Graduate School of Business, University of Chicago
Business 41202, Spring Quarter 2007, Mr. Ruey S. Tsay

Midterm

GSB 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.
- Write your answer in the blank space provided for each question.
- **Manage** your time carefully and answer as many questions as you can.
- The exam has 7 pages and the R output has 7 pages. The output of S-Plus will be provided in class. Please **check** to make sure that you have all the pages.
- For simplicity, **ALL** tests use the 5% significance level.
- Round your answer to 2 significant digits.

Circle the output used in your answer: (a) R, (b) S-Plus.

Problem A: (30 pts) Answer briefly the following questions.

1. Describe a situation under which the R^2 defined as

$$R^2 = \frac{\text{(Sum of squares of total) - (Sum of squares of residuals)}}{\text{Sum of squares of total}},$$

is not informative in evaluating a fitted time series model.

2. Consider a linear regression model with time-series errors. Why is the Durbin-Watson statistics not sufficient in model checking?

3. **For questions 3 to 5**, consider the AR(1)-IGARCH(1,1) model

$$r_t = 0.02 + 0.2r_{t-1} + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0,1).$$

$$\sigma_t^2 = 0.1a_{t-1}^2 + 0.9\sigma_{t-1}^2.$$

What is the expected value of r_t?, i.e., $E(r_t) =$?

4. Suppose that $r_{h-1} = 0.04$, what are the 1-step ahead forecast and its forecast error of r_t at the forecast origin $h-1$?, i.e. $r_{h-1}(1) =$? and $e_{h-1}(1) =$?

5. In addition to the information of the prior question, suppose we also observe that $r_h = -0.012$ and $\sigma_h^2 = 0.25$. What are the 1-step and 2-step ahead volatility forecasts of the model at time origin h? That is, what are $\sigma_h^2(1)$ and $\sigma_h^2(2)$?

6. Give two advantages of EGARCH models over the GARCH models.

7. **For problems 7 to 9**, consider the daily exchange rate between U.S. dollar and U.K. pound from January 2001 to April 26, 2007. Descriptive statistics of the daily log returns are given in the attached output. Is the mean of the log return different from zero? Why?

8. Is the distribution of the log return symmetric with respect to its mean? Why?

9. Does the distribution of the log return have heavy tails? Why?

10. Suppose that the monthly time series r_t follows the model

$$r_t = (1 - \theta_2 B^2)(1 - \theta_{12} B^{12})a_t, \quad a_t \sim N(0, \sigma_a^2),$$

where θ_2 and θ_{12} are non-zero real numbers satisfying $|\theta_2| < 1$ and $|\theta_{12}| < 1$, and $\sigma_a^2 > 0$. List all non-zero autocorrelations of r_t.

11. Give two reasons that observed daily returns of an asset are serially correlated even though the true underlying returns are serially uncorrelated.

12. To test for ARCH effect, one often employs the Ljung-Box statistics $Q(m)$ of the squared residuals of the mean equation. Write down the null and alternative hypotheses for $Q(10)$ statistic in ARCH-effect testing.

13. Assume that time series x_t and y_t follow the following models,

$$
\begin{aligned}
x_t &= 0.5x_{t-1} + a_t, \\
y_t &= 1.3y_{t-1} - 0.4y_{t-2} + a_t,
\end{aligned}
$$

where $\{a_t\}$ are *iid* $N(0, \sigma_a^2)$ with $\sigma_a^2 > 0$. Both series are mean reverting. What is the half-life for x_t? What is the half-life of y_t?

14. Suppose that your average daily balance of a credit card is \$1000. Suppose also that the card charges an interest rate of 22.5% per annum (daily compounding). How much is your financial charge in a 30-day billing cycle?

15. Suppose that the monthly log returns of an asset are normally distributed with mean 0.08 and standard deviation 0.12. What is the mean of the monthly simple return of the asset?

Problem B. (20 pts) Consider Moody's seasoned AAA and BAA corporate bond yields from January 5, 1962 to April 20, 2007. The data are averages of daily yields and obtained from the Federal Reserve Bank of St. Louis. Denote the bond yields by AAA and BAA, respectively. To find the relationship between the two bond yields, we conduct certain analysis. The output is attached. Answer the following questions.

1. Write down the fitted linear regression with BAA and AAA representing the dependent and independent variable, respectively. What is the R^2 of the linear regression? Is the fitted model adequate? Why?

