Graduate School of Business, University of Chicago
Business 41202, Spring Quarter 2008, 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 8 pages. 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.

Problem A: (30 pts) Answer briefly the following questions. Each question has two points.

1. Describe two methods for choosing a time series model.

2. Describe two applications of volatility in finance.

3. Give two applications of seasonal time series models in finance.

4. Describe two weaknesses of the ARCH models in modelling stock volatility.

5. Give two empirical characteristics of daily stock returns.

6. The daily simple returns of Stock A for the last week were 0.02, 0.01, -0.005, -0.01, and 0.025, respectively. What is the weekly log return of the stock last week? What is the weekly simple return of the stock last week?

7. Suppose the closing price of Stock B for the past three trading days were \$100, \$120, and \$100, respectively. What is the arithmetic mean of the simple return of the stock for the past three days? What is the geometric average of the simple return of the stock for the past three days?

8. Consider the AR(1) model $r_t = 0.02 + 0.8r_{t-1} + a_t$, where the shock a_t is normally distributed with mean zero and variance 1. What are the variance and lag-1 autocorrelation function of r_t?

9. **For problems 6 and 7**, suppose the daily return r_t, in percentages, of Stock A follows the model $r_t = 1.0 + a_t + 0.3a_{t-1}$, where $a_t = \sigma_t\epsilon_t$ with $\sigma_t^2 = 1.0 + 0.4a_{t-1}^2$ and ϵ_t being standard normal. What is the unconditional variance of a_t? What is the variance of r_t?

10. Suppose that $a_n = 3.0$, what is the 1-step ahead forecast for r_{n+1} at the forecast origin n? What is the 1-step ahead volatility forecast of r_t at the forecast origin n?

11. Consider the simple AR(1) model $r_t = 100+0.8r_{t-1}+a_t$, where a_t is normally distributed with mean zero and variance 10. Is the r_t series mean-reverting? If yes, what is the half-life of the series?

12. Describe two test statistics for testing the ARCH effect of an asset return series. Write down the associated null hypotheses.

13. Consider the following two IGARCH(1,1) models for percentage log returns:

$$\begin{aligned} \text{Model A}: \quad \sigma_t^2 &= 1.0 + 0.1a_{t-1}^2 + 0.9\sigma_{t-1}^2 \\ \text{Model B}: \quad \sigma_t^2 &= 0.1a_{t-1}^2 + 0.9\sigma_{t-1}^2. \end{aligned}$$

Suppose that $\sigma_{100}^2 = 20$ and $a_{100} = -2.0$. What are the 3-step ahead volatility forecasts for Models A and B?

14. Consider the following two models for the log price of an asset:

$$\begin{aligned} \text{Model A}: \quad p_t &= p_{t-1} + a_t \\ \text{Model B}: \quad p_t &= 0.00001 + p_{t-1} + a_t \end{aligned}$$

where the shock a_t is normally distributed with mean zero and variance $\sigma^2 > 0$. Suppose further that $p_{100} = 5$. Let $p_n(\ell)$ be the ℓ-step ahead forecast at the forecast origin n. What are the point forecasts $p_{100}(\ell)$ for both models as $\ell \to \infty$?

15. Suppose that we have $T = 1000$ daily log returns for the Decile 1 portfolio. Suppose further that the sample autocorrelation at lag-12 is $\hat{\rho}_{12} = 0.15$. Test the hypothesis $H_o : \rho_{12} = 0$ against the alternative hypothesis $H_a : \rho_{12} \neq 0$. Compute the test statistic and draw your conclusion.

