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

Midterm

ChicagoBooth Honor Code:
I pledge my honor that I have not violated the Honor Code during this examination.

Signature: **Name**: **ID**:

Notes:

- Open notes and books. Exam time: 180 minutes.
- You may use a calculator or a PC. However, turn off Internet connection and cell phones. **Internet access and any communication are strictly prohibited during the exam.**
- The exam has 9 pages and the R output has 8 pages. Please **check** that you have all 17 pages.
- For each question, write your answer in the blank space provided.
- **Manage** your time carefully and answer as many questions as you can.
- For simplicity, if not specifically given, use 5% Type-I error in hypothesis testings.
- Round your answer to 3 significant digits.

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?

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.

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?

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 $\text{Var}(a_t)$.

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

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

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

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

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

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

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 $\sigma_{1000}(5)$?

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

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?

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

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

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.

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).

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

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?

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

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

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

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

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.

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.

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

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

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.

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

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?

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.

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.

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.

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

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`.

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

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

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.

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?

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

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

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.

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

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?

R output: edited to shorten the output

```
###### Problem A ###########
> da <- read.table("m-ibm6116.txt",header=T)
> require(rugarch)
> ibm <- log(da$RET+1)
> summer <- c(0,0,0,0,0,1,1,1,0,0,0,0)
> Sum <- rep(summer,56)
> Sum <- as.matrix(Sum)
> spec7 <- ugarchspec(variance.model=list(model="sGARCH",garchOrder=c(1,1),
external.regressors=Sum), mean.model=list(armaOrder=c(0,0)), distribution.model="std")
> k3 <- ugarchfit(data=ibm,spec=spec7)
> k3
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*
Conditional Variance Dynamics
-----------------------------------
GARCH Model     : sGARCH(1,1)
Mean Model      : ARFIMA(0,0,0)
Distribution    : std

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.008394    0.002338  3.59067 0.000330
omega   0.000302    0.000190  1.58972 0.111897
alpha1  0.100171    0.034104  2.93722 0.003312
beta1   0.827267    0.058591 14.11942 0.000000
summer  0.000133    0.000353  0.37792 0.705492
shape   9.070224    2.832075  3.20268 0.001362

LogLikelihood : 887.9616

Information Criteria
------------------------------------
Akaike       -2.6249
Bayes        -2.5846
Shibata      -2.6250
Hannan-Quinn -2.6093

### Problem B #########
> getSymbols("GOOG",from="XXXX'',to=''XXXX'')
> rtn <- diff(log(as.numeric(GOOG[,6])))
> Box.test(rtn,lag=10,type="Ljung")
        Box-Ljung test
data:  rtn
X-squared = 9.3271, df = 10, p-value = 0.5014
```

```
> basicStats(rtn)
