

Graduate School of Business, University of Chicago  
Business 41202, Spring Quarter 2008, Mr. Ruey S. Tsay

**Final Exam**

**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.
- There are 16 pages, including the R output.
- For simplicity, **ALL** tests use the 5% significance level, and all **VaR** calculations use 1% tail probability.
- Round your answer to 3 significant digits.

**Problem A:** (34 pts) Answer briefly the following questions.

1. Consider the monthly simple returns of Decile 1 and Decile 10 portfolios from January 1988 to December 2007 for 240 observations. Due to January effects, Decile 1 has some serial correlations at seasonal lags. To remove the effect, an ARIMA model is used. Write down the fitted model for the Decile 1 portfolio returns, including the variance of the innovations.
2. Again, consider the monthly simple returns of Decile 1 of the prior question. Are there any serial correlations in the residuals? Are there any ARCH-effects in the residuals? Why?

3. The residuals of the prior model for Decile 1 returns are combined with Decile 10 returns to form a bivariate time series. The multivariate Ljung-Box statistics are applied to detect any serial or cross-correlations between the two time series. Write down the null hypothesis for  $Q_2(10)$  statistic, where 10 denotes the first 10 lags of serial and cross correlations. What is the asymptotic distribution of  $Q_2(10)$ ? Is the test significant for the two series? Why?
  
4. Describe conditions under which nonsynchronous trading leads to the existence of serial correlations at all lags in the returns of an asset, even though the returns are in fact serially uncorrelated.
  
5. In addition to GARCH-type models, give two alternative approaches to modeling daily stock volatilities.
  
6. The CBOE VIX index is used as a measure of market volatility. We can use time series models to forecast the VIX index. Write down the fitted model for the  $\log(\text{VIX})$  series, including residual variance. Compute 1-step and 5-step ahead forecasts of the VIX index (not  $\log(\text{VIX})$ ).
  
7. Suppose that the price  $P_t$  of a stock follows the stochastic diffusion model
 
$$dP_t = \mu P_t dt + \sigma P_t^{0.5} dw_t,$$
 where  $\mu$  and  $\sigma$  are constant and  $w_t$  is the standard Brownian motion. What is the stochastic model for the log price  $G(P_t) = \ln(P_t)$ ?
  
8. Consider a nondividend-paying stock. If the current price of the stock is \$45.50 and the risk-free interest rate is 5% per annum. What is the minimum price of a European call option on the stock with time-to-expiration 3 months and strike price \$43.00?

9. Describe a consequence in linear regression analysis if the serial correlations of the residuals are overlooked.
  
10. Describe briefly two methods that can be used to model the stock price changes in tick-by-tick transaction data.
  
11. Suppose that the monthly log returns, in percentages, of an asset follow the model
 
$$r_t = 0.1 + 0.2r_{t-1} + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, 1), \quad \sigma_t^2 = 0.95\sigma_{t-1}^2 + 0.05a_{t-1}^2.$$

Suppose that at the forecast origin  $T$ ,  $r_T = 1.0$ ,  $r_{T-1} = -0.5$ , and  $\sigma_T^2 = 2.05$ . What is the 2-step ahead forecast  $r_T(2)$  of the return? What is the 2-step ahead volatility forecast  $\sigma_T(2)$ ?
  
12. Give one advantage and one disadvantage of using the empirical quantiles to calculate Value at Risk (VaR) of a financial position.
  
13. Consider the daily log returns, in percentages, of Caterpillar Stock from January 1993 to December 2007. The summary statistics are given. Is the distribution of the log returns symmetric against its mean? Perform a proper test to justify your answer.
  
14. Again, consider the daily log returns of Caterpillar stock of the prior question. Is the mean of the returns significantly different from zero? Why?
  
15. Realized volatility has been used as a measure of daily stock volatility. However, market microstructure may affect the quality of the measure. Give two approaches that can overcome the effects of microstructure noises on the realized volatility.

