Graduate School of Business, University of Chicago
Business 41910, Time Series Analysis, Autumn Quarter 2008, Mr. Ruey S. Tsay

Solutions to Midterm

Problem A: (54 pts) Answer briefly the following questions. Each question has three points.

1. Find the general solution of the difference equation $(1 - 1.3B + 0.4B^2)X_t = 0$, where $X_0 = 2$ and $X_1 = 5$.

   A: The roots of $(1 - 1.3B + 0.4B^2) = 0$ are 0.8 and 0.5. Therefore, the general solution is $X_t = c_1(0.8)^t + c_2(0.5)^t$. Using $X_0 = 2$ and $X_1 = 5$, we obtain $X_t \approx 13.33(0.8)^t - 11.33(0.5)^t$.

2. Obtain the mean and the first two ACFs of the time series X_t satisfying $X_t = 1.0 + 1.3X_{t-1} - 0.4X_{t-2} + a_t$, where a_t is a white noise with variance 1.

   A: (a) $E(X_t) = \frac{1}{1-1.3+0.4} = 10$. For AR(2) models, $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$. Letting $k = 1$, we have $\rho_1 = \frac{1.3}{1.4} = 0.93$. Letting $k = 2$, we have $\rho_2 = 0.81$.

3. Write down the exact likelihood function of the data $\{X_1, X_2, \ldots, X_n\}$ of the model $X_t = \phi_0 + \phi_1 X_{t-1} + a_t$, where the white noise a_t is Gaussian with variance σ_a^2.

   A: First, using $f(x, y) = f(y|x)f(x)$, we have

$$f(X_1, \ldots, X_n) = \prod_{t=2}^{n} f(X_t|X_{t-1}, \ldots, X_1)f(X_1).$$

   Next, X_1 is $N(\frac{\phi_0}{1-\phi_1}, \frac{\sigma_a^2}{1-\phi_1^2})$ and $X_t|X_{t-1}, \ldots, X_1$ is $N(\phi_0 + \phi_1 X_{t-1}, \sigma_a^2)$. Therefore,

$$\begin{aligned} f(X_1, \ldots, X_n) &= \prod_{t=2}^{n} \frac{1}{\sqrt{2\pi}\sigma_a} \exp[-(X_t - \phi_0 - \phi_1 X_{t-1})^2/(2\sigma_a^2)] \\ &\times \frac{\sqrt{1-\phi_1^2}}{\sqrt{2\pi}\sigma_a} \exp\{-(1-\phi_1^2)[X_1 - (\phi_0/(1-\phi_1))]^2/(2\sigma_a^2)\}. \end{aligned}$$

4. Consider two time series

$$\begin{aligned} X_t &= 0.8X_{t-1} + a_t, \\ Y_t &= 1.2Y_{t-1} - 0.32Y_{t-2} + b_t \end{aligned}$$

   where $\{a_t\}$ and $\{b_t\}$ are two independent white noises with unit variance. What is the model ARMA model of $Z_t = X_t + Y_t$?

   A: Rewrite the model of Y_t as $(1 - 0.8B)(1 - 0.4B)Y_t = b_t$. The least comon multiple of the two AR polynomials is $(1 - 0.8B)(1 - 0.4B)$ so that

$$(1 - 0.8B)(1 - 0.4B)Z_t = (1 - 0.4B)a_t + b_t,$$

where the right hand side follows an MA(1) model. Thus, Z_t is an ARMA(2,1) process such that $(1-0.8B)(1-0.4B)Z_t = c_t - \theta c_{t-1}$, where c_t is a white noise series. The values of σ_c^2 and θ can be obtained using $(1+\theta^2)\sigma_c^2 = 2.16$ and $-\theta\sigma_c^2 = -0.4$.

5. Obtain the moment generating function of the model $(1-0.6B)X_t = 1 + (1+0.4B)a_t$, where a_t is a white noise with variance σ_a^2.

   A: MGF $= \Gamma(z) = \frac{\sigma_a^2(1+0.4z)(1+0.4z^{-1})}{(1-0.6z)(1-0.6z^{-1})}$.

6. Obtain the non-zero autocorrelation function of $X_t = (1-0.4B)(1-0.6B^4)a_t$, where a_t is a white noise series.

   A: $\rho_0 = 1$, $\rho_1 = -0.4/1.16 = -0.34$, $\rho_4 = -0.6/1.36 = -0.44$, $\rho_3 = \rho_5 = \rho_1\rho_4 = 0.15$, and $\rho_j = 0$ all other positive j.

7. Suppose that X_t follows the ARMA(1,1) model $(1-\phi B)X_t = (1-\theta B)a_t$, where a_t is a white noise. Let $Z_t = (1-\gamma B)X_t$. What is the model for Z_t?

