Booth School of Business, University of Chicago
Business 41202, Spring Quarter 2017, Mr. Ruey S. Tsay

Solutions to Midterm

Problem A: (36 pts) Answer briefly the following questions. Each question has two points.

1. **(For Questions 1 to 4)** Let r_t denote the daily log return of an asset. Describe a procedure for testing the existence of serial correlations in r_t. What is the reference distribution of the test statistic used?

   A: Consider the null hypothesis $H_0 : \rho_1 = \cdots = \rho_m = 0$ vs $H_a : \rho_i \neq 0$ for some $1 \leq i \leq m$. The test statistic is $Q(m)$ and the reference distribution is χ^2_m.

2. Let $\mu_t = E(r_t|F_{t-1})$, where F_{t-1} denotes the information available at time $t-1$. Write the return as $r_t = \mu_t + a_t$. Describe the null hypothesis for testing the ARCH effect of r_t, including definition of the statistics involved in H_0.

   A: $H_0 : \rho_1 = \cdots = \rho_m = 0$, where ρ_i is the lag-i ACF of a_t^2.

3. Let $a_t = \sigma_t \epsilon_t$, where $\sigma_t^2 = E(a_t^2|F_{t-1})$ and ϵ_t are iid random variates with mean zero and variance 1. Describe a statistic discussed in class for testing the null hypothesis that ϵ_t is normally distribution. What is the reference distribution of the test statistic?

   A: The JB statistic, and the reference distribution is χ^2_2.

4. Suppose that σ_t^2 of Question 3 satisfies the model

   $$\sigma_t^2 = 0.01 + 0.1a_{t-1}^2 + 0.8\sigma_{t-1}^2.$$

   Compute $E(a_t)$ and $\text{Var}(a_t)$.

   A: $E(a_t) = 0$, $\text{Var}(a_t) = \frac{0.01}{1-0.1-0.8} = 0.1$.

5. Provide two reasons that may lead to serial correlations in the observed asset returns even when the underlying *true* returns are serially uncorrelated.

   A: Any two of the followings: (a) nonsynchronous trading, (b) bid-ask bounce, (c) data smoothing.

6. Provide two methods that can be used to specify the order of an autoregressive time series.

   A: PACF or information criteria (such as AIC, BIC). The latter can be done with the command `ar` in R.

7. Describe two statistics that can be used to measure dependence between variables.

   A: Any two of the following: (a) Pearson correlation coefficient, (b) Kendall's tau, and (c) Spearman's rho.

8. Provide two volatility models that can be used to model the `leverage` effect of asset returns.

   A: Any two of the following models: (a) EGARCH, (b) TGARCH (also known as GJR-GARCH), and (c) APARCH.

9. Describe a nice feature and a drawback of using GARCH models to modeling asset volatility.

   A: Nice feature can be (a) volatility clustering or (b) heavy tails. Drawback can be (a) no leverage effect or (b) restrictive.

10. Give two potential impacts on the linear regression analysis if the serial dependence in the residuals is overlooked.

    A: Any two of the followings: (a) supurious regression (or bias in coefficient estimates), (b) errornous standard errors, or (c) misleading t-ratios.

11. **(Questions 11-12)** Suppose that the log return r_t follows the model

$$\begin{aligned} r_t &= 0.0061 + a_t, \\ a_t &= \sigma_t \epsilon_t, \quad \epsilon_t \sim t_7^*(0.9) \\ \sigma_t^2 &= 0.08 a_{t-1}^2 + 0.92 \sigma_{t-1}^2 \end{aligned}$$

    where $t_v^*(\xi)$ denotes a skew standardized Student-t distribution with v degrees of freedom and skew parameter ξ. Suppose further that $a_{1000} = -0.02$ and $\sigma_{1000} = 0.15$. Consider the forecast origin $t = 1000$. What is the 5-step ahead prediction of the log return? What is the 5-step ahead volatility prediction? That is, $r_{1000}(5)$ and $\sigma_{1000}(5)$?

    A: $r_{1000}(5) = 0.0061$, $\sigma_{1000}(5) = \sqrt{0.08 * (-0.02)^2 + 0.92 * (0.15)^2} = 0.144$.

12. Compute a 95% interval prediction for the cumulative returns $r_{1001} + \cdots + r_{1005}$ at the forecast origin $t = 1000$.