16. Consider two stocks. Let their log prices be  $p_{1t}$  and  $p_{2t}$ , respectively. Describe two conditions needed for implementing pairs trading of the two stocks to obtain a positive profit.
  
17. Consider a univariate AR(3) model, say  $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + a_t$ , where  $\{a_t\}$  is a sequence of independent and identically distributed random variables with mean zero and variance  $\sigma^2 > 0$ . Write down an explicit error-correction form for the model to perform unit-root test. What is the null hypothesis?

**Problem B.** (30 points) Consider the daily log returns, in percentages, of the Caterpillar and Adobe System Stocks from January 1993 to December 2007. The sample size is 3778 and the tick symbols are CAT and ADBE, respectively. Suppose that you hold a long position of the stocks valued at \$1 million dollars each. (Total is 2 millions.) Answer the following questions based on the attached output.

1. Focus on the CAT stock. A GARCH(1,1) model is fitted to the data. Write down the fitted model, including innovation distributions. Is the fitted model adequate? Why?
  
2. Use the fitted GARCH(1,1) model, calculate the VaR of the CAT position for the next trading day.
  
3. A special IGARCH(1,1) model is also fitted to the CAT returns. Use the information to calculate VaR of the CAT position via the RiskMetrics method. What is the corresponding VaR for the next 10 trading days?

4. Apply the traditional extreme value theory to the negative log returns with block size 63. What are the estimates of the three parameters  $k$ ,  $\alpha$ , and  $\beta$ ? Are these estimates statistically significant? Why?
5. Based on the EVT estimates of block size 63, what is the VaR of the CAT position for the next trading day?
6. Turn to the approach of peaks over the threshold. Using threshold 2.5, we estimate the parameters of generalized Pareto distribution for the negative CAT returns. Does the log returns have a heavy left tail? Why?
7. Based on estimated generalized Pareto distribution, what is the VaR of the CAT position for the next trading day? What is the VaR of the position for the next 10 trading days?
8. Again, based on the fitted generalized Pareto distribution with threshold 2.5, what is the expected shortfall when the 1% VaR is used?
9. Turn next to the ADBE stock. Again, a special IGARCH(1,1) model is fitted. Based on the information available, what is the VaR of the ADBE position for the next trading day.
10. The correlation between the CAT and ADBE stocks is 0.208. What is the VaR of your portfolio of the two stocks for the next trading day?

**Problem C.** (16 pts) Again, consider the daily log returns, in percentages, of the Adobe System stock from January 1993 to December 2007. Now, let us focus on predicting the direction of price movement. Define

$$Y_t = \begin{cases} 1 & \text{if } r_t > 0 \\ 0 & \text{otherwise} \end{cases}$$

where  $r_t$  is the log return of ADBE stock at time index  $t$ .

1. A 5-2-1 neural network is employed for  $Y_t$ . The five input variables are the first five lagged values of the returns, i.e.  $x_{1,t} = r_{t-1}, x_{2,t} = r_{t-2}, \dots, r_{t-5}$ . Write down the fitted model for the first node ( $h_1$ ) of the hidden layer.
2. Write down the fitted model for the output node.
3. Alternatively, a logistic linear regression can be used to model  $Y_t$  using the five input variables. Write down the fitted logistic linear regression model. Are the estimated coefficients significant? Why?
4. To simplify the model, we only keep the lag-1 variable ( $r_{t-1}$ ) as the independent variable. Write down the fitted logistic linear regression model. If  $r_T = -0.47$ , what is the probability that the price of the stock would increase at the time index  $T + 1$ ?

**Problem D.** (20 points) Consider the adjusted daily closing prices of BHP and RIO stocks from March 21, 2002 to May 30, 2008 as discussed in class. Let the log price be  $p_{1t}$  and  $p_{2t}$ , respectively. Based on the augmented Dickey-Fuller unitroot test, both series  $\{p_{1t}\}$  and  $\{p_{2t}\}$  have a unitroot.