   A: From the two equations, we have $X_t = \frac{1-\theta B}{1-\phi B}a_t$ and $\frac{1}{1-\gamma B}Z_t = X_t$. Therefore, $(1-\phi B)Z_t = (1-\gamma B)(1-\theta B)a_t$, which is an ARMA(1,2) model.

8. Consider the AR(1) time series $X_t = 0.8X_{t-1} + a_t$, where a_t is a Gaussian white noise with variance 1. Let $Z_t = X_{3t} + X_{3t-1} + X_{3t-2}$ be the 3-aggregate of X_t. What is the order of the ARMA model for Z_t?

   A: An ARMA(1,1) model. Consider the model $(1-0.8^2B)Z_t = (Z_{3t} + Z_{3t-1} + Z_{3t-2}) - 0.8^3(Z_{3(t-1)} + Z_{3(t-1)-1} + Z_{3(t-1)-2}) = (Z_{3t} - 0.8^3Z_{3t-3}) + (Z_{3t-1} - 0.8^3Z_{3t-4}) + (Z_{3t-2} - 0.8^3Z_{3t-5})$. Now, for each term, one can use "addition" and "subtration" to work out the MA part. For instance, $Z_{3t} - 0.8^3Z_{3t-3} = Z_{3t} - 0.8Z_{3t-1} + 0.8Z_{3t-1} - 0.8^2Z_{3t-2} + 0.8^2Z_{3t-2} - 0.8^3Z_{3t-3}$.

9. Suppose that X_t follows the model $(1-0.87B+0.27B^2)X_t = a_t$, where a_t is a white noise with variance σ_a^2, and $X_{100} = 3.0$ and $X_{99} = -1.0$. What are the 2-step ahead forecast of X_{102} at the forecast origin $t = 100$? What is the variance of the associated forecast error?

   A: $X_{100}(2) = 1.70$ and the variance of forecast error is $(1+\psi_1^2)\sigma_a^2 = (1+.87^2)\sigma_a^2 = 1.76\sigma_a^2$.

10. Consider a stationary ARMA(p,q) model $\phi(B)X_t = \theta(B)a_t$, where a_t is a Gaussian white noise with variance 1. Write down the generalized Yule-Walker equations of X_t.

    A: $\rho_k = \phi_1\rho_{k-1} + \ldots + \phi_p\rho_{k-p}$ for $k = q+1, \ldots, q+p$.

11. Define an additive outlier and a level shift of a time series X_t at time index $t = d$.

    A: Additive outlier: $Y_t = X_t + \omega I_t^{(d)}$ and level shift is $Y_t = X_t + \frac{\omega}{1-B}I_t^{(d)}$, where $I_t^{(d)}$ is the indicator variable for time index $t = d$.

12. Suppose an innovational outlier of magnitude 3 occurs at $t = 100$ to the AR(1) time series $X_t = 0.6X_{t-1} + a_t$. Let Y_t be the observed time series. What is the relationship between Y_{100} and X_{100}? How about that between Y_{102} and X_{102}? Finally, what is the relationship between X_{150} and Y_{150}?

    A: $Y_{100} = X_{100} + 3$, $Y_{102} = X_{102} + .36 \times 3$, and $Y_{150} = X_{150} + 0.6^{50} \times 3 \approx X_{150}$.

13. Give two sources of uncertainty in time series forecasting.

    A: Parameter uncertainty and model (order) uncertainty.

14. Consider the Gaussian ARMA(p, q) time series X_t satisfying $\phi(B)X_t = \theta(B)a_t$, where a_t is a white noise with variance σ_a^2. State the stationarity and invertibility conditions for X_t.

    A: Stationarity: All zeros of $\phi(B)$ are outside the unit circle. Invertibility: All zeros of $\theta(B)$ are outside the unit circle.

15. Consider the time series $X_t = 0.6X_{t-1} + a_t$ where a_t is a Gaussian white noise with variance 1. Let $Z_t = X_{2t}$. What is the model for Z_t?

    A: $Z_t - 0.36Z_{t-1} = X_{2t} - 0.36Z_{2t-2} = X_{2t} - 0.6Z_{2t-1} + 0.6X_{2t-1} - 0.36X_{2t-2} = a_t + 0.6a_{t-1}$. The right hand side does not involve any time lag on the time scale of Z_t. More specifically, (a_t, a_{t-1}) and (a_{t-2}, a_{t-3}) are uncorrelated. Thus, the model for Z_t is an AR(1) model.

16. Describe briefly four statistics (or tools) for model checking in time series analysis.

    A: (1) Residual ACF, (b) Ljung-Box statistics of residual ACF, (c) outlier test statistics, and (d) residual plot. [You may also include normality test, residuals PACF, or residual EACF.]

17. Describe briefly two methods that can be used to make inference in a regression analysis when the errors might have serial correlations and conditional heteroscedasticity.