2. Let $Y_t = BAA_t - BAA_{t-1}$ and $X_t = AAA_t - AAA_{t-1}$ be the differenced series. Consider the linear regression $Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$. What is the fitted model? What is the residual standard deviation of the model?

3. The residuals of the prior linear regression show certain serial correlations. A linear regression model with time series errors is employed. Write down the fitted model. Based on the available output, is this model adequate? Why?

4. Consider the above linear regression model with time-series error. One way to confirm that the MA(2) model is needed is to test the lag-2 MA coefficient. Write down the null and alternative hypotheses for such a test. What is the test statistic? Drawn your conclusion.

5. Construct a 95% confidence interval for the coefficient β_1 (the slope parameter of the linear regression model with time-series errors). Is the estimate 0.719 (see Question 2) in the 95% confidence interval? Discuss the implication of the result.

Problem C. (30 pts) Consider the daily closing values of the VIX index (which is an implied volatility for the S&P 500 index) of CBOE from January 2, 2004 to April 5, 2007. The index appears to have a unit root so that we analyze its log return series. The relevant compute output is attached. Answer the following questions.

1. (4 points) Write down the fitted mean equation for the log return series, including the residual variance. Is the model adequate in handling the serial correlations? Why?

2. Is there any ARCH effect in the log return series? Why?

3. A GARCH(1,1) model is used in the volatility equation. Write down the fitted model, including the degrees of freedom of the Student-t innovations.

4. Based on the output, what are the estimated standard errors of ARCH (α_1) and GARCH (β_1) coefficients?

5. (8 points) A GJR (or TGARCH) model is also fitted to the log return series. Write down the fitted model.

6. Is the fitted GJR (or TGARCH) model adequate? Why?

7. (4 points) Between the GARCH(1,1) and GJR(1,1) models, which one is preferred? Why?

8. Is the leverage effect of the GJR model significant? Why? Why is the leverage parameter negative?

9. (5 points) To better understand the leverage effect, use the fitted GJR or TGARCH model to calculate the ratio $\frac{\sigma_t^2(\epsilon_{t-1}=-2)}{\sigma_t^2(\epsilon_{t-1}=2)}$, where $\{\epsilon_t\}$ denotes the standardized innovation. For simplicity, you may ignore the constant term of the volatility equation.

10. (4 points) Based on the fitted GJR or TGARCH model, what are the 1-step and 5-step ahead forecasts of the log return and its volatility at the forecast origin $T = 820$, the last data point?

Problem D. (20 pts) Consider the quarterly earnings per share of the FedEx stock from the fourth quarter of 1991 to the last quarter of 2006. The data were obtained from First Call. To take the log transformation, we add one to all data points. Compute output is attached. Let $x_t = \ln(y_t + 1)$ be the transformed earnings, where y_t is the actual earnings per share.

1. (5 points) Write down the fitted model for x_t, including the variance of the residuals.

2. (4 points) Is there any significant serial correlation in the residuals of the fitted model? Why?

3. (4 points) Let $T = 62$ be the forecast origin. Based on the fitted model, and, for simplicity, use the relationship $y_t = \exp(x_t) - 1$, what are the 1-step and 2-step ahead forecasts of earnings per share for the FedEx stock?

4. (3 points) Obtain a 95% interval forecast for x_{63} at the forecast origin $T = 62$.

5. Test the null hypothesis $H_o : \theta_4 = 0$ vs $H_a : \theta_4 \neq 0$. What is the test statistic? Draw your conclusion.

R output. (S-Plus output will be given in class.)

Questions 7-9, Problem A.

```
> da=read.table("d-usuk0107.txt")
> dim(da)
[1] 1588    4
> da[1,]
    V1 V2 V3     V4
1 2001  1  2 1.4977
> fx=da[,4]
> fx=log(fx)
> basicStats(diff(fx))
            round.ans..digits...6.
nobs                   1587.000000
NAs                       0.000000
Minimum                  -0.021707
Maximum                   0.020930
1. Quartile              -0.002747
3. Quartile               0.003338
Mean                      0.000179
Median                    0.000281
Sum                       0.284355
SE Mean                   0.000129
LCL Mean                 -0.000074
UCL Mean                  0.000432
Variance                  0.000026
Stdev                     0.005139
Skewness                 -0.142401
Kurtosis                  0.597221
```

