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

Signature: Name: ID:

Notes:
- This is a 3-hour, open-book, and open-note exam.
- Write your answer in the blank space provided for each question.
- There are 18 pages, including some R output.
- For simplicity, **ALL** tests use the 5% significance level, and all risk-measure calculations use 1% tail probability. Furthermore, unless specified, all VaR and expected shortfall are for the next trading day.
- Round your answer to 3 significant digits.
- You may bring a PC or calculator to the exam, but **no Internet or email or phone is allowed during the exam!**

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

1. Describe the two key assumptions used by the RiskMetrics to justify the square root of time rule in VaR calculation.

2. Suppose that $X_t$ follows a 2-dimensional stationary vector AR(1) model,

\[
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix} = \begin{bmatrix}
    1.5 \\
    0.5
\end{bmatrix} + \begin{bmatrix}
    0.2 & 0.0 \\
    0.6 & 0.8
\end{bmatrix} \begin{bmatrix}
    x_{1,t-1} \\
    x_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
    a_{1t} \\
    a_{2t}
\end{bmatrix},
\]

where $a_t$ is a sequence of independent and identically distributed Gaussian random vectors with mean zero and covariance matrix $\Sigma > 0$. Does the model imply the existence of Granger causality? Why? If there is a Granger causality, state the relation.
3. Describe two applications of seasonal time-series models in finance.

4. Suppose the price $P_t$ of a stock follows the stochastic diffusion equation (SDE) $dP_t = P_t dt + 2P_t dw_t$, where $w_t$ denotes the standard Brownian motion. What is the stochastic diffusion equation for the square-root process $G_t = \sqrt{P_t}$? Justify briefly your answer.

5. Provide two volatility models with leverage effect for the daily log return $r_t$ of a stock.

6. Provide two potential sources of serial correlations in observed log returns of an asset even though the true log returns are serially uncorrelated.

7. The methods of empirical quantile and quantile regression can be used to estimate VaR. Give two reasons that the quantile regression approach is preferred.

8. Describe two assumptions used in pairs trading in the equity market.

9. Consider an AR(2) time series $(1 - 0.805B - 0.1215B^2)x_t = 0.134 + a_t$, where $B$ is the backshift operator and $\{a_t\}$ are iid $N(0, 1)$ variables. Is $x_t$ stationary? Why? If it is stationary, calculate $E(x_t)$, the mean of $x_t$.

10. Describe two methods that can mitigate the effects of market micro-structure noises in calculating daily realized volatility of a stock.
11. Suppose that the true (unobserved) log price $p^*_t$ of an asset follows the model $p^*_t = p^*_{t-1} + a_t$, where $\{a_t\}$ is sequence of iid $N(0, 1)$ random variables. Suppose that the observed log price $p_t$ follows the model $p_t = p^*_t + \epsilon_t$, where $\{\epsilon_t\}$ is another sequence of iid $N(0, 0.25)$ random variables and $\{\epsilon_t\}$ and $\{a_t\}$ are independent. Let $r_t = p_t - p_{t-1}$ be the observed log return. Calculate the lag-1 and lag-2 autocorrelations of $r_t$.

12. Let $C_i$ be the price change of the $i$th trade for a stock in a given day. For a heavily traded stock, describe two empirical characteristics of $C_i$.

13. Let $r_t$ and $r_{m,t}$ be the monthly log returns of an asset and the market index, respectively. Suppose we like to extend the well-known Market model $r_t = \alpha + \beta r_{m,t} + \epsilon_t$ so that the intercept $\alpha$ or the slope $\beta$ may depend on the performance of the market. Describe one extension that allows for the intercept, but not the slope, to change depending on the sign of the market. Also, describe another model that allows for the slope, but not the intercept, to change depending on the sign of the market.

14. (For Questions 14 to 18) Consider the U.S. monthly growth rate of consumer price index, seasonally adjusted, from February 1947 to April 2013 for 795 observations. Analysis of the series is provided in the attached R output. Use the output to answer Questions 14 to 18. Write down the fitted pure time series model for the series $x_t$, including the residual variance. (See model m1.)

15. Based on the provided model checking statistics, is there any serial correlation in the residuals? Is there any ARCH effect in the residuals? Justify your answer briefly.
16. A Gaussian GARCH(1,1) model is entertained, write down the fitted model for the residuals.