                  rtn
nobs      2705.000000
NAs          0.000000
..
Mean         0.000516
...
SE Mean      0.000354
LCL Mean    -0.000177
UCL Mean     0.001210
....
Skewness     0.612749
Kurtosis    11.758266
> m1 <- garchFit(~garch(1,1),data=rtn,trace=F)
> summary(m1)
Title: GARCH Modelling
Call:  garchFit(formula = ~garch(1, 1), data = rtn, trace = F)

Mean and Variance Equation:
 data ~ garch(1, 1) [data = rtn]

Conditional Distribution: norm

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     6.891e-04   2.963e-04    2.326     0.02 *
omega  1.085e-05   2.115e-06    5.128 2.94e-07 ***
alpha1 8.124e-02   1.508e-02    5.387 7.18e-08 ***
beta1  8.904e-01   1.728e-02   51.528  < 2e-16 ***
---
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  16500.32  0
 Ljung-Box Test     R    Q(10)  4.986361  0.8920873
 Ljung-Box Test     R    Q(20)  13.17736  0.8696453
 Ljung-Box Test     R^2  Q(10)  3.296533  0.9735597
 Ljung-Box Test     R^2  Q(20)  6.529325  0.9979671

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-5.338750 -5.330021 -5.338754 -5.335594

> m2 <- garchFit(~garch(1,1),data=rtn,trace=F,cond.dist="std")
> summary(m2)
Title: GARCH Modelling
Call: garchFit(formula=~garch(1,1), data=rtn, cond.dist="std",trace = F)
```

```
Mean and Variance Equation:
 data ~ garch(1, 1) [data = rtn]

Conditional Distribution: std

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     7.419e-04   2.398e-04    3.094  0.00198 **
omega  1.976e-06   8.012e-07    2.466  0.01367 *
alpha1 3.182e-02   6.459e-03    4.927 8.35e-07 ***
beta1  9.618e-01   7.395e-03  130.069  < 2e-16 ***
shape  3.856e+00   2.911e-01   13.247  < 2e-16 ***
---
Standardised Residuals Tests:
                                Statistic p-Value
 Ljung-Box Test     R    Q(10)  3.719213  0.9591261
 Ljung-Box Test     R    Q(20)  11.02247  0.9456378
 Ljung-Box Test     R^2  Q(10)  1.448737  0.9990863
 Ljung-Box Test     R^2  Q(20)  3.600967  0.9999806

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-5.565413 -5.554502 -5.565420 -5.561468

> m3 <- garchFit(~garch(1,1),data=rtn,trace=F,cond.dist="sstd")
> summary(m3)
Title: GARCH Modelling
Call: garchFit(formula=~garch(1,1), data=rtn, cond.dist="sstd", trace=F)

Mean and Variance Equation: data ~ garch(1, 1) [data = rtn]

Conditional Distribution: sstd

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     8.717e-04   2.774e-04    3.142  0.00168 **
omega  2.021e-06   8.231e-07    2.455  0.01409 *
alpha1 3.252e-02   6.700e-03    4.854 1.21e-06 ***
beta1  9.612e-01   7.622e-03  126.111  < 2e-16 ***
skew   1.024e+00   2.631e-02   38.931  < 2e-16 ***
shape  3.834e+00   2.895e-01   13.242  < 2e-16 ***
---
Standardised Residuals Tests:
                                Statistic p-Value
  Ljung-Box Test     R    Q(10)  3.746294  0.9580672
 Ljung-Box Test     R    Q(15)  4.57212   0.9951711
```

```
 Ljung-Box Test     R^2  Q(10)  1.479565  0.9989976
 Ljung-Box Test     R^2  Q(20)  3.658553  0.9999778

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-5.564992 -5.551899 -5.565002 -5.560257

> m4 <- garchFit(~garch(1,1),data=rtn,trace=F,cond.dist="std",leverage=T)
> summary(m4)
Title: GARCH Modelling
Call: garchFit(formula = ~garch(1, 1), data = rtn, cond.dist = "std",
    leverage = T, trace = F)

Mean and Variance Equation: data ~ garch(1, 1)[data = rtn]

Conditional Distribution: std

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     6.774e-04   2.380e-04    2.846  0.00443 **
omega  2.481e-06   9.704e-07    2.557  0.01057 *
alpha1 2.875e-02   6.331e-03    4.541 5.60e-06 ***
gamma1 4.412e-01   1.024e-01    4.310 1.63e-05 ***
beta1  9.583e-01   8.438e-03  113.570  < 2e-16 ***
shape  3.892e+00   2.938e-01   13.244  < 2e-16 ***
---
Standardised Residuals Tests:
                                Statistic p-Value
 Ljung-Box Test     R    Q(10)  4.404431  0.9272639
 Ljung-Box Test     R    Q(20)  11.98351  0.9166422
 Ljung-Box Test     R^2  Q(10)  2.453224  0.99154
 Ljung-Box Test     R^2  Q(20)  4.124367  0.9999403

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-5.575489 -5.562396 -5.575499 -5.570755

> predict(m4,5)
  meanForecast  meanError standardDeviation
1 0.0006773902 0.01019445        0.01019445
2 0.0006773902 0.01025011        0.01025011
3 0.0006773902 0.01030477        0.01030477
4 0.0006773902 0.01035843        0.01035843
5 0.0006773902 0.01041113        0.01041113

> m5 <- garchM(rtn*100)  #### Percentage return
Maximized log-likehood:  -5232.231
```

```
Coefficient(s):
        Estimate  Std. Error  t value   Pr(>|t|)
mu    0.04916860  0.05390606  0.91212    0.36171
gamma 0.00798344  0.01821644  0.43825    0.66120
omega 0.10894921  0.02149213  5.06926 3.9937e-07 ***