    A: mean = 0.0061*5 = 0.0305, sig = $\sqrt{5} \times 0.144 = 0.322$. The quantiles are 1.883918 and -2.104275, respectively. The 95% forecast interval is $(-0.647, 0.637)$.

13. **(Questions 13-14)** Consider the time series model $(1-0.7B+0.8B^2)r_t = 0.3+(1-0.5B)a_t$, where $a_t \sim_{iid} N(0,1)$. Is the model stationary? Why?

    A: Yes, the solutions of $1 - 0.7x + 0.8x^2 = 0$ are $0.438 \pm 1.029i$. The absolute value is 1.118 which is greater than 1.

14. Does the model imply existence of business cycles? If yes, what is the average length of the business cycle?

    A: Yes. The average length is approximately 5.38.

15. Write down the Airline model for a monthly time series. Why is the model useful in modeling monthly time series?

    A: The model is $(1-B)(1-B^{12})r_t = (1-\theta_1 B)(1-\theta_{12}B^{12})a_t$, where a_t is an iid sequence. The model is useful because it is a double exponential smoothing model; one for regular dependence and the other for seasonal dependence.

16. **(Questions 16-18)** Consider the monthly log returns of IBM stock from January 1961 to December 2016. The sample size is 672. The sample excess kurtosis is 1.83. Do the monthly IBM log returns have heavy tails? Perform a test to justify your answer.

    A: The test statistic is $1.83/sqrt24/672 \approx 9.68$, which is greater than 1.96.

17. Using a GARCH(1,1) model with standardized Student-t distribution as the baseline model, we like to explore whether the volatilities of the log returns are lower in the summer (June, July, August). To this end, a summer dummy variable is created for the volatility equation. Write down the fitted model. (See R output).

    A: The model is

$$\begin{aligned} r_t &= 0.00839 + a_t \\ a_t &= \sigma_t \epsilon_t, \quad \epsilon_t \sim_{iid} t^*_{9.07} \\ \sigma_t^2 &= 3.02 \times 10^{-4} + 1.33 \times 10^{-4} I_t^{(summer)} + 0.10 a_{t-1}^2 + 0.827\sigma_{t-1}^2. \end{aligned}$$

18. Based on the fitted model, were the volatilities lower in the summer? Why?

    A: No, the coefficient of the summer dummy variable has a t-ratio of 0.378 with p-value 0.71, which is not statistically significant.

Problem B. (35 points) Consider the daily log returns of Google stock for a period with 2705 observations. Analysis of the return via R is attached. Use the output to answer the following questions.

1. (2 points) Are there serial correlations in the dialy log returns? Why?

    A: No, the Ljung-Box statistics of the return series show Q(10) =9.33 with p-value 0.50.

2. (2 points) Is the expected log return different from zero? Why?

   A: No, because the 95% confidence interval of the mean is $(-0.177, 1.21)\times 10^{-3}$, which contains zero.

3. (3 points) Is the distribution of the log return skew? Perform a test to justify your answer.

   A: The t-ratio for skewness is $0.613/\sqrt{6/2705} = 13.0$, which is greater than 1.96.

4. (2 points) A Gaussian GARCH(1,1) model, called `m1`, is fitted. Write down the fitted model.

   A: The model is

$$\begin{aligned} r_t &= 6.89 \times 10^{-4} + a_t \\ a_t &= \sigma_t \epsilon_t, \quad \epsilon_t \sim_{iid} N(0,1) \\ \sigma_t^2 &= 1.09 \times 10^{-5} + 0.0812 a_{t-1}^2 + 0.89 \sigma_{t-1}^2. \end{aligned}$$

5. (3 points) A Student-t GARCH(1,1) model, called `m2`, is also fitted. Write down the fitted model.

   A: The model is

$$\begin{aligned} r_t &= 7.42 \times 10^{-4} + a_t \\ a_t &= \sigma_t \epsilon_t, \quad \epsilon_t \sim_{iid} t_{3.86}^* \\ \sigma_t^2 &= 1.98 \times 10^{-6} + 0.0318 a_{t-1}^2 + 0.962 \sigma_{t-1}^2. \end{aligned}$$

6. (2 points) Let v denote the degrees of freedom of a standardized Student-t distribution. Based on the fitted model `m2`, test $H_0 : v = 5$ versus $H_a : v \neq 5$ and draw your conclusion.

