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

Solutions to Final Exam

Problem A: (40 points) Answer briefly the following questions.

1. Describe two characteristics commonly encountered in modeling high-frequency log returns in finance.

   A: Any two of (a) impact of market microstructure, (b) discreteness in price change, and (c) heavy tails.

2. Give one nice feature and one drawback of using RiskMetrics in financial risk management.

   A: Nice feature: simplicity or transparency. Drawback: strong assumptions or model is often rejected by asset returns.

3. Assume that x_t follows the stochastic diffusion equation $dx_t = \mu x_t dt + \sigma x_t dw_t$, where w_t is a Wiener process. Let $G(x_t) = \sqrt{x_t}$. Derive the stochastic diffusion equation for $G(x_t)$.
$$dG = \left(\frac{\mu}{2} - \frac{\sigma^2}{8}\right)\sqrt{x_t}dt + \frac{\sigma}{2}\sqrt{x_t}dw_t.$$

4. Describe two approaches to calculate value at risk using the extreme value theory.

   A: (a) block maximum and (b) Peaks over the threshold or generalized Pareto distribution.

5. Give two volatility models that can estimate the leverage effect in modeling daily asset returns.

   A: Any two of (a) EGARCH, (b) TGARCH, (c) GJR-GARCH, and (d) asymmetric power ARCH.

6. Describe two statistics that can be used to specify the order of an autogressive time series.

   A: PACF and information criteria such as AIC, BIC, and HQ.

7. Describe two statistics that can be used to test for ARCH effects in the log returns of an asset.

   A: Ljung-Box Q-statistics of squared returns and Lagrange multiplier test (or F-test).

8. Give two potential problems of over-looking the serial correlations in a linear regression model.

   A: Any two of (a) supurious regression, (b) in-accurate estimates of parameter variance, and (c) erroneous inference.

9. Describe two alternative approaches (vs econometric modeling) to calculate asset volatility.

   A: Any two of (a) high-frequency log returns, (b) use daily open, high, low and close prices, and (c) moving-window estimation.

10. Asset returns typically do not follow the Gaussian distribution. Provide two alternative statistical distributions that can improve volatility modeling.

    A: Generalized error distribution and Student-t distribution.

11. Many statistical models have been introduced in this course. Describe two models that can be used to model the price changes in high-frequency financal trading.

    A: ADS decomposition, i.e. logistic regression and (b) ordered probit model

12. Give one similarity and one difference between threshold autoregressive models and Markov switching models.

    A: Similarity: piecewise linear models. Difference: stochastic vs determinstic way to determine the regimes.

13. In empirical data analysis, one often ecounters certain outliers. Describe two methods to mitigate the effects of outliers in financial time series analysis.

    A: Use dummy variables (indicator variables) or employ a heavy-tail distribution.

14. Describe two methods that can be used to compare different statistical models in analysis of financial data.

    A: Information criteria and out-of-sample prediction (or backtesting).

15. Give two reasons for analyzing two time series jointly.

    A: To find the dynamic relationship between variables and to improve accuracy in prediction.

16. Give two multivariate volatility models that always give positive-definite volatility matrices.

    A: Any two of (a) exponentially weighted moving average, (b) BEKK model, (c) Dynamic conditional correlation (DCC) model, and (d) models based on Cholesky decomposition.

17. (**For Questions 17-20**). Consider the daily simple returns of Qualcomm (QCOM) stock and the S&P composite index for a period of 2769 trading days. Some analysis of the data are given in the attached R output. Let μ be the daily expected return of the S&P composite index. Test $H_0 : \mu = 0$ versus $H_a : \mu \neq 0$. What is your conclusion? Why?

    A: The null hypothesis cannot be rejected because the 95% confidence interval of the mean contains zero.

18. Is the CAPM model, **n1**, adequate for the QCOM return? Why?

    A: No, because there exists some serial correlations in the residuals with Q(10) = 19.0 with p-value 0.04.

19. A regression model with time series errors is entertained. After removing insignificant parameters. A final model is obtained. Write down the fitted model, **n3**.

    A: The model is $\text{qcom}_t = 0.99\text{sp}_t + a_t - 0.0472a_{t-2}$, with $\sigma_a^2 = 0.000227$.

20. Does the final model, **n3**, successfully capture the serial correlations in the residuals? Why?

    A: Yes, the Ljung-Box statistics of the residuals show Q(10) = 14.25 with p-value 0.16. (The p-value is still greater than 0.05 if degrees of freedom are adjusted.)

Problem B. (12 points) Exchange rate series.

1. Write down the sample cross-correlation matrix of $\boldsymbol{x}_t$. Why is the correlation between the log returns of exchange rates negative?

$$\widehat{\boldsymbol{\rho}}_0 = \left[\begin{array}{rr} 1 & -0.245 \\ -0.245 & 1 \end{array} \right].$$

   The correlation is negative because the Dollar and Yen exchange rate is measured in Yen.