17. A GARCH(1,1) model with Student-$t$ innovations is also entertained. Let $v$ be the degrees of freedom of the Student-$t$ innovation. Consider the hypothesis $H_0 : v = 5$ versus $H_a : v \neq 5$. Calculate the test statistic and draw your conclusion.

18. One may combine the time series model with a volatility model to describe the dynamics of U.S. monthly CPI growth rates. Which volatility model is preferred? Why?

19. Give two types of nonlinear time series models discussed in the lectures that are useful in finance.

20. Give two models that can be used to model price changes in high-frequency financial transactions.
Problem B. (26 points) Consider the daily adjusted close prices of Apple (AAPL) and McDonald’s (MCD) stocks, starting from some day in January 2002 for 2850 observations. We compute the daily log returns from the price series, resulting in 2849 returns. Consider a long position of $1 million on Apple stock and $1 million on McDonald’s stock. Use the attached output to answer the following questions.

1. (4 points) If RiskMetrics is used, calculate the VaR for Apple and McDonald’s positions separately.

2. (3 points) A Gaussian GARCH(1,1) models is applied to the Apple log returns. Based on the model, compute 1-day and 5-day VaR of the Apple position.

3. (3 points) A GARCH(1,1) model with Student-$t$ innovations is applied to the Apple stock returns. Based on the model, what are the VaR and expected shortfall for the Apple position?

4. (4 points) If the traditional method of block maximum is used to study the extreme-value behavior of AAPL log returns, what are the parameter estimates? Are the estimates statistically significant? Why?

5. (2 points) What are the 1-day and 10-day VaRs for the Apple position based on the traditional block-maximum method?
6. (2 points) What are the VaR and expected shortfall for the Apple position based on the POT (peaks over threshold) approach?

7. (4 points) Consider the approach based on generalized Pareto distribution. The threshold of 3% is used for the AAPL log returns. Write down the parameter estimates for the stock? Are these estimates significantly different from zero? Why?

8. (4 points) Consider, again, the RiskMetrics method. If constant correlation is used, what is the VaR of the combined position? If time-varying correlation is used, what is the VaR of the combined position?

Problem C. (18 points). Consider the intraday trading of the Caterpillar (CAT) stock on December 5, 2012 with 44,334 trades during the regular trading hours. Let $C_i$ be the price change of the $i$th trade. Define

$$A_i = \begin{cases} 1 & \text{if } C_i \neq 0 \\ 0 & \text{if } C_i = 0 \end{cases} \quad D_i = \begin{cases} 1 & \text{if } C_i > 0 \\ -1 & \text{if } C_i < 0 \end{cases}.$$

In addition, we classify the size of the price change into seven categories, namely $S_i = 1, 2, \ldots, 7$ for $(-\infty, -0.02), [-0.02, -0.01), [-0.01, 0), [0, 0.01), [0.01, 0.02), (0.02, \infty)$, respectively. Logistic regressions are used to model $A_i$ and $D_i$. The output is attached. Answer the following questions:

1. (2 points) Write down the fitted model $md1$ for $A_i$. Are the estimates significantly different from zero? Why?
2. (2 points) Based on the fitted model, calculate the probabilities \(P(A_i = 1|A_{i-1} = 1)\) and \(P(A_i = 1|A_{i-1} = 0)\).

3. (4 points) Write down the fitted model for \(D_i\) conditional on \(A_i = 1\) when both \(D_{i-1}\) and \(C_{i-1}\) are used. [That is, \text{md2}.] Are the estimates significantly different from zero? Why?

4. (3 points) Based on the fitted model \(\text{md2}\), compute the probabilities \(P(D_i = 1|A_i = 1, D_{i-1} = -1, C_{i-1} = -0.02)\) and \(P(D_i = 1|A_i = 1, D_{i-1} = 1, C_{i-1} = 0.01)\).

5. (3 points) Another model is employed for \(D_i\) conditional on \(A_i = 1\) and using \(D_{i-1}\) and \(S_{i-1}\) as explanatory variables. [That is, \text{md3}.] Based on the model, compute \(P(D_i = 1|A_i = 1, D_{i-1} = -1, S_{i-1} = 2)\) and \(P(D_i = 1|A_i = 1, D_{i-1} = 1, S_{i-1} = 5)\).

6. (2 points) Consider the probabilities in questions 4 and 5. Since \(C_{i-1} = -0.02\) corresponds to \(S_{i-1} = 2\), why the two models give different probabilities?