Problem B

```
> da=read.table("w-aaa.txt")
> aaa=da[,4]
> da1=read.table("w-baa.txt")
> baa=da1[,4]

> m0=lm(baa~aaa)
> summary(m0)

lm(formula = baa ~ aaa)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.030487   0.019636  -1.553    0.121
```

```
aaa          1.128573   0.002369 476.464   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.279 on 2362 degrees of freedom
Multiple R-Squared: 0.9897,     Adjusted R-squared: 0.9897
F-statistic: 2.27e+05 on 1 and 2362 DF,  p-value: < 2.2e-16

> Box.test(m0$residuals,lag=10,type='Ljung')
        Box-Ljung test
data:  m0$residuals
X-squared = 16920.04, df = 10, p-value < 2.2e-16

> y=diff(baa)
> x=diff(aaa)
> plot(x,y)
> m1=lm(y~x)
> summary(m1)

lm(formula = y ~ x)

Residuals:
       Min         1Q     Median         3Q        Max
-0.3083274 -0.0217853 -0.0002261  0.0196215  0.3625531

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.0002261  0.0009079   0.249    0.803
x           0.7186425  0.0095404  75.326   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04413 on 2361 degrees of freedom
Multiple R-Squared: 0.7062,     Adjusted R-squared: 0.706
F-statistic:  5674 on 1 and 2361 DF,  p-value: < 2.2e-16

> m2=arima(y,order=c(0,0,2),xreg=x)
> m2

arima(x = y, order = c(0, 0, 2), xreg = x)

Coefficients:
         ma1     ma2  intercept      x
      0.2475  0.0974     0.0002  0.692
```

```
s.e.  0.0209  0.0199     0.0012  0.010

sigma^2 estimated as 0.001823:  log likelihood = 4099.37,  aic = -8188.74

> Box.test(m2$residuals,lag=10,type='Ljung')
        Box-Ljung test

data:  m2$residuals
X-squared = 9.483, df = 10, p-value = 0.487
```

