $p$-values for all lags, $k$ = 12, 24, 36, 48, are greater than 0.05 (i.e., 0.276, 0.481, 0.407, and 0.558), meaning that we do not reject $H_0$: no signs of remaining autocorrelation in the residuals, and hence the model is adequate.

c. Random shock assumptions:

- $E(\varepsilon_i) = 0$: Violated. Mean of residuals = 0.0002528 (from the A-D test result). Close to zero, but not zero.
- Normality: Not violated. Residuals (random shocks) have a normal distribution, since $p$-value of A-D test is $0.887 > 0.05$. We don't reject $H_0$: Random shocks are normal.

- Independence: Can be considered independent. The RSAC and the RSPAC both have no spikes out of the bounds, implying that the residuals are stationary and independent. Not a strong result, though.

- Identical distribution: Seems violated. Residual vs fits plot depicts some sort of pattern. Also, there are some unusual observations.

8. a. $\ln y_t = \ln \beta_0 \beta_1^t e^{\varepsilon}$ => $y_t^* = \ln \beta_0 + t \ln \beta_1 + \varepsilon \ln e$.

Since $\hat{y}_t^* = 10 - 0.05\,t$, $\ln \hat{\beta}_0 = 10$ and $\ln \hat{\beta}_1 = -0.05$ => $\hat{\beta}_1 = e^{-0.05} = 0.951229$. It means $y_t$ is estimated to be approximately 0.951229 times $y_{t-1}$. Thus, $100(\hat{\beta}_1 - 1)\% = -4.87706\%$ is the percentage change (decrease) from $y_{t-1}$ to $y_t$.

b. Sum of initial seasonal factors should be zero. Hence, $sn_0 = -83$.

c. False

d. Time-series regression model: there is a linear trend and seasonality where $y_t$ : Millions of dollars of beverage shipments in month $t$. We use $t = 1,2,\ldots,180$ for the linear trend, and 11 indicator variables for the seasonal factors.

$y = \beta_0 + \beta_1 t + \beta_2 M_2 + \beta_3 M_3 + \ldots + \beta_{12} M_{12} + \varepsilon$, where $M_1$: January is the base month.

$$M_j = \begin{cases} 1, & \text{if } t \text{ is month } j \\ 0, & \text{otherwise} \end{cases} \quad , j = 2,3,\ldots,12.$$

e. Half−length = (13.78− (−7.78))/2 = 10.78 => $SE$ = 10.78/1.96 = 5.5.

Midpoint = 13.78−10.78 = 3.

A 99% prediction interval: $[3 \pm (2.576)(5.5)] = [-11.168, 17.168]$.