2. The lag-1 sample cross-correlation matrix is

$$\boldsymbol{\rho}_1 = \left[\begin{array}{rr} 0.0060 & 0.0109 \\ 0.0049 & -0.0169 \end{array} \right].$$

   What is the meaning of the correlation -0.0169? What is the meaning of the correlation 0.0109?

   A: -0.0169 measures the linear dependence of r_{2t} on $r_{2,t-1}$, where r_{2t} denotes the log return of dollar-yen exchange rate. The estimate 0.0109 measures the linear dependence of r_{1t} on $r_{2,t-1}$, where r_{1t} is the log return of Euro-dollar exchange rate.

3. Consider the null hypothesis $H_0 : \boldsymbol{\rho}_1 = \cdots = \boldsymbol{\rho}_{10} = \mathbf{0}$ versus $H_a : \boldsymbol{\rho}_i \neq \mathbf{0}$ for some $1 \leq i \leq 10$. Perform a proper test and state the p-value of the test statistic. Also, draw your conclusion.

   A: Q(10) = 42.0 with p-value 0.38 so that the null hypothesis cannot be rejected at the 5% level.

4. State the implication of the conclusion of the hypothesis testing of Question (3).

   A: There is no significant serial or cross- correlations between the two series.

5. (3 points) A vector autoregressive model of order 1 is entertained for $\boldsymbol{z}_t$. Write down the fitted model, including the residual covariance matrix.

   A: Let $\boldsymbol{r}_t = (r_{1t}, r_{2t})'$ be the two log return series. The model is

$$\boldsymbol{r}_t = \begin{bmatrix} 1.0 & 8.85 \times 10^{-6} \\ 0.0514 & 0.999 \end{bmatrix} \boldsymbol{r}_{t-1} + \boldsymbol{a}_t, \quad \boldsymbol{\Sigma}_a = \begin{bmatrix} 6.75 \times 10^{-5} & -0.00142 \\ -0.00142 & 0.473 \end{bmatrix}$$

6. (1 point) Is the VAR(1) model adequate? Why?

   A: Yes, the residuals show no significant serial and cross-correlations based on the Q(m) statistics.

Problem C. (38 points) Consider the daily log returns of Qualcomm (QCOM) and Pfizer (PFE) stocks.

1. (**Questions 1 to 4**) What are the VaR of the individual stocks using RiskMetrics?

   A: VaR = \$35,485 for QCOM and \$28,807 for PFE.

2. What is the VaR of the portfolio using RiskMetrics?

   A: \$ $\sqrt{35485^2 + 28807^2 + 2 * 0.42 * 35485 * 28807} =$ \$54,293.

3. What is the VaR of the portfolio for the next 10 trading days using RiskMetrics?

   A: \$ $\sqrt{10} \times 54293 =$ \$171,690.

4. Suppose the position on Pfizer stock is changed to a short position. What is the VaR of the portfolio for the next trading day?

   A: \$ $\sqrt{35485^2 + 28807^2 - 2 * 0.42 * 35485 * 28807} =$ \$35,077.

5. A Gaussian GARCH(1,1) model is ftted to the QCOM returns, **m3**. What are the VaR and expected shortfall of the stock?

   A: VaR = \$37,160 and ES = \$42,643 (use RMeasure.R)

6. A Gaussian GARCH(1,1) model is also fitted to the PFE returns, **m4**. What are the VaR and expected shortfall of the stock?

   A: VaR = \$23,484 and ES = \$26,962.

7. What is the VaR of the portfolio for the next trading day using the Gaussian GARCH(1,1) models?

   A: $\sqrt{37160^2 + 23484^2 + 2 * 0.42 * 37160 * 23484} =$ \$51,628.

8. A GARCH(1,1) model with standardized Student-t innovations is entertained for the QCOM stock returns, **m5**. What are the VaR and expected shortfall of the QCOM position?

   A: VaR = \$40,836 and ES = \$56,011.

9. What are the VaR and expected shortfall of QCOM position for the next five (5) trading days?

   A: VaR(5) = \$90,781 and ES(5) = \$125,094.

10. A GARCH(1,1) model with standardized Student-t innovations is also fitted to the PFE stock returns, **m6**. What are the VaR and expected shortfall of PFE position?

    A: VaR = \$25,192 and ES = \$32,476.

11. What is the VaR of the portfolio for the next trading day using Student-t GARCH(1,1) models?

    A: $\sqrt{40836^2 + 25192^2 + 2 * 0.42 * 40836 * 25192} = \$56{,}270$.

12. Turn to extreme value theory. A extreme value distribution is fitted to the QCOM stock returns with block size 21, **m8**. Write down the three parameter estimates.

    A: The estimates of $(\xi, \sigma, \mu) = (0.298, 0.0130, 0.0238)$.

13. Let ξ be the shape parameter. Consider the null hypothesis $H_0 : \xi = 0$ versus $H_a : \xi \neq 0$. Perform a proper test and draw your conclusion.