7. (2 points) Compare the two models \(\text{md2}\) and \(\text{md3}\). Which model is preferred? Why?
**Problem D.** (16 points). Consider two financial assets. The first asset is the Vanguard energy ETF (VDE) and the second is energy select sector SPDR (XLE) of State Street Global Advisors. The sample period is from July 1, 2009 to June 30, 2011 for 505 observations. Denote the log price (namely, daily adjusted close price) of the asset by $v_{de_t}$ and $x_{le_t}$, respectively. Some analysis of the data is given. Answer the following questions:

1. (2 points) Are the two log price series co-integrated? Why?

2. (2 points) Let $w_t = v_{de_t} - x_{le_t}$. Write down the fitted model for $w_t$.

3. (2 points) What is the approximate half-life for the $w_t$ series?

4. (2 points) The time plot of $w_t$ is shown in Figure 1. The horizontal line denotes the sample mean of $w_t$. From the plot and the information provided, is there any pairs trading opportunities? Why?

5. (2 points) Suppose that the transaction cost is 2%. Are there sufficient opportunities to conduct pairs trading? Why?

6. (2 points) Suppose that the transaction cost is reduced to 1%. Are there sufficient opportunities to conduct pairs trading? Why?
7. (2 points) We extend the sample period so that the data span is from July 1, 2009 to December 31, 2011. Again, let \( w_t = vde_t - xle_t \). Write down the model for \( w_t \) and calculate the approximate half-life of this new \( w_t \) series.

8. (2 points) Compare the two models for \( w_t \) and discuss the implication, if any, on pairs trading.
### R Analysis ####

```r
> da=read.table("m-cpiaucsl.txt",header=T)
> head(da)
   year mon day   cpi   
 1 1947   1   1 21.48
 6 1947   6   1 22.08
> cpi=da$cpi ### CPI level
> xt=diff(log(cpi)) #### CPI growth rate (inflation rate)
> plot(xt,type='l'); acf(xt); pacf(xt)
> m1=arima(xt,order=c(3,0,1),seasonal=list(order=c(1,0,1),period=12))
> m1
Call:arima(x=xt,order=c(3,0,1), seasonal=list(order = c(1, 0, 1), period = 12))
Coefficients:
(Intercept) ar1 ar2 ar3 ma1 sar1 sma1
1.2427  -0.3222  0.0639  0.1539  0.0029 0.0007
s.e.  0.0498  0.0567  0.0409  0.0363 0.1068 0.0968
sigma^2 estimated as 7.257e-06: log likelihood = 3575.03, aic = -7134.06
> resi=m1$residuals
> Box.test(resi,lag=12,type="Ljung")
Box-Ljung test
   data: resi
X-squared = 14.8517, df = 12, p-value = 0.2496
> Box.test(resi^2,lag=12,type="Ljung")
Box-Ljung test
   data: resi^2
X-squared = 285.9342, df = 12, p-value < 2.2e-16
> m2=garchFit(~garch(1,1),include.mean=F,data=resi,trace=F)
> summary(m2)
Title: GARCH Modelling
Call: garchFit(formula = ~garch(1, 1), data = resi, include.mean = F, trace = F)
            Estimate Std. Error   t value   Pr(>|t|)
omega 3.397e-07  9.525e-08  3.567      0.000362 ***
alpha1 2.159e-01  3.608e-02  5.983      2.2e-09 ***
beta1 7.492e-01  3.474e-02 21.565 < 2e-16 ***
```

---
Standardised Residuals Tests:

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jarque-Bera Test R Chi²</td>
<td>146.8382</td>
<td>0</td>
</tr>
<tr>
<td>Shapiro-Wilk Test R W</td>
<td>0.9762105</td>
<td>4.358221e-10</td>
</tr>
<tr>
<td>Ljung-Box Test R Q(10)</td>
<td>15.43711</td>
<td>0.1169196</td>
</tr>
<tr>
<td>Ljung-Box Test R² Q(10)</td>
<td>7.948971</td>
<td>0.6338215</td>
</tr>
</tbody>
</table>