   A: The t-ratio is $\frac{3.86-5}{0.291} = -3.92$, which is less than -1.96. Thus, the degrees of freedom is not 5.

7. (3 points) A skew Student-t GARCH(1,1) model, called `m3`, is fitted. Based on the fit, is the distribution of the standardized residuals symmetric? Perform a test to justify your answer.

   A: The t-ratio is $(1.024-1)/0.0263 = 0.913$, which is smaller than 1.96. Thus, the symmetry of the standardized residuals cannot be rejected at the 5% level.

8. (4 points) A TGARCH(1,1) model with Student-t distributions is also considered, called `m4`. Write down the fitted model.

   A: The model as

$$\begin{aligned} r_t &= 6.77 \times 10^{-4} + a_t \\ a_t &= \sigma_t \epsilon_t, \quad \epsilon_t \sim_{iid} t_{3.89}^* \\ \sigma_t^2 &= 2.48 \times 10^{-6} + 0.0288(|a_{t-1}| - 0.441 a_{t-1})^2 + 0.958 \sigma_{t-1}^2. \end{aligned}$$

However, since I stated TGARCH model, you can write the volatility equation as

$$\sigma_t^2 = 2.48 \times 10^{-6} + (0.0288 + 0.441 N_{t-1}) a_{t-1}^2 + 0.958 \sigma_{t-1}^2,$$

where $N_{t-1} = 1$ if and only if $a_{t-1} < 0$.

9. (2 points) Based on the fitted TGARCH model, is the leverage effect statistically significant? Why?

   A: Yes, the t-ratio of the leverage effect is 4.31 with p-value 1.63×10^{-5}.

10. (3 points) Based on the prediction of the model `m4`, compute a 5-step ahead 95% interval forecast for the log return at the forecast origin t = 2705.

    A: The 97.5% quantile is $t^*_{3.89}$ is 1.957. Therefore, the 95% interval forecast is $0.000677 \pm 1.957 \times 0.0104$. That is, $(-0.0197, 0.021)$.

11. (2 points) Among the fitted models, `m1, m2, m3, m4`, which model is preferred? Why?

    A: `m4`, because it has the smallest AIC.

12. (2 points) For numerical stability, returns are multiplied by 100, i.e. in percentages. A GARCH-M model is fitted, called `m5`. Is the risk premium statistically signigicant? Why?

    A: No, the estimate of risk premium is 0.00798 with t-ratio 0.438, which is not statistically significant.

13. (2 points) Let $a_t = (r_t - \text{mean}(r_t)) * 100$ be the residuals of the mean equation for the returns. An IGARCH(1,1) model is fitted to a_t, called `m6`. Write down the fitted model.

    A: The model is

$$\begin{aligned} r_T &= a_t \\ a_t &= \sigma_t \epsilon_t, \quad \epsilon_t \sim_{iid} N(0,1) \\ \sigma_t^2 &= 0.0798 + (1 - 0.878) a_{t-1}^2 + 0.878 \sigma_{t-1}^2. \end{aligned}$$

14. (3 points) Based on the fitted IGARCH(1,1) model, compute the 1-step to 3-step ahead volatility forecasts of the log return at forecast origin t = 2705.

    A: The volatility forecasts are 0.97, 1.01, and 1.05, respectively. The one-step ahead prediction is computed from $\sqrt{0.0798 + (1 - 0.878)(-0.144)^2 + 0.878(0.989)^2} = 0.97$.

Problem C. (20 points) Consider the monthly Moody's seasoned Aaa and Baa corporate bond yields from July 1954 to March 2005 with 609 observations.

1. (2 points) A simple linear regression is used to find the relationship between Aaa and Baa bond yields. Write down the fitted model, including R^2.

   A: $\text{Baa}_t = 0.1 + 1.115\text{Aaa}_t + e_t$ with standard error of residuals being 0.283 and $R^2 = 99.12\%$.

2. (2 points) Is the fitted linear regression model adequate? Why?.

   A: No, the Ljung-Box statistics of the residuals show Q(12) = 2608.2 which is highly significant. Thus, there are serial correlations in the residuals.