1. Use  $p_{1t}$  as the dependent variable and  $p_{2t}$  as the independent variable. What is the linear combination that gives rise to a mean-reverting process?
  
2. Denote the mean-reverting linear combination by  $w_t$ . Unitroot test confirms that it is indeed stationary. The values of  $w_t$  for  $t$  from 330 to 340 are shown in the attached output. Consider the position: long one share of BHP stock and short  $\gamma$  shares of RIO stock. Let  $\delta = 0.03$ . Was there an trading opportunity within the time period (assume no problem with order execution)? If yes, what is the resulting profit?
  
3. Alternatively, use  $p_{2t}$  as the dependent variable and  $p_{1t}$  as the independent variable. What is the linear combination that gives rise to a mean-reverting process?
  
4. The values of the linear combination of Question 3 (denoted by  $w1$  in the output) are given for  $t$  from 605 to 630. Based on the values, construct a portfolio and chosen threshold  $\delta$  that will provide a maximum profit. What is the resulting profit?
  
5. Compare the two linear combinations in Question 1 and 3. The  $R^2$  measure of the two simple linear regressions is the same. Similarly, the two  $t$ -ratios for regression coefficients are the same. For the purpose of pairs trading, which linear combination do you prefer? Why?

Computer output.

```
*** Problem A ***
```

```
> x=read.table("m-dec1n10.txt")
```

```
> dim(x)
```

```
[1] 240  3
```

```
> dec1=x[,2]
```

```
> dec10=x[,3]
```

```
> mm=arima(dec1,order=c(0,0,1),seasonal=list(order=c(1,0,1),period=12))
```

```
> mm
```

```
arima(x = dec1, order = c(0, 0, 1), seasonal = list(order = c(1, 0, 1), period = 12))
```

```
Coefficients:
```

|      | ma1    | sar1   | sma1    | intercept |
|------|--------|--------|---------|-----------|
|      | 0.1927 | 0.9985 | -0.9757 | 0.0148    |
| s.e. | 0.0630 | 0.0061 | 0.0487  | 0.0092    |

```
sigma^2 estimated as 0.002966:  log likelihood = 350.82,  aic = -691.64
```

```
> Box.test(mm$residuals,24,type='Ljung')
```

```
Box-Ljung test
```

```
data: mm$residuals
```

```
X-squared = 15.1279, df = 24, p-value = 0.917
```

```
> Box.test(mm$residuals^2,24,type='Ljung')
```

```
Box-Ljung test
```

```
data: mm$residuals^2
```

```
X-squared = 10.4837, df = 24, p-value = 0.9922
```

```
> rr=cbind(mm$residuals,dec10)
```

```
> source("mqstat.txt")
```

```
> mq(rr,10)
```

```
[1] "m,      Q(m)  p-value:"
```

```
[1] 1.0000  9.1862  0.0566
```

```
[1]  2.000 11.521  0.174
```

```
[1]  3.000 14.281  0.283
```

```
.....
```

```
[1] 10.000 32.420  0.797
```

```
**** VIX ****
```

```
> x=read.table("vix08.txt",header=T)
```

```
> vix=log(x[,7])
```

```
> pacf(diff(vix))
```

```

> mm=arima(vix,order=c(0,1,2))
> tsdiag(mm) % Shows some serial correlations in the residuals.
> acf(mm$residuals)

> mm=arima(vix,order=c(0,1,2),seasonal=list(order=c(0,0,1),period=8))
> tsdiag(mm,gof.lag=15) % The model is adequate.
> mm
Call:
arima(x = vix, order = c(0, 1, 2), seasonal = list(order = c(0, 0, 1), period = 8))

Coefficients:
          ma1          ma2          sma1
      -0.1611   -0.0932   -0.1045
s.e.   0.0305    0.0328    0.0305

sigma^2 estimated as 0.003732:  log likelihood = 1485.1,  aic = -2962.2
> predict(mm,5)
$pred
Time Series:
Start = 1081
End = 1085
Frequency = 1
[1] 3.029399 3.030089 3.034245 3.028598 3.026774

*** CAT stock returns ***
> x=read.table("d-cat9307.txt")
> cat=log(x[,2]+1)*100
> library(fBasics)
> basicStats(cat)

          cat
nobs      3778.000000
NAs         0.000000
Minimum    -15.685851
Maximum     10.295994
Mean         0.071379
Median       0.005050
SE Mean     0.032052
LCL Mean    0.008538
UCL Mean    0.134220
Variance    3.881240
Stdev       1.970086
Skewness    -0.185373
Kurtosis    3.808928