Problem B. (20 pts) It is well-known in economics that growth rate of the domestic gross product (GDP) is negatively correlated with the change in unemployment rate. Consider the U.S. quarterly real GDP and unemployment rate from the first quarter of 1948 to the first quarter of 2008. Let dgdp_t be the growth rate of the GDP, i.e. $\text{dgdp}_t = \ln(GDP_t) - \ln(GDP_{t-1})$, and dun_t be the change in unempolyment rate, i.e. $\text{dun}_t = U_t - U_{t-1}$ with U_t being the civilian unemployment rate. The data were seasonally adjusted and obtained from the Federal Reserve Bank at St. Louis. The sample size after the differencing is 240. Use the attached R output to answer the following questions.

1. (5 points) Write down the fitted linear regression model with dgdp_t and dun_t representing the dependent and independent variable, respectively, including residual standard error. What is the R^2 of the linear regression? Is the fitted model adequate? Why?

2. (5 points) To take care of the serial correlations in the residuals, a linear regression model with time-series errors is built for the two variables. Write down the fitted model, including the residual variance.

3. (2 points) Is the model in Question 2 adequate? Why?

4. (4 points) Based on the fitted model in Question 2, is the growth rate of GDP negatively correlated with the change in unempolyment rate? Why?

5. (4 points) To check the predictive power of the model, it was re-estimated using the first 236 data points. This re-fitted model is used to produce 1-step to 4-step ahead forecasts at the forecast origin $t = 236$. The actual value of the GDP growth rates are also given. Construct the 1-step ahead 95% interval forecast of the model. Is the actual growth rate in the forecasting interval?

Problem C. (16 pts) Consider the quarterly earnings per share of the Microsoft stock from the first quarter of 1992 to the first quarter of 2008. The data were obtained from First Call. To take the log transformation, **we add 0.5 to all data points**. The R output is attached. Let $x_t = \ln(y_t + 0.5)$ 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. (2 points) Is there any significant serial correlation in the residuals of the fitted model? Why?

3. (4 points) Let $T = 65$ be the forecast origin, where T is the sample size. Based on the fitted model, and, for simplicity, use the relationship $y_t = \exp(x_t) - 0.5$, what are the 1-step and 2-step ahead forecasts of earnings per share for the Microsoft stock?

4. (2 points) Test the null hypothesis $H_o : \theta_4 = 0$ vs $H_a : \theta_4 \neq 0$. What is the test statistic? Draw your conclusion.

5. (3 points) Consider the regular (i.e., non-seasonal) part of the MA model. Is it invertible? Why?

Problem D. (34 pts) Consider the daily log returns of the Starbucks stock, in percentages, from January 1993 to December 2007. The relevant R output is attached. Answer the following questions.

1. (2 points) Is the mean log return significant different from zero? Why?

2. (2 points) Is there any serial correlation in the log return series? Why?

3. (2 points) An MA model is used to handle the mean equation, which appears to be adequate. Is there any ARCH effect in the return series? Why?

4. (6 points) A GARCH(1,1) model with Student-t distribution is used for the volatility equation. Write down the fitted model, including the degrees of freedom of the Student-t innovations and mean equation.

5. (4 points) Since the constant term of the GARCH(1,1) model is not significantly different from zero at the 1% level, an IGARCH(1,1) model is used. Write down the fitted IGARCH(1,1) model, including the mean equation.

6. (3 points) Is the IGARCH(1,1) model adequate? Why? What is the 3-step ahead volatility forecast with the last data point as the forecast origin?

7. (5 points) A GJR (or TGARCH) model with Student-t distribution is also fitted to the log return series. Write down the fitted model, including the mean equation and all parameters.

8. (2 points) Is the fitted GJR (or TGARCH) model adequate? Why?

9. (2 points) Among the GARCH(1,1), IGARCH(1,1) and GJR(1,1) models, which one is preferred? Why?

10. (2 points) Is the leverage effect of the GJR model significant? Why?