Information Criterion Statistics:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>AIC</th>
<th>BIC</th>
<th>SIC</th>
<th>HQIC</th>
</tr>
</thead>
</table>

```R
> m3=garchFit(~garch(1,1),include.mean=F,data=resi,trace=F,cond.dist="std")
> summary(m3)
Title: GARCH Modelling
Call:
garchFit(formula ="garch(1,1),data=resi, cond.dist="std",include.mean=F, trace=F)

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

Conditional Distribution: std
Error Analysis:

| Parameter | Estimate | Std. Error | t value | Pr(>|t|) |
|-----------|----------|------------|---------|----------|
| omega     | 3.793e-07| 1.327e-07  | 2.858   | 0.00426  ** |
| alpha1    | 2.238e-01| 5.157e-02  | 4.340   | 1.43e-05 *** |
| beta1     | 7.402e-01| 4.856e-02  | 15.242  | < 2e-16 *** |
| shape     | 5.252e+00| 9.604e-01  | 5.468   | 4.55e-08 *** |

Standardised Residuals Tests:

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ljung-Box Test R Q(10)</td>
<td>15.22552</td>
<td>0.1240537</td>
</tr>
<tr>
<td>Ljung-Box Test R² Q(10)</td>
<td>7.833404</td>
<td>0.6451059</td>
</tr>
</tbody>
</table>

Information Criterion Statistics:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>AIC</th>
<th>BIC</th>
<th>SIC</th>
<th>HQIC</th>
</tr>
</thead>
</table>
```

##### Problem B ##############
```R
> getSymbols("AAPL",from="XXXX",to="XXXX")
[1] "AAPL"
> getSymbols("MCD",from="XXXX",to="XXXX")
[1] "MCD"
> aapl=diff(log(as.numeric(AAPL$AAPL.Adjusted)))
> mcd=diff(log(as.numeric(MCD$MCD.Adjusted)))

```

####### Apple Stock ############## (MCD on page 14-15) ###
```R
> xt=-aapl
```

11
> m1=lgarch(xt)
Estimates: 0.9655594
Maximized log-likehood: -6676.829
Coefficient(s):

| Estimate | Std. Error | t value | Pr(>|t|)       |
|----------|------------|---------|---------------|
| beta 0.96555937 | 0.00429533 | 224.793 | < 2.22e-16 ***|

---

> xt[2849]
[1] -0.02901629
> m1$volatility[2849]
[1] 0.02089391

> m2=garchFit(~garch(1,1),data=xt,trace=F)
> summary(m2)

Title: GARCH Modelling
Call: garchFit(formula = ~garch(1, 1), data = xt, trace = F)

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

Conditional Distribution: norm
Error Analysis:

| Estimate | Std. Error | t value | Pr(>|t|)       |
|----------|------------|---------|---------------|
| mu -2.084e-03 | 4.027e-04 | -5.173 | 2.30e-07 ***   |
| omega 8.001e-06 | 2.543e-06 | 3.146 | 0.00165 **     |
| alpha1 5.425e-02 | 8.510e-03 | 6.375 | 1.83e-10 ***   |
| beta1 9.336e-01 | 1.080e-02 | 86.474 | < 2e-16 ***    |

---

Standardised Residuals Tests:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jarque-Bera Test R Chi^2</td>
<td>1080.226</td>
</tr>
<tr>
<td>Shapiro-Wilk Test R W</td>
<td>0.9734941</td>
</tr>
<tr>
<td>Ljung-Box Test R Q(10)</td>
<td>14.90746</td>
</tr>
<tr>
<td>Ljung-Box Test R^2 Q(10)</td>
<td>4.866153</td>
</tr>
</tbody>
</table>

Information Criterion Statistics:

<table>
<thead>
<tr>
<th>AIC</th>
<th>BIC</th>
<th>SIC</th>
<th>HQIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.704425</td>
<td>-4.696064</td>
<td>-4.704429</td>
<td>-4.701410</td>
</tr>
</tbody>
</table>

> predict(m2,5)

<table>
<thead>
<tr>
<th>meanForecast</th>
<th>meanError</th>
<th>standardDeviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 -0.002083516</td>
<td>0.02198204</td>
<td>0.02198204</td>
</tr>
<tr>
<td>2 -0.002083516</td>
<td>0.02203044</td>
<td>0.02203044</td>
</tr>
<tr>
<td>3 -0.002083516</td>
<td>0.02207816</td>
<td>0.02207816</td>
</tr>
<tr>
<td>4 -0.002083516</td>
<td>0.02212519</td>
<td>0.02212519</td>
</tr>
<tr>
<td>5 -0.002083516</td>
<td>0.02217155</td>
<td>0.02217155</td>
</tr>
</tbody>
</table>
m3 = garchFit(~garch(1,1), data = xt, trace = F, cond.dist = "std")
summary(m3)

Title: GARCH Modelling
Call: garchFit(formula = ~garch(1,1), data = xt, cond.dist = "std", trace = F)