Problem C

```
> da=read.table("vix07.txt",header=T)
> dim(da)
[1] 821   7

> vix=log(da[,7])
> acf(diff(vix))
> rtn=diff(vix)
> m1=arima(rtn,order=c(0,0,2))
> m1

Call:
arima(x = rtn, order = c(0, 0, 2))

Coefficients:
          ma1      ma2  intercept
      -0.1025  -0.1167    -0.0004
s.e.   0.0349   0.0367     0.0016

sigma^2 estimated as 0.00326:  log likelihood = 1184.15,  aic = -2360.3

> Box.test(m1$residuals,lag=5,type='Ljung')
        Box-Ljung test

data:  m1$residuals
X-squared = 4.6809, df = 5, p-value = 0.4561

> Box.test(m1$residuals,lag=10,type='Ljung')
        Box-Ljung test

data:  m1$residuals
X-squared = 17.4477, df = 10, p-value = 0.06503

> Box.test(m1$residuals^2,lag=10,type='Ljung')
```

```
        Box-Ljung test

data:  m1$residuals^2
X-squared = 58.3179, df = 10, p-value = 7.531e-09

> m2=garchOxFit(formula.mean=~arma(0,2),formula.var=~garch(1,1),series=rtn,cond.dist="t")

 ********************
 ** SPECIFICATIONS **
 ********************
Dependent variable : X
Mean Equation : ARMA (0, 2) model.
No regressor in the mean
Variance Equation : GARCH (1, 1) model.
 No regressor in the variance
The distribution is a Student distribution, with 4.9587 degrees of freedom.

Strong convergence using numerical derivatives
Log-likelihood = 1283.29

 Maximum Likelihood Estimation (Std.Errors based on Second derivatives)
                  Coefficient  Std.Error  t-value  t-prob
Cst(M)              -0.002884  0.0012583   -2.292  0.0222
MA(1)               -0.102151   0.034944   -2.923  0.0036
MA(2)               -0.110580   0.036894   -2.997  0.0028
Cst(V)               2.093952    0.86713    2.415  0.0160
ARCH(Alpha1)         0.086887   ????????    3.156  0.0017
GARCH(Beta1)         0.844596   ????????    19.07  0.0000
Student(DF)          4.958702    0.82290    6.026  0.0000

No. Observations :       820  No. Parameters  :         7
Mean (Y)         :  -0.00039  Variance (Y)    :   0.00333
Skewness (Y)     :   1.03411  Kurtosis (Y)    :  11.83534
Log Likelihood   :  1283.292  Alpha[1]+Beta[1]:   0.93148

Warning : To avoid numerical problems, the estimated parameter
Cst(V), and its std.Error have been multiplied by 10^4.

AIC = -3.1129.

> m3=garchOxFit(formula.mean=~arma(0,2),formula.var=~gjr(1,1),series=rtn,cond.dist="t")

 ********************
 ** SPECIFICATIONS **
```

```
 ********************
Dependent variable : X
Mean Equation : ARMA (0, 2) model.
No regressor in the mean
Variance Equation : GJR (1, 1) model.
 No regressor in the variance
The distribution is a Student distribution, with 5.09874 degrees of freedom.

Strong convergence using numerical derivatives
Log-likelihood = 1287.68

 Maximum Likelihood Estimation (Std.Errors based on Second derivatives)
                  Coefficient  Std.Error  t-value  t-prob
Cst(M)              -0.002423  0.0012987   -1.865  0.0625
MA(1)               -0.097371   0.034658   -2.809  0.0051
MA(2)               -0.099173   0.037309   -2.658  0.0080
Cst(V)               2.026803    0.93400    2.170  0.0303
ARCH(Alpha1)         0.120076   0.038681    3.104  0.0020
GARCH(Beta1)         0.867060   0.048342    17.94  0.0000
GJR(Gamma1)         -0.138640   0.047364   -2.927  0.0035
Student(DF)          5.098737    0.85461    5.966  0.0000

No. Observations :       820  No. Parameters  :         8
Mean (Y)         :  -0.00039  Variance (Y)    :   0.00333
Skewness (Y)     :   1.03411  Kurtosis (Y)    :  11.83534
Log Likelihood   :  1287.682

Warning : To avoid numerical problems, the estimated parameter
Cst(V), and its std.Error have been multiplied by 10^4.

***************
** FORECASTS **
***************
Number of Forecasts: 15

Horizon       Mean   Variance
      1  0.0005288   0.003373
      2  -0.001627   0.003096
      3  -0.002423   0.002841
      4  -0.002423   0.002608
      5  -0.002423   0.002393
     .....
     15  -0.002423   0.001015
---------------
```

```
***********
** TESTS **
***********
                  Statistic       t-Test      P-Value
Skewness             1.6790       19.664  4.3931e-086
Excess Kurtosis      12.176       71.387      0.00000
Jarque-Bera          5450.5         .NaN      0.00000
---------------
Information Criterium (to be minimized)
Akaike         -3.121175  Shibata        -3.121363
Schwarz         -3.075231  Hannan-Quinn    -3.103546
---------------
Q-Statistics on Standardized Residuals
  --> P-values adjusted by 2 degree(s) of freedom
  Q( 10) =  13.1406   [0.1071025]
  Q( 15) =  16.1934   [0.2388416]
  Q( 20) =  17.7944   [0.4692696]
H0 : No serial correlation ==> Accept H0 when prob. is High [Q < Chisq(lag)]
--------------
Q-Statistics on Squared Standardized Residuals
  --> P-values adjusted by 2 degree(s) of freedom
  Q( 10) =  2.49384   [0.9620177]
  Q( 15) =  3.15659   [0.9973083]
  Q( 20) =  3.75639   [0.9998498]
H0 : No serial correlation ==> Accept H0 when prob. is High [Q < Chisq(lag)]
--------------
```

Problem D

```
> da=read.table("q-earn-fdx.txt")
> fdx=da[,4]
> plot(fdx,type='l')
> min(fdx)
[1] -0.07
> x=log(fdx+1)
> plot(x,type='l')
> acf(x)
> acf(diff(x))
> acf(diff(diff(x),4))

> m4=arima(x,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=4))
> m4

arima(x = x, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 4))
```

```
Coefficients:
          ma1     sma1
      -0.7215  -0.382
s.e.   0.0937   0.116

sigma^2 estimated as 0.007214:  log likelihood = 58.88,  aic = -111.76
> tsdiag(m4,gof.lag=12)
> Box.test(m4$residuals,lag=12)

        Box-Pierce test

data:  m4$residuals
X-squared = 9.9519, df = 12, p-value = 0.6202

> 1-pchisq(9.95,10)
[1] 0.4448909

> predict(m4,5)
$pred
Time Series:
Start = 63
End = 67
Frequency = 1
[1] 0.9722516 1.1631526 1.0554812 1.1813262 1.0954462

$se
Time Series:
Start = 63
End = 67
Frequency = 1
[1] 0.08493732 0.08816912 0.09128657 0.09430102 0.12120565
```