11. (4 points) To better understand the leverage effect, use the fitted GJR to calculate the ratio $\frac{\sigma_t^2(a_{t-1}=-5.10)}{\sigma_t^2(a_{t-1}=5.10)}$, assuming $\sigma_{t-1}^2 = 7.5$.

```
**************PROBLEM B ***************** PROBLEM B
> x=read.table("q-gdp4808.txt")
> gdp=log(x[,4])
> x=read.table("q-unrate4808.txt")
> un=x[,1]

> dgdp=diff(gdp)
> dun=diff(un)

> m1=lm(dgdp~dun)
> summary(m1)
Call:
lm(formula = dgdp ~ dun)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0214767 -0.0056257 -0.0008208  0.0047532  0.0342197

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.016744   0.000566   29.59   <2e-16 ***
dun         -0.017437   0.001475  -11.82   <2e-16 ***
---
Residual standard error: 0.008767 on 238 degrees of freedom
Multiple R-squared:  0.37,      Adjusted R-squared: 0.3673
F-statistic: 139.8 on 1 and 238 DF,  p-value: < 2.2e-16

> Box.test(m1$residuals,lag=12,type='Ljung')

        Box-Ljung test

data:  m1$residuals
X-squared = 219.3962, df = 12, p-value < 2.2e-16

> m2=arima(dgdp,xreg=dun,order=c(2,0,0),seasonal=list(order=c(1,0,1),period=4))
> m2
Call:
arima(x = dgdp, order = c(2, 0, 0), seasonal = list(order = c(1, 0, 1), period = 4),
    xreg = dun)

Coefficients:
         ar1     ar2    sar1     sma1  intercept      dun
      0.2057  0.1236  0.8563  -0.7183     0.0165  -0.0178
s.e.  0.0665  0.0693  0.0761   0.0996     0.0014   0.0014
```

```
sigma^2 estimated as 6.077e-05:  log likelihood = 824.14,  aic = -1634.28

> Box.test(m2$residuals,lag=12,type='Ljung')

        Box-Ljung test

data:  m2$residuals
X-squared = 17.5298, df = 12, p-value = 0.1307

> source("r-fore.txt")
> forecast(m2,dgdp,236,4,xre=dun)
$pred
Time Series:
Start = 237
End = 240
Frequency = 1
[1] 0.01197725 0.01000377 0.01067890 0.01137371
$se
Time Series:
Start = 237
End = 240
Frequency = 1
[1] 0.007729502 0.007837210 0.007929924 0.007971967

*** Actual values *****
> dgdp[237:240]
[1] 0.015878404 0.014542801 0.007395370 0.007855834


**************** PROBLEM C ******************* PROBLEM C
> x=read.table("q-earn-msft92.txt")
> earn=x[,4]
> y=log(earn+0.5)

> m3=arima(y,order=c(0,1,2),seasonal=list(order=c(0,0,1),period=4))
> m3
Call:
arima(x = y, order = c(0, 1, 2), seasonal = list(order = c(0, 0, 1), period = 4))

Coefficients:
          ma1     ma2    sma1
      -0.6953  0.3889  0.3912
s.e.   0.1244  0.1219  0.1442
```

```
sigma^2 estimated as 0.00164:  log likelihood = 113.68,  aic = -219.37
> tsdiag(m3,gof.lag=12)  % Not shown, but checked.

> Box.test(m3$residuals,lag=12,type='Ljung')

        Box-Ljung test

data:  m3$residuals
X-squared = 9.6495, df = 12, p-value = 0.6467

> predict(m3,3)
$pred
Time Series:
Start = 66
End = 68
Frequency = 1
[1] 0.05458936 0.04296237 0.06145849

$se
Start = 66
End = 68
Frequency = 1
[1] 0.04049557 0.04233408 0.05080440

> mp=predict(m3,3)

> exp(mp$pred)-0.5
Time Series:
Start = 66
End = 68
Frequency = 1
[1] 0.5561068 0.5438986 0.5633864

***** PROBLEM D ************************ PROBLEM D
> x=read.table("d-sbuxsp9307.txt")
> sbux=log(x[,2]+1)*100

> basicStats(sbux)
                   sbux
nobs        3778.000000
NAs            0.000000
Minimum      -33.248699
Maximum       13.720128
```

```