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

Conditional Distribution: std

Error Analysis:

| Variable | Estimate  | Std. Error | t value | Pr(>|t|)  |
|----------|-----------|------------|---------|-----------|
| mu       | -1.711e-03| 3.698e-04  | -4.627  | 3.71e-06  *** |
| omega    | 6.235e-06 | 2.499e-06  | 2.495   | 0.0126    *  |
| alpha1   | 4.833e-02 | 1.027e-02  | 4.707   | 2.51e-06  *** |
| beta1    | 9.421e-01 | 1.227e-02  | 76.812  | < 2e-16   *** |
| shape    | 5.483e+00 | 5.379e-01  | 10.192  | < 2e-16   *** |

---

Standardised Residuals Tests:

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ljung-Box Test R</td>
<td>14.9856</td>
<td>0.1325876</td>
</tr>
<tr>
<td>Ljung-Box Test R^2</td>
<td>5.575123</td>
<td>0.849608</td>
</tr>
</tbody>
</table>

Information Criterion Statistics:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>-4.780660</td>
</tr>
<tr>
<td>BIC</td>
<td>-4.770210</td>
</tr>
<tr>
<td>SIC</td>
<td>-4.780667</td>
</tr>
<tr>
<td>HQIC</td>
<td>-4.776892</td>
</tr>
</tbody>
</table>

> predict(m3, 1)

meanForecast  meanError  standardDeviation
1  -0.001711227  0.02180995  0.02180995

#

> m4 = gev(xt, block = 21)

$n.all
[1] 2849
$n
[1] 136
...
$block
[1] 21
$par.est$sigma     0.14973960 0.01535382 0.03200342
$pars.$par.ses     0.057148183 0.001080040 0.001457007

13
> m5=pot(xt,thres=0.03)
> m5
$n
[1] 2849
$period
[1]  1 2849
....
$span
[1] 2848
$threshold
[1] 0.03
$p.less.thresh
[1] 0.9248859
$n.exceed
[1] 214
$par.est
  xi  sigma    mu     beta
0.113680697 0.011370103 -0.004217072 0.015259923
$par.ses
  xi  sigma    mu
0.048694978 0.001860720 0.003777981
> riskmeasures(m5,c(0.99))
   p  quantile  sfall
[1,] 0.99  0.06458284 0.08623568

> m6=gpd(xt,thres=0.03)
> m6
$n
[1] 2849
...
$threshold
[1] 0.03
$p.less.thresh
[1] 0.9248859
$n.exceed
[1] 214
$par.est
  xi    beta
0.11382866 0.01525516
$par.ses
  xi    beta
0.064025413 0.001407045
> riskmeasures(m6,c(0.99))
   p  quantile  sfall
[1,] 0.99  0.06457741 0.08623568

################################################################# McDonald's stock ##############
> yt=-mcd
> n1=Igarch(yt)
Estimates:  0.9594923
Maximized log-likehood:  -8285.379
Coefficient(s):

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| beta 0.95949234 | 0.00402332 | 238.483 | < 2.22e-16 *** |

---

> mcd[2849]
[1] -0.0003945941
> n1$volatility[2849]
[1] 0.008514932

> cor(xt,yt)  ####### Correlation
[1] 0.2970319
>
####
> S=xt+yt
> D=xt-yt
> n1=Igarch(S)
Estimates:  0.953611
Maximized log-likehood:  -5942.07
Coefficient(s):

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| beta 0.95361096 | 0.00548611 | 173.823 | < 2.22e-16 *** |

---

> v1=(1-0.95361)*S[2849]^2+0.95361*n1$volatility[2849]^2
> v1
[1] 0.0005940488

> n2=Igarch(D)
Estimates:  0.9798481
Maximized log-likehood:  -6675.77
Coefficient(s):

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| beta 0.97984806 | 0.00274522 | 356.929 | < 2.22e-16 *** |

---

> v2=(1-0.97985)*D[2849]^2+0.97985*n2$volatility[2849]^2
> v2
[1] 0.0004558032
>
############ Problem C ############
> da=read.csv("taq-cat-t-dec052012.csv",header=T)
> dim(da)
[1] 44633 6
> m1=hfchg(da)  ### Process price change series

