TS Module 14 Model diagnostics residual analysis practice problems

(The attached PDF file has better formatting.)

Cryer and Chan chapter 8 model diagnostics residual analysis

** Exercise 14.1: MA(2) process

An AR(2) process is $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \theta_0 + e_t$.

$\phi_1 = 0.5$, $\phi_2 = 0.3$, and $\theta_0 = 1$

A. What is the expected value for Y_T in terms of Y_{t-1} and Y_{t-2}?
B. What is the residual in Period T in terms of Y_T, Y_{t-1}, and Y_{t-2}?
C. If $Y_1 = 4$, $Y_2 = 6$, and $Y_3 = 5$, what is e_3?

Part A: The expected value of $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \theta_0 = 0.5\ Y_{t-1} + 0.3\ Y_{t-2} + 1$.

Part B: The residual in Period T = $\epsilon_t = Y_t - (0.5\ Y_{t-1} + 0.3\ Y_{t-2} + 1)$.

Part C: $5 = 0.5 \times 4 + 0.3 \times 6 + 1 + e_3 = 4.8 \Rightarrow e_3 = -0.2$

This problem is simple. More complex problems use an ARMA(1,1) process. You figure out the residual in Period T from the actual and expected values and then use the residual to forecast the Period T+1 value.

** Exercise 14.2: Residuals

If a stationary ARMA process with *n* elements is correctly specified, which of the following are true?

A. All the residuals = 0
B. The sum of the residuals = 0
C. $\sum$ (residual – mean residual)^2 / (n – 1) = σ^2_e
D. The residuals have (approximately) expected values of zero and constant variance
E. The residuals approach zero asymptotically as n → ∞

Part A: The residual $\epsilon_t = y_t - \hat{y}_t$.

y_t is a random variable with a variance, so ϵ_t is also a variance with a variance. It does not equal zero.

Part B: Each residual has a variance of σ^2_ε. The residuals are independent, so the sum of the residuals has a variance of n × σ^2_ε.

Jacob: In regression analysis, the sum of the residuals is zero. Why is that not true here?

Rachel: If we know the true regression coefficients, the sum of the residuals is not zero. If we don't know the true regression coefficients, ordinary least squares estimation chooses values that force the sum of residuals to be zero. This exercise assumes the ARMA process is correctly specified: that is, the autoregressive and moving average parameters are correct, not the estimates from a sample.

Part C: As n → ∞, $\sum$ (residual – mean residual)^2 / (n – 1) = σ^2_e. In a finite sample with n < ∞, the equation is not exact.

Part D: The statement is correct. Each estimated value is the true expected value, so the residuals have expected values of zero. The variance of the residuals is constant: even if the residual is large in Period T, the residual in Period T+1 has an expected value of zero and the same variance as other residuals.

Part E: The residuals have a constant variance; they don't approach zero asymptotically.

See Cryer & Chan, page 176: "If the model is correctly specified and the parameter estimates are reasonably close to the true values, then the residuals should have nearly the properties of white noise. They should behave roughly like independent, identically distributed normal variables with zero means and common standard deviations."

** Exercise 14.3: An AR(1) process of 100 values (n = 100) has $\phi = 0.9$.

A. What is the standard deviation of $\hat{r}_1$?

B. What is the standard deviation of $\hat{r}_2$?

C. What is the standard deviation of $\hat{r}_{40}$?

Part A: The standard deviation of $\hat{r}_1$ is $\phi / \sqrt{n} = 0.9 / 10 = 0.090$.

See Cryer and Chan, page 180, equation 8.1.5.

Part B: The variance of $\hat{r}_2$ is $[1 - (1 - \phi^2) \times \phi^2] / n = [1 - (1 - 0.9^2) \times 0.9^2] / 100 = 0.00846$.

The standard deviation of $\hat{r}_2$ is $0.00846^{0.5} = 0.09198$.

See Cryer and Chan, page 180, equation 8.1.6.

Part C: The standard deviation of $\hat{r}_{40}$ is approximately $1 / \sqrt{n} = 0.10$. This approximation is true for large n.

See Cryer and Chan, page 181.

** Exercise 14.4: Box-Pierce Q statistic

A. What is the Box-Pierce Q statistic for a time series with *n* elements and *k* lags?
B. What is the Ljung-Box Q_* statistic for a time series with *n* elements and *k* lags?
C. Which is larger: the Box-Pierce Q statistic or the Ljung-Box Q_* statistic?
D. What distribution does the Box-Pierce Q statistic have?
E. How many degrees of freedom does the Box-Pierce Q statistic have?

Part A: The Box-Pierce Q statistic is

$$n \times \left(\hat{r}_1^2 + \hat{r}_2^2 + \ldots + \hat{r}_K^2\right)$$

See Cryer and Chan, page 183, equation 8.1.11.

Part B: The Ljung-Box Q_* statistic is

$$Q_* = n \times (n+2) \times \left(\frac{\hat{r}_1^2}{n-1} + \frac{\hat{r}_2^2}{n-2} + \ldots + \frac{\hat{r}_k^2}{n-k}\right)$$

See Cryer and Chan, page 184, equation 8.1.12.

Part C: The Ljung-Box Q_* statistic is larger, since (n+2) > (n-1), (n-2), …, or (n-k).

See Cryer and Chan, page 184, equation 8.1.12: "Since (n+2)/(n-k) > 1 for every K ≥ 1, we have Q* > Q, which partly explains why the original statistic Q tended to overlook inadequate models."

Part D: The Box-Pierce Q statistic has an approximate chi-square distribution.

Jacob: The Box-Pierce Q statistic is a number, not a distribution. What is this question asking?

Rachel: The Box-Pierce Q statistic is the sum of *k* random variables, each of which has a distribution. The sum of these random variables is also a random variable, and it has a χ-squared distribution.

Part E: The degrees of freedom = k – p – q = 15 – 3 – 1 = 11.

See Cryer and Chan, page 183, equation 8.1.11.