Mean           0.076066
SE Mean        0.044029
LCL Mean      -0.010258
UCL Mean       0.162390
Variance       7.324015
Stdev          2.706292
Skewness      -0.304982
Kurtosis       9.159948

> Box.test(sbux,lag=15,type='Ljung')

        Box-Ljung test

data:  sbux
X-squared = 38.3907, df = 15, p-value = 0.0007898

> acf(sbux)  % Not shown, but shows two significant ACFs at lags 1 and 2.

> m1=arima(sbux,order=c(0,0,2))
> m1
Call:
arima(x = sbux, order = c(0, 0, 2))

Coefficients:
          ma1      ma2  intercept
      -0.0666  -0.0450     0.0761
s.e.   0.0163   0.0165     0.0390

sigma^2 estimated as 7.275:  log likelihood = -9109.24,  aic = 18226.49

> Box.test(m1$residuals,lag=15,type='Ljung')

        Box-Ljung test

data:  m1$residuals
X-squared = 14.4624, df = 15, p-value = 0.4908

> Box.test(m1$residuals^2,lag=15,type='Ljung')

        Box-Ljung test

data:  m1$residuals^2
X-squared = 112.6057, df = 15, p-value < 2.2e-16
```

```
> v1=garchOxFit(formula.mean=~arma(0,2),formula.var=~garch(1,1),series=sbux,
cond.dist="t")
 ********************
 ** SPECIFICATIONS **
 ********************
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 5.2718 degrees of freedom.

 Maximum Likelihood Estimation (Std.Errors based on Second derivatives)
                  Coefficient  Std.Error  t-value  t-prob
Cst(M)               0.036952   0.029768    1.241  0.2146
MA(1)               -0.042578   0.015755   -2.703  0.0069
MA(2)               -0.047814   0.015792   -3.028  0.0025
Cst(V)               0.012330  0.0073447    1.679  0.0933
ARCH(Alpha1)         0.026007  0.0062032    4.192  0.0000
GARCH(Beta1)         0.972732  0.0064010    152.0  0.0000
Student(DF)          5.271797    0.43966    11.99  0.0000

No. Observations :      3778  No. Parameters  :         7
Mean (Y)         :   0.07607  Variance (Y)    :   7.32208
Skewness (Y)     :  -0.30510  Kurtosis (Y)    :  12.16639
Log Likelihood   : -8602.449  Alpha[1]+Beta[1]:   0.99874

** TESTS **
***********
Information Criterium (to be minimized)
Akaike          4.557676  Shibata         4.557669
Schwarz         4.569232  Hannan-Quinn    4.561784
---------------
Q-Statistics on Standardized Residuals
  --> P-values adjusted by 2 degree(s) of freedom
  Q( 10) =  2.38597   [0.9668363]
  Q( 15) =  11.3430   [0.5821095]
  Q( 20) =  19.1817   [0.3807155]
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) =  6.00286   [0.6469113]
  Q( 15) =  9.28699   [0.7509393]
  Q( 20) =  11.1975   [0.8857877]
```

```
H0 : No serial correlation ==> Accept H0 when prob. is High [Q < Chisq(lag)]
--------------
ARCH 1-2 test:    F(2,3771)=  0.15601 [0.8556]
ARCH 1-5 test:    F(5,3765)=  0.62824 [0.6782]
ARCH 1-10 test:   F(10,3755)=  0.58855 [0.8247]
--------------

> v2=garchOxFit(formula.mean=~arma(0,2),formula.var=~igarch(1,1),
series=sbux,include.var=F)
 ********************
 ** SPECIFICATIONS **
 ********************
Mean Equation : ARMA (0, 2) model.
No regressor in the mean
Variance Equation : IGARCH (1, 1) model.
 No regressor in the variance
The distribution is a Gauss distribution.

 Maximum Likelihood Estimation (Std.Errors based on Second derivatives)
                  Coefficient  Std.Error  t-value  t-prob
