THE UNIVERSITY OF CHICAGO
Graduate School of Business
Business 41202, Spring Quarter 2006, Mr. Ruey S. Tsay

Solutions to Homework Assignment #3

1. Problem 1. For your information, I included the R commands used in the analysis. Taking the log transformation, the series shows a downward trend, indicating non-stationarity. However, the ACF of the first differenced log series suggests an MA(3) model. [If PACF is used, an AR(3) is suggested.] Let r_t be the log BAA yields. The fitted ARIMA(0,1,3) model is

$$(1-B)r_t = (1+.3299B+.0167B^2+.1028B^3)a_t,$$

where the estimated variance of a_t is 8.6×10^{-5}. Model checking, using **tsdiag** command, indicates that the fitted model is adequate.

Based on the fitted model, the 1-step to 4-step ahead forecasts are 1.8885, 1.8903, 1.8912, and 1.8912, respectively. The standard errors of the prediction are 0.00927, 0.01543, 0.01985, 0.02398, respectively.

Notice that the 2nd MA coefficient is insignificant at the 5% level. One can drop it by fixing the lag-2 coefficient to zero. For this particular instance, deleting the MA(2) coefficient has essentially no impact on the forecasts.

If an AR(3) model is entertained for the differenced series, then the model become

$$(1-.3277B-0.0844B^2-0.1220B^3)(1-B)r_t = a_t,$$

where the variance of a_t is $8.6{\times}10^{-05}$.

The minimum AIC criterion slightly prefers the ARIMA(0,1,3) model.

```
> da=read.table("w-baa.txt")
> baa=log(da[,4])
> plot(baa,type='l')
> acf(baa)
> acf(diff(baa))
> m1=arima(baa,order=c(0,1,3))
> m1
Call:
arima(x = baa, order = c(0, 1, 3))

Coefficients:
```

```
         ma1     ma2     ma3
      0.3299  0.0167  0.1028
s.e.  0.0281  0.0284  0.0281

sigma^2 estimated as 8.602e-05:  log likelihood = 4125.73,  aic = -8243.46
> tsdiag(m1)
> predict(m1,4)
$pred
Start = 1267
[1] 1.888475 1.890294 1.891214 1.891214
$se
[1] 0.00927487 0.01543287 0.01985397 0.02397713

> m1=arima(baa,order=c(0,1,3),fixed=c(NA,0,NA))
> tsdiag(m1)
> predict(m1,4)
$pred
Start = 1267
[1] 1.888199 1.889788 1.890669 1.890669
$se
[1] 0.009276146 0.015398004 0.019701527 0.023716125

> m2=arima(baa,order=c(3,1,0))
> m2

Call:
arima(x = baa, order = c(3, 1, 0))

Coefficients:
         ar1      ar2     ar3
      0.3277  -0.0844  0.1220
s.e.  0.0279   0.0293  0.0279

sigma^2 estimated as 8.606e-05:  log likelihood = 4125.46,  aic = -8242.92
> tsdiag(m2)
> predict(m2,4)
$pred
Start = 1267
[1] 1.888238 1.890040 1.892081 1.892926
$se
[1] 0.009276859 0.015419557 0.019868817 0.024007410
```

2. The log series of the two time series are unit-root nonstationary so that we consider the first differenced series. Let y_t be the log BAA yield and x_t the log federal fund rate. If you regress y_t on x_t and check the residuals of the fitted model, the residuals

would have strong serial correlations. Thus, we focus on the differenced data. Let $c_t = (1 - B)y_t$ and $d_t = (1 - B)x_t$. We regress c_t on d_t and find the model to be adequate. The fitted model is

$$c_t = -0.00074 + 0.01313 d_t + e_t,$$

where the estimated variance of e_t is 0.0098. The R^2 of the prior linear regression is low at 0.32%. The two estimated coefficients are statistically significant at the 5% level. However, ACF of the residual shows some serial correlations and suggests an MA(3) model for the residuals. The refined model is

$$c_t = -0.0007 + 0.0076 d_t + e_t,$$

$$e_t = (1 + 0.326B + .014B + .099B^3)a_t,$$

and the estimated variance of a_t is 8.56×10^{-5}. The Ljung-Box statistics of the residuals indicate that the model is adequate.

Finally, t-ratio of the coefficient of d_t is $0.0076/0.0052 = 1.462$ with p-value 0.144. Consequently, there is no significant concurrent linear relationship between c_t and d_t.

Note that one can specify an AR(3) model for the residual series and would obtain the same conclusion.

```
> da=read.table("w-ff.txt")
> ff=log(da[,4])
> plot(ff,type='l')
> plot(ff,baa)
> m3=lm(baa~ff)
> summary(m3)
lm(formula = baa ~ ff)

Residuals:
     Min       1Q   Median       3Q      Max
-0.35092 -0.12977  0.01017  0.13148  0.36710

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.715881   0.012634  135.81   <2e-16 ***
ff          0.309230   0.007382   41.89   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1632 on 1264 degrees of freedom
Multiple R-Squared: 0.5813,     Adjusted R-squared: 0.5809
```

```
> Box.test(m3$residuals,lag=12)
X-squared = 13710.86, df = 12, p-value < 2.2e-16

> x1=diff(ff)
> y1=diff(baa)
> m3=lm(y1~x1)
> summary(m3)
lm(formula = y1 ~ x1)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0434914 -0.0059899 -0.0003439  0.0057974  0.0488949

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0007367  0.0002756  -2.673   0.0076 **
x1           0.0131264  0.0065291   2.010   0.0446 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.009799 on 1263 degrees of freedom
Multiple R-Squared: 0.00319,    Adjusted R-squared: 0.002401
> acf(m3$residuals)
> m4=arima(y1,order=c(0,0,3),xreg=x1)
> m4
arima(x = y1, order = c(0, 0, 3), xreg = x1)

Coefficients:
         ma1     ma2     ma3  intercept      x1
      0.3262  0.0137  0.0987     -7e-04  0.0076
s.e.  0.0282  0.0284  0.0282      4e-04  0.0052

sigma^2 estimated as 8.561e-05:  log likelihood = 4128.75,  aic = -8245.5
> Box.test(m4$residuals,lag=12,type='Ljung')
X-squared = 8.5719, df = 12, p-value = 0.739

> t=.0076/.0052
> pv=2*(1-pnorm(t))
> pv
[1] 0.1438677
> t
[1] 1.461538
>
> m4=arima(y1,order=c(3,0,0),xreg=x1)
> m4
```

```
arima(x = y1, order = c(3, 0, 0), xreg = x1)

Coefficients:
         ar1      ar2     ar3  intercept      x1
      0.3239  -0.0854  0.1174     -7e-04  0.0081
s.e.  0.0280   0.0293  0.0280      4e-04  0.0053

sigma^2 estimated as 8.568e-05:  log likelihood = 4128.27,  aic = -8244.55
> Box.test(m4$residuals,lag=12,type='Ljung')
X-squared = 9.0534, df = 12, p-value = 0.6984

> t=.0081/.0053
> pv=2*(1-pnorm(t))
> t
[1] 1.528302
> pv
[1] 0.1264376
```

3. Let ew_t be the equal-weighted index returns, in percentages, and mo_t be the indicator variable for Monday. The simple lienar regression is

$$ew_t = 0.1159 - 0.0981 mo_t + e_t,$$

where the estimated variance of e_t is 0.75. The Box-Ljung statistics of the residuals indicate that the residuals are serially correlated with a small p-value for Q(10).

The ACF of residuals also suggests an MA(4) model. Therefore, we entertain a regression model with an MA(4) residuals. The fitted model