```

```

**** Problem B ****
> x=read.table("d-cat9307.txt")
> cat=log(x[,2]+1)*100
> m1=garch0xFit(formula.mean=~arma(0,0),formula.var=~garch(1,1),series=cat,
cond.dist="t")
*****
** SPECIFICATIONS **
*****
Dependent variable : X
Mean Equation : ARMA (0, 0) model.
No regressor in the mean
Variance Equation : GARCH (1, 1) model.
No regressor in the variance
The distribution is a Student distribution, with 5.90009 degrees of freedom.

```

Maximum Likelihood Estimation (Std.Errors based on Second derivatives)

|              | Coefficient | Std.Error | t-value | t-prob |
|--------------|-------------|-----------|---------|--------|
| Cst(M)       | 0.082924    | 0.027293  | 3.038   | 0.0024 |
| Cst(V)       | 0.017036    | 0.0084250 | 2.022   | 0.0432 |
| ARCH(Alpha1) | 0.021360    | 0.0048529 | 4.402   | 0.0000 |
| GARCH(Beta1) | 0.974316    | 0.0061420 | 158.6   | 0.0000 |
| Student(DF)  | 5.900092    | 0.53214   | 11.09   | 0.0000 |

```

No. Observations :      3778  No. Parameters   :          5
Mean (Y)          :    0.07138  Variance (Y)    :    3.88021
Skewness (Y)     :   -0.18545  Kurtosis (Y)   :    6.81253
Log Likelihood   : -7648.597  Alpha[1]+Beta[1]: 0.99568

```

\*\* TESTS \*\*

-----  
Information Criterium (to be minimized)

```

Akaike           4.051666  Shibata           4.051662
Schwarz          4.059920  Hannan-Quinn     4.054600

```

-----  
Q-Statistics on Standardized Residuals

```

Q( 10) =  8.60388  [0.5700636]
Q( 15) = 12.3913  [0.6492001]
Q( 20) = 16.5084  [0.6846259]

```

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) =  3.96433  [0.8603270]
Q( 15) =  5.84873  [0.9514853]

```

Q( 20) = 7.96943 [0.9790882]  
H0 : No serial correlation ==> Accept H0 when prob. is High [Q < Chisq(lag)]