    A: t-ratio $= \frac{0.298}{0.0796} = 3.74$, which is greater than 1.96. The null hypothesis is rejected.

14. Based on the fitted EVT, calculate the VaR for the QCOM position for the next trading day. What is the VaR for the next ten (10) trading days?

    A: VaR = \$49,503. For 10 days, the VaR = \$ $10^{0.298} \times 49503 = \$98{,}367$.

15. A peaks over the threshold approach is used. With threshold 0.02, we fit a POT model to the QCOM stock returns, **m9**. Write down the estimate of the shape parameter ξ. What are the VaR and expected shortfall for the QCOM position under the fitted POT mode (next trading day).

    A: $\hat{\xi} = 0.251$. VaR = \$55,776 and ES = \$83,000.

16. The threshold is changed to 0.03, **m10**. What are the VaR and expected shortfall for the QCOM position for the next trading day?

    A: VaR = \$55,663 and ES = \$84,681.

17. A generalized Pareto distribution with threshold 0.02 is also fitted to the QCOM returns, **m11**. What are the VaR and expected shortfall for the QCOM position? Compared with results of model **m9**. Is there any significant difference? Why?

    A: VaR = \$55,769 and ES = \$82,995. These quantities are close to those of model **m9**. Therefore, there is no significant difference. The two methods are based on the same theory.

18. Turn to quantitle regression. The daily log returns of the S&P composite index during the sample period is used to obtain a proxy of market volatility; see the model **W1** in the R output. We then employ a quantile regression for QCOM returns using (1) squares of lag-1 return, (2) lag-1 market volatility and (3) lag-1 QCOM volatility as explanatory variables. However, the lag-1 squared return appears to be statistically insignificant; see model **mm**. A simplified quantitle regression is fitted, **mm1**. Based on the fitted model, is lag-1 S&P volatility helpful in estimating the 95 percentiles of QCOM returns? Why?

    A: Yes, the coefficient estimate 1.07 is highly significant with p-value close to zero.

19. Given that lag-1 volatilities of the QCOM return and market index are 0.01626 and 0.00609, respectively, what is the fitted value of the 95th percentile?

    A: 0.00602+0.536*0.0163+1.074*0.00609 = 0.0213.

Problem D. (10 points) Consider, again, the daily log returns of Qualcomm stock of Problem C. Let r_t and m_t be the daily log returns of QCOM and S&P composite index. Define

$$Q_t = \begin{cases} 1 & \text{if } r_t > 0 \\ 0 & \text{otherwise,} \end{cases} \qquad M_t = \begin{cases} 1 & \text{if } m_t > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The goal is to predict Q_t using Q_{t-i} and M_{t-j}, where $i, j > 0$.

1. A logistic regression is employed using $\{Q_{t-i}, M_{t-i} | i = 1, 2, 3\}$ as explanatory variables. It turns out that only Q_{t-2} is statistically significant. See models **g1 and g3**. Write down the fitted model **g3**.

    A: The fitted model is

    $$P(Q_t = 1|F_{t-1}) = \frac{\exp(0.536 - 0.0687\mathrm{Q}_{t-2})}{1 + \exp(0.536 - 0.0687\mathrm{Q}_{t-2})}.$$

    Note: There was an error in the computer command. The correct coefficients are 0.144 and -0.275, respectively. You will receive the credits even if you used the incorrect coefficients. The correct answer is

    $$P(Q_t = 1|F_{t-1}) = \frac{\exp(0.144 - 0.275\mathrm{Q}_{t-2})}{1 + \exp(0.144 - 0.275\mathrm{Q}_{t-2})}.$$

2. What is the meaning of the fitted coefficient -0.0687?

    A: It is the logarithm of the odds-ratio.

3. What is $P(Q_t = 1|Q_{t-2} = 1)$?

    A: 0.615. Note: The answer with correct coefficient estimates is 0.467.

4. A 6-2-1 neural network with inputs $\{Q_{t-i}, M_{t-i} | i = 1, 2, 3\}$ is applied. See **g4**. Write down the model for the hidden node h_1.

    A: The model is

    $$h_{1t} = \frac{\exp(-20.2 - 2.78Q_{t-1} + 16.9Q_{t-2} - 4.62Q_{t-3} + 8.16M_{t-1} + 8.08M_{t-2} + 12.0M_{t-3})}{1 + \exp(-20.2 - 2.78Q_{t-1} + 16.9Q_{t-2} - 4.62Q_{t-3} + 8.16M_{t-1} + 8.08M_{t-2} + 12.0M_{t-3})}.$$

5. Write down the model for the output node.

   A:

$$O_t = \begin{cases} 1 & \text{if } L_t > 0 \\ 0 & \text{otherwise,} \end{cases}$$

   where $L_t = 0.29 + 0.74h_{1t} - 0.68h_{2t} - 0.15Q_{t-1} - 0.82Q_{t-2} - 0.03Q_{t-3} - 0.10M_{t-1} - 0.04M_{t-2} - 0.03M_{t-3}$.