3. (3 points) Denote the first-differenced bond yields by dA and dB, respectively. A regression model with time series errors is employed for dB with dA as the regressor. Write down the fitted model, called `n3`.

   A: The model is

$$\begin{aligned}\text{dB}_t &= 0.794\text{dA}_t + a_t + 0.294a_{t-1} + 0.028a_{t-2} \\ &- 0.001a_{t-3} + 0.006a_{t-4} + 0.087a_{t-5}, \quad \sigma_a^2 = 0.00751.\end{aligned}$$

4. (3 points) The model `n3` contains several insignificant parameter estimates. A modified model, called `n4`, is used. Write down the modified model.

   A: The model is

$$\text{dB}_t = 0.798\text{dA}_t + (1 + 0.289B)(1 + 0.0932B^5)a_t,$$

   where $\sigma_a^2 = 0.00751$.

5. (2 points) Compare models `n3` and `n4`. Provide a justification that the insignificant parameters of `n3` can indeed be removed.

   A: The AIC of model `n4` is smaller than that of `n3`.

6. (2 points) Model checking indicates several possible outliers. To study the impact of outliers, we tried to handle the largest outlier (in absolute value). The resulting model is called `n5`. Compare models `n4` and `n5`. Describe a clear impact of the outlier on the fit.

   A: The σ_a^2 reduces from 0.00751 to 0.00709. Alternativley, the AIC drops from -1240 to -1273.

7. (2 points) Backtest is used to compare models `n4` and `n5`. Based on the results shown, does the outlier have any impact on prediction? Why?

   A: No, the outlier has no impact on out-of-sample prediction. The two models provide exactly the same out-of-sample predictions.

8. (2 points) A pure time series model is also identified for the dB series (`n6`). Write down the fitted model.

   A: The model is

$$(1 - 0.194B)\text{dB}_t = (1 + 0.323B)a_t, \quad \sigma_a^2 = 0.0321.$$

9. (2 points) Backtest is also applied to the pure time series model. Based on the result, is dA helpful in predicting dB?

   A: Yes, the RMSE of model `n5` is 0.0912 whereas that of model `n6` is 0.209, which is much larger.

Problem D. (9 points) Consider the daily log returns of Coke (KO) from January 4, 2004 to April 27, 2017 for 3351 observations.

1. (3 points) An EGARCH model in the form

$$\ln(\sigma_t^2) = \omega + \alpha_1 \epsilon_{t-1} + \gamma_1(|\epsilon_{t-1}| - 0.798) + \ln(\sigma_{t-1}^2),$$

   is entertained, where $a_{t-1} = \sigma_{t-1}\epsilon_{t-1}$ and $\epsilon_t \sim_{iid} N(0, 1)$. Write down the fitted model.

   A: The model is

$$\begin{aligned} r_t &= 2.13 \times 10^{-4} + a_t \\ a_t &= \sigma_t \epsilon_t, \quad \epsilon_t \sim_{iid} N(0, 1) \\ \ln(\sigma_t^2) &= -0.314 - 0.0847\epsilon_{t-1} + 0.179(|\epsilon_{t-1}| - 0.798) + 0.965 \ln(\sigma_{t-1}^2). \end{aligned}$$

2. (2 points) Based on the fitted model, is the leverage effect significant? Why?

   A: Yes, the t-ratio of α_1 is -7.25, which is highly significant with p-value close to zero.

3. (4 points) Based on the fitted model, compute the ratio

$$\frac{\sigma_t^2(\epsilon_{t-1} = -3)}{\sigma_t^2(\epsilon_{t-1} = 3)} = ?$$

   What is the implication of the ratio?

   A: The ratio is

$$\frac{\sigma_t^2(\epsilon_{t-1} = -3)}{\sigma_t^2(\epsilon_{t-1} = 3)} = \frac{e^{-0.2637 \times (-3)}}{e^{0.0943 \times 3}} \approx 1.66.$$

   Thus, dropping 3 standard error has a 66% larger impact on the conditional variance than increasing by 3 standard error.

   Note that $-0.2637 = -0.0847 - 0.178$ and $0.0943 = -0.0847 + 0.179$.

   Also, a short cut to the solution is $e^{0.0847 \times 2 \times 3} = 1.66$.