```
> m1$condvars[3778]
[1] 2.8458
> m1$residuals[3778]
[1] -0.90641
>
> m2=garch0xFit(formula.mean=~arma(0,0),formula.var=~igarch(1,1),series=cat,
include.mean=F,include.var=F)
*****
** SPECIFICATIONS **
*****
Dependent variable : X
Mean Equation : ARMA (0, 0) 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 |
|--------------|-------------|-----------|---------|--------|
| ARCH(Alpha1) | 0.012178    | 0.0019016 | 6.404   | 0.0000 |
| GARCH(Beta1) | 0.988022    |           |         |        |

```
> source("r-garch11v.txt")
> m3=garch11v(cat,0,0,0.012,0.988)
> names(m3)
[1] "condvars" "atsq"
> m3$condvars[3778]
[1] 2.572129
> m3$atsq[3778]
[1] 0.6781215
>
> ncat=-cat
> library(evir)
> m4=gev(ncat,63)
> m4
$n.all
[1] 3778
$n
[1] 60
$data
[1] 2.715128 2.000069 2.402019 2.444434 3.261720 4.602614 3.444653
```

```

(edited)
[57] 3.679677 3.105526 4.443058 5.410356
$block
[1] 63

$par.ests
      xi      sigma      mu
0.2473358 1.5297742 3.9135484

$par.ses
      xi      sigma      mu
0.1191271 0.1866184 0.2279482

> m5=gpd(ncat,2.5)
> m5
$n
[1] 3778
$data
 [1] 2.715128 2.655036 2.766312 2.558246 3.261720 2.779472 4.602614
(edited)
[267] 3.096965 5.410356 2.828528 3.806226 3.085722
$threshold
[1] 2.5
$p.less.thresh
[1] 0.928269
$n.exceed
[1] 271

$par.ests
      xi      beta
0.1537228 1.1388052

$par.ses
      xi      beta
0.06507928 0.10081661

> riskmeasures(m5,c(0.99,0.999))
      p quantile      sfall
[1,] 0.990 5.120717 6.942424
[2,] 0.999 9.379966 11.975349

> y=read.table("d-adbe9307.txt")
> adbe=log(y[,2]+1)*100
> g1=garch0xFit(formula.mean=~arma(0,0),formula.var=~igarch(1,1),series=adbe,

```

```
include.mean=F,include.var=F)
```

```
*****
```

```
** SPECIFICATIONS **
```

```
*****
```

```
Dependent variable : X
```

```
Mean Equation : ARMA (0, 0) 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 |
|--------------|-------------|-----------|---------|--------|
| ARCH(Alpha1) | 0.022568    | 0.0032825 | 6.875   | 0.0000 |
| GARCH(Beta1) | 0.977632    |           |         |        |

```
No. Observations :      3778  No. Parameters   :          2
```

```
Mean (Y)          :    0.08302  Variance (Y)     :   12.07482
```

```
Skewness (Y)     :   -0.27535  Kurtosis (Y)    :   11.85749
```

```
Log Likelihood   : -9684.285
```

```
> g2=garch11v(adbe,0,0,0.0226,0.9776)
```

```
> g2$condvars[3778]
```

```
[1] 2.771387
```

```
> g2$atsq[3778]
```

```
[1] 0.2180784
```

```
>
```

```
> cor(cat,adbe)
```

```
[1] 0.2081141
```

```
*** Problem C ***
```

```
> y=adbe[6:3778]
```

```
> xx=cbind(adbe[5:3777],adbe[4:3776],adbe[3:3775],adbe[2:3774],adbe[1:3773])
```

```
> dy=ifelse(y>0,1,0)
```

```
> library(nnet)
```

```
> n1=nnet(xx,dy,size=2,skip=T,linout=F,maxit=3000)
```

```
# weights: 20
```

```
initial value 1360.069317
```

```
....
```

```
final value 926.854281
```

```
converged
```

```
> summary(n1)
```

```
a 5-2-1 network with 20 weights
```

```
options were - skip-layer connections
```

|         |         |         |         |          |         |       |       |
|---------|---------|---------|---------|----------|---------|-------|-------|
| b->h1   | i1->h1  | i2->h1  | i3->h1  | i4->h1   | i5->h1  |       |       |
| 2312.27 | -443.21 | -635.03 | 990.14  | -1152.05 | -272.51 |       |       |
| b->h2   | i1->h2  | i2->h2  | i3->h2  | i4->h2   | i5->h2  |       |       |
| 518.57  | 1029.97 | 98.24   | -172.38 | -356.23  | -737.76 |       |       |
| b->o    | h1->o   | h2->o   | i1->o   | i2->o    | i3->o   | i4->o | i5->o |
| -0.68   | 0.53    | 0.47    | -0.05   | 0.00     | -0.03   | 0.05  | 0.03  |