15
> names(m1)
[1] "pchange" "duration" "size"
> m2=pchcls(m1$pchange) #### Classify price changes into categories.
> dim(m2)
[1] 44332 3
> nT=dim(m2)[1]
> head(m2)
     Ai  Di  Si
[1,] 0   0  4
.....
[6,] 0   0  4
> Ai=m2[2:nT,1]; Aim1=m2[1:(nT-1),1] ### define variables
> Di=m2[2:nT,2]; Dim1=m2[1:(nT-1),2]
> Si=m2[2:nT,3]; Sim1=m2[1:(nT-1),3]
> md1=glm(Ai~Aim1,family=binomial)
> summary(md1)
Call: glm(formula = Ai ~ Aim1, family = binomial)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.0236   0.01360  -75.28  <2e-16 *** 
Aim1       1.2434   0.02074   59.96  <2e-16 ***
---
> di=Di[Ai==1]; dim1=Dim1[Ai==1]
> di=(di+abs(di))/2
> Ci=m1$pchange[2:nT]
> Cim1=m1$pchange[1:(nT-1)]
> cim1=Cim1[Ai==1]
> md2=glm(di~dim1+cim1,family=binomial)
> summary(md2)
Call: glm(formula = di ~ dim1 + cim1, family = binomial)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.0016   0.01645  -0.100   0.921
dim1       -0.7886   0.03434  -22.965  < 2e-16 ***
cim1     -13.2597   2.56560  -5.168  2.36e-07 ***
---
Null deviance: 22892 on 16512 degrees of freedom
Residual deviance: 21097 on 16510 degrees of freedom; AIC: 21103
> sim1=Sim1[Ai==1]
> md3=glm(di~dim1+sim1,family=binomial)
> summary(md3)
Call: glm(formula = di ~ dim1 + sim1, family = binomial)
Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.39272  0.15743  2.495 0.0126 *
dim1       -0.79592  0.05560 -14.316 <2e-16 ***
sim1        -0.09853  0.03916  -2.516 0.0119 *
---
 Null deviance: 22892 on 16512 degrees of freedom
 Residual deviance: 21121 on 16510 degrees of freedom; AIC: 21127

########## Problem D ###############################
> getSymbols("VDE", from="2009-07-01", to="2011-6-30")
[1] "VDE"
> getSymbols("XLE", from="2009-07-01", to="2011-6-30")
[1] "XLE"
> vde = log(as.numeric(VDE$VDE.Adjusted))
> xle = log(as.numeric(XLE$XLE.Adjusted))
> plot(vde - xle, type='l')
> acf(vde - xle)
> m1 = lm(vde ~ xle)
> m1
Call: lm(formula = vde ~ xle)
Coefficients:
(Intercept)         xle
  0.3721     1.0023

> wt = vde - xle
> m2 = ar(wt, method="mle")
> m2$order
[1] 2
> m3 = arima(wt, order=c(2,0,0))
> m3
Call: arima(x = resi, order = c(2, 0, 0))
Coefficients:
            ar1       ar2     intercept
ar1       0.6473   0.2870     0.3810
s.e. 0.0426  0.0427    0.0008
sigma^2 estimated as 1.529e-06: log likelihood = 2663.65, aic = -5319.29

> p1 = c(1, -m3$coef[1:2]); root = polyroot(p1)
> root
[1] 1.053103+0i -3.308222+0i
> 1/Mod(root)
[1] 0.9495747 0.3022772

> getSymbols("XLE", from="2009-07-01", to="2011-12-31")
[1] "XLE"
> getSymbols("VDE", from="2009-07-01", to="2011-12-31")
[1] "VDE"
> vde=log(as.numeric(VDE$VDE.Adjusted))
> xle=log(as.numeric(XLE$XLE.Adjusted))
> m1=lm(vde~xle)
> m1
Call: lm(formula = vde ~ xle)
Coefficients:
(Intercept) xle
 0.3864 0.9986

> wt=vde-xle
> plot(wt,type='l')
> m2=ar(wt,method="mle")
> m2$order
[1] 3
> m3=arima(wt,order=c(3,0,0))
> m3
Call: arima(x = wt, order = c(3, 0, 0))
Coefficients:
          ar1    ar2    ar3 intercept
 0.6250 0.2457 0.0761  0.3803
s.e.  0.0396 0.0458 0.0398  0.0009
sigma^2 estimated as 1.604e-06: log likelihood = 3318.65, aic = -6627.3
> p2=c(1,-m3$coef[1:3]); root2=polyroot(p2)
> root2
[1] 1.039052+0.000000i -2.132856+2.844579i -2.132856-2.844579i
> 1/Mod(root2)
[1] 0.9624154 0.2812642 0.2812642