alpha 0.08174536  0.01529616  5.34417 9.0830e-08 ***
beta  0.88982594  0.01756281 50.66534 < 2.22e-16 ***
---
> at <- (rtn-mean(rtn))*100
> m6 <- Igarch(at,include.mean=F,volcnt=T)
Estimates:  0.07984549 0.8783669
Maximized log-likehood:  5241.394

Coefficient(s):
       Estimate  Std. Error  t value   Pr(>|t|)
omega 0.0798455   0.0154414  5.17086 2.3301e-07 ***
beta  0.8783669   0.0160542 54.71260 < 2.22e-16 ***
---
> at[2705]
[1] -0.1438779
> m6$volatility[2705]
[1] 0.9892536
>
##### Problem C ######
> da <- read.table("m-bnd.txt",header=T)
> Aaa <- da$Aaa; Baa <- da$Baa
> n1 <- lm(Baa~Aaa)
> summary(n1)
Call: lm(formula = Baa ~ Aaa)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.100177   0.033594   2.982  0.00298 **
Aaa         1.114989   0.004254 262.080  < 2e-16 ***
---
Residual standard error: 0.2828 on 607 degrees of freedom
Multiple R-squared:  0.9912,    Adjusted R-squared:  0.9912
F-statistic: 6.869e+04 on 1 and 607 DF,  p-value: < 2.2e-16

> Box.test(n1$residuals,lag=12,type="Ljung")
        Box-Ljung test
data:  m1$residuals
X-squared = 2608.2, df = 12, p-value < 2.2e-16

> dA <- diff(Aaa); dB <- diff(Baa)
> n2 <- lm(dB~-1+dA)
> summary(n2)
```

```
Call: lm(formula = dB ~ -1 + dA)

Coefficients:
   Estimate Std. Error t value Pr(>|t|)
dA  0.81739    0.01663   49.17   <2e-16 ***
---
Residual standard error: 0.09082 on 607 degrees of freedom
Multiple R-squared:  0.7993,    Adjusted R-squared:  0.799

> Box.test(n2$residuals,lag=12,type="Ljung")
        Box-Ljung test
data:  m2$residuals
X-squared = 63.481, df = 12, p-value = 5.197e-09
> acf(n2$residuals)
> pacf(n2$residuals)
> n3 <- arima(dB,order=c(0,0,5),xreg=dA,include.mean=F)
> n3
Call: arima(x = dB, order = c(0, 0, 5), xreg = dA, include.mean = F)

Coefficients:
         ma1     ma2      ma3     ma4     ma5      dA
      0.2942  0.0278  -0.0008  0.0060  0.0869  0.7938
s.e.  0.0419  0.0442   0.0419  0.0436  0.0407  0.0188

sigma^2 estimated as 0.007513:  log likelihood = 624.12,  aic = -1234.23
> n4 <-arima(dB,order=c(0,0,1),seasonal=list(order=c(0,0,1),period=5),xreg=dA,
include.mean=F)
> n4
Call: arima(x=dB, order=c(0,0,1), seasonal=list(order = c(0,0,1), period=5),
    xreg = dA, include.mean = F)
Coefficients:
         ma1    sma1      dA
      0.2889  0.0932  0.7980
s.e.  0.0383  0.0398  0.0178

sigma^2 estimated as 0.007509:  log likelihood = 624.29,  aic = -1240.58
> tsdiag(n4,gof=24)
> which.max(abs(n4$residuals))
[1] 312
> length(dB)
[1] 608
> I312 <- rep(0,608)
> I312[312] <- 1
> X <- cbind(dA,I312)
> n5 <- arima(dB,order=c(0,0,1),seasonal=list(order=c(0,0,1),period=5),
xreg=X,include.mean=F)
> n5
```

```
Call:arima(x=dB, order=c(0,0,1), seasonal=list(order=c(0,0,1), period=5),
    xreg = X, include.mean = F)

Coefficients:
         ma1    sma1      dA     I312
      0.3079  0.1410  0.8060  -0.4873
s.e.  0.0409  0.0407  0.0174   0.0807

sigma^2 estimated as 0.007091:  log likelihood = 641.66,  aic = -1273.32
> tsdiag(n5,gof=24)
>
> source("backtest.R")
> backtest(n4,dB,orig=550,xre=dA,include.mean=F)
[1] "RMSE of out-of-sample forecasts"
[1] 0.09120028
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.06315827
> backtest(n5,dB,orig=550,xre=X,include.mean=F)
[1] "RMSE of out-of-sample forecasts"
[1] 0.09120028
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.06315827

> require(forecast)
> auto.arima(dB)
Series: dB
ARIMA(1,0,1) with zero mean
> n6 <- arima(dB,order=c(1,0,1),include.mean=F)
> n6
Call:
arima(x = dB, order = c(1, 0, 1), include.mean = F)

Coefficients:
         ar1     ma1
      0.1939  0.3229
s.e.  0.0782  0.0744

sigma^2 estimated as 0.03212:  log likelihood = 182.35,  aic = -358.71
> backtest(n6,dB,orig=550,include.mean=F)
[1] "RMSE of out-of-sample forecasts"
[1] 0.208794
[1] "Mean absolute error of out-of-sample forecasts"
[1] 0.1574626

##### Problem D #######
> getSymbols("KO",from="2004-01-03")
> ko <- diff(log(as.numeric(KO[,6])))
```

```
> spec3 <- ugarchspec(variance.model=list(model="eGARCH"),
 mean.model=list(armaOrder=c(0,0))
> kk <- ugarchfit(data=ko,spec=spec3)
> kk
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*
Conditional Variance Dynamics
-----------------------------------
GARCH Model     : eGARCH(1,1)
Mean Model      : ARFIMA(0,0,0)
Distribution    : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000213    0.000154   1.3797  0.16769
omega  -0.314077    0.025828 -12.1603  0.00000
alpha1 -0.084652    0.011678  -7.2491  0.00000
beta1   0.965124    0.002714 355.5522  0.00000
gamma1  0.178557    0.020001   8.9273  0.00000

LogLikelihood : 10827.23

Information Criteria
------------------------------------
Akaike       -6.4591
Bayes        -6.4500
Shibata      -6.4591
Hannan-Quinn -6.4558
```