```
>
> n3=glm(dy~xx[,1]+xx[,2]+xx[,3]+xx[,4]+xx[,5],family=binomial)
```

```
> summary(n3)
Call:
glm(formula = dy ~ xx[, 1] + xx[, 2] + xx[, 3] + xx[, 4] + xx[,
5], family = binomial)
```

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.058773    0.032675  -1.799  0.0721 .
xx[, 1]      -0.023368    0.009497  -2.461  0.0139 *
xx[, 2]      -0.011216    0.009436  -1.189  0.2346
xx[, 3]      -0.015253    0.009459  -1.613  0.1068
xx[, 4]       0.006230    0.009433   0.660  0.5090
xx[, 5]      -0.010918    0.009438  -1.157  0.2474
```

```
---
Null deviance: 5226.7 on 3772 degrees of freedom
Residual deviance: 5214.9 on 3767 degrees of freedom
AIC: 5226.9
```

```
Number of Fisher Scoring iterations: 3
```

```
> n4=glm(dy~xx[,1],family=binomial)
> summary(n4)
Call:
glm(formula = dy ~ xx[, 1], family = binomial)
```

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.061280    0.032610  -1.879  0.0602 .
xx[, 1]      -0.023255    0.009457  -2.459  0.0139 *
---
```

```
Null deviance: 5226.7 on 3772 degrees of freedom
Residual deviance: 5220.6 on 3771 degrees of freedom
AIC: 5224.6
```

```
**** Pairs trading ***
```

```

> x=read.table("d-bhp0208r.txt")
> bhp=log(x[,9])
> y=read.table("d-rio0208r.txt")
> rio=log(y[,9])
> library(FinTS)

> adfTest(rio,type=c("c"))
Title:
  Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
  Lag Order: 1
STATISTIC:
  Dickey-Fuller: 0.1375
P VALUE:
  0.9667

> adfTest(bhp,type=c("c"))
Title:
  Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
  Lag Order: 1
STATISTIC:
  Dickey-Fuller: -0.2018
P VALUE:
  0.9317
>
> m3=lm(bhp~rio)
> summary(m3)
lm(formula = bhp ~ rio)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.816399   0.004241   428.3  <2e-16 ***
rio           0.718123   0.001961   366.3  <2e-16 ***
---
Residual standard error: 0.07404 on 1558 degrees of freedom
Multiple R-squared:  0.9885,    Adjusted R-squared:  0.9885

> w=m3$residuals
> adfTest(w,lags=2)

```

Title:  
Augmented Dickey-Fuller Test

Test Results:

PARAMETER:

Lag Order: 2

STATISTIC:

Dickey-Fuller: -4.0677

P VALUE:

0.01

```
> print(w[330:340],digits=3)
      330      331      332      333      334      335      336      337
-0.01431 -0.00216 -0.04059 -0.01960 -0.01329 -0.00258 -0.03232  0.00450
      338      339      340
 0.00759  0.01961  0.04060
```

```
> m4=lm(rio~bhp)
> summary(m4)
lm(formula = rio ~ bhp)
```

Coefficients:

|             | Estimate  | Std. Error | t value | Pr(> t )   |
|-------------|-----------|------------|---------|------------|
| (Intercept) | -2.478055 | 0.012339   | -200.8  | <2e-16 *** |
| bhp         | 1.376532  | 0.003758   | 366.3   | <2e-16 *** |

---

Residual standard error: 0.1025 on 1558 degrees of freedom  
Multiple R-squared: 0.9885, Adjusted R-squared: 0.9885

```
> w1=m4$residuals
> print(w1[605:630],digits=3)
      605      606      607      608      609      610      611      612
-0.06428 -0.06946 -0.03435 -0.01126 -0.01767 -0.03348 -0.05349 -0.03299
      613      614      615      616      617      618      619      620
-0.00644 -0.01654 -0.01426  0.01946  0.01425  0.00266 -0.00975  0.01809
      621      622      623      624      625      626      627      628
 0.05738  0.05534  0.05879  0.04013  0.04380  0.05841  0.08031  0.03897
      629      630
 0.02378  0.02515
```