    A: (a) Use HAC-type covariance matrix of the fitted parameters and (b) use regression model with time-series errors.

18. Suppose that X_t follows the model $X_t = .6X_{t-1} + a_t$ and $Y_t = b_t$, where a_t and b_t are two independent white noise series with variances σ_a^2 and σ_b^2, respectively. What is the model for $Z_t = X_t + Y_t$, including parameter values.

    A: ARMA(1,1) model, $(1 - 0.6B)Z_t = c_t - \theta c_{t-1}$, where c_t is a white noise with variance σ_c^2. The values of θ and σ_c can be determined by using $(1 + \theta^2)\sigma_c^2 = \sigma_a^2 + 1.36\sigma_b^2$ and $-\theta\sigma_c^2 = -0.6\sigma_b^2$.

Problem B. (20 pts) In this tough economic environment, it is timely to analyze the U.S. monthly total civilian unemployment rate. The time span is from January 1948 to September 2008 for 729 observations. The data are *seasonally adjusted* and are obtained from the Federal Reserve Bank at St. Louis database. I use R to perform the analysis. First, I fit a non-seasonal model, but found that it is inadequate. Second, I employed a seasonal model and found that it passed the usual model checking measures. Based on the output attached, answer the following questions:

1. (4 points) Write down the fitted non-seasonal model, including residual variance, and explain why the model is inadequate.

   A: Let X_t be the monthly U.S. unemployment rate. Then, the non-seasonal model is $(1-0.52B-0.25B^2)(1-B)X_t = (1-0.51B)a_t$, with $\sigma_a^2 = 0.040$. This model is inadequate because the Ljung-Box statistics of the residuals show significant serial correlations inthe residuals, i.e. p-value of $Q(12)$ is 0.0004.

2. (5 points) Write down the fitted seasonal model for the series, including the residual variance.

   A: $(1-0.23B^2)(1-0.61B^{12})(1-B)X_t = (1+0.15B^3+0.17B^5)(1-0.86B^{12})a_t$ with $\sigma_a^2 = 0.037$.

3. (5 points) Give two reasons that the seasonal model is preferred over the non-seasonal model, even though the data were seasonally adjusted.

   A: (a) The Ljung-Box statistics show that the residuals of the seasonal model have no significant serial correlations for the first 24 lags. (b) The estimated parameters of the seasonal part are highly significant.

4. (4 points) Based on the seasonal model, compute 95% interval forecasts of U.S. unemployment rates for October and November 2008, respectively. [You may assume the conditional normaility of the unemployment rate.]

   A: October: $6.29 \pm 1.86 \times 0.19$; November: $6.33 \pm 1.96 \times 0.27$.

5. (2 points) Based on the forecasts of the seasonal model, do you think the U.S. economy will improve soon? Why?

   A: The forecasts of unemployment rate continue to increase for at least the next 5 quarters. Thus, it indicates that the U.S. economy will not improve soon.

Problem C. (26 pts) Another economic time series of interest is the U.S. monthly total non-farm payrolls (in thousands). I took the natural log transformation. Again, the data are also obtained from the Federal Reserve Bank at St. Louis and the time span is from January 1939 to September 2008 for 837 observations. The series is NOT seasonally adjusted. Again, I use R to perform the analysis and keep output of several models I tried. Use the attached output to answer the following questions.

1. (4 points) Consider the model I called "p1". Give two reasons that can be used to justify my simplification of the model.

   A: (a) The estimated AR-2 and AR-3 coefficients are not significantly different from zero, (b) the AIC of the simplified model is lower than that of the model before simplication.

2. (5 points) Write down the fitted final "p1" model, including the residual variance.

   A: $(1-0.79B)(1-B)(1-B^{12})Z_t = (1-0.49B)(1-0.72B^{12})a_t$ with $\sigma_a^2 = 1.9\times 10^{-5}$.

3. (5 points) Consider the model I called "p2". Write down the fitted model, including the residual variance.

   A: $(1-0.97B)(1-B)(1-B^{12})X_t = (1-0.72B-0.19B^6-0.07B^7)(1-0.68B^{12})a_t$ with $\sigma_a^2 = 1.85\times 10^{-5}$.

4. (5 points) Give two justifications that "p2" is preferred over "p1".

   A: (a) The model has a lower AIC value. (b) The Ljung-Box statistics indicate the residuals have no significant zerial correlations. $Q(24) = 30.58$ with p-value 0.17.

5. (3 points) Is the final "p2" model adequate? Why?

   A: Yes, because (a) the residuals have no significant serial correlations (see the Q-statistics), (b) all parameter estimates are significant at the 5% level.

6. (4 points) Why do the standard errors of the forecasts increase as forecast horizon

   A: Because the fitted model contains the unit root. increases?