Cst(M)               0.076850   0.033538    2.291  0.0220
MA(1)               -0.028510   0.017066   -1.671  0.0949
MA(2)               -0.044489   0.017175   -2.590  0.0096
ARCH(Alpha1)         0.022409  0.0030556    7.334  0.0000
GARCH(Beta1)         0.977791

Log Likelihood   : -8766.558

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

Horizon       Mean   Variance
      1     0.1169      3.779
      2    0.01172      ?????
      3    0.07685      ?????

***********
** TESTS **
***********
Information Criterium (to be minimized)
Akaike          4.642963  Shibata         4.642961
Schwarz         4.649567  Hannan-Quinn    4.645311
---------------
```

```
Q-Statistics on Standardized Residuals
  --> P-values adjusted by 2 degree(s) of freedom
  Q( 10) =  1.76955   [0.9872800]
  Q( 15) =  10.7899   [0.6284105]
  Q( 20) =  18.6587   [0.4131178]
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) =  7.61622   [0.4718271]
  Q( 15) =  10.7587   [0.6310203]
  Q( 20) =  12.5260   [0.8189285]
H0 : No serial correlation ==> Accept H0 when prob. is High [Q < Chisq(lag)]
--------------
ARCH 1-2 test:    F(2,3771)=  0.25569 [0.7744]
ARCH 1-5 test:    F(5,3765)=   1.0421 [0.3909]
ARCH 1-10 test:   F(10,3755)=  0.75004 [0.6775]
--------------

> v3=garchOxFit(formula.mean=~arma(0,2),formula.var=~gjr(1,1),
series=sbux,cond.dist="t")
 ********************
 ** SPECIFICATIONS **
 ********************
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 5.30675 degrees of freedom.

 Maximum Likelihood Estimation (Std.Errors based on Second derivatives)
                  Coefficient  Std.Error  t-value  t-prob
Cst(M)               0.031637   0.029810    1.061  0.2886
MA(1)               -0.043451   0.015757   -2.758  0.0059
MA(2)               -0.047573   0.015828   -3.006  0.0027
Cst(V)               0.015363  0.0084419    1.820  0.0689
ARCH(Alpha1)         0.020823  0.0060791    3.425  0.0006
GARCH(Beta1)         0.969739  0.0073203    132.5  0.0000
GJR(Gamma1)          0.017052  0.0084921    2.008  0.0447
Student(DF)          5.306752    0.44563    11.91  0.0000

Log Likelihood   : -8599.726
***************
** FORECASTS **
```

```
***************
Number of Forecasts: 15

Horizon       Mean   Variance
      1    0.05403       5.06
      2   -0.03868       5.07
      3    0.03164      5.081
     (edited)
***********
** TESTS **
***********
Information Criterium (to be minimized)
Akaike          4.556763  Shibata         4.556755
Schwarz         4.569970  Hannan-Quinn    4.561459
---------------
Q-Statistics on Standardized Residuals
  --> P-values adjusted by 2 degree(s) of freedom
  Q( 10) =  2.05833   [0.9791711]
  Q( 15) =  11.4473   [0.5733998]
  Q( 20) =  19.4943   [0.3619947]
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) =  5.32602   [0.7222298]
  Q( 15) =  8.43218   [0.8143503]
  Q( 20) =  10.6425   [0.9089071]
H0 : No serial correlation ==> Accept H0 when prob. is High [Q < Chisq(lag)]
--------------
ARCH 1-2 test:    F(2,3771)= 0.030543 [0.9699]
ARCH 1-5 test:    F(5,3765)=  0.46169 [0.8050]
ARCH 1-10 test:   F(10,3755)=  0.51667 [0.8796]

> mean(v3$condvars)
[1] 7.497493
> quantile(v3$residuals,0.025)
    2.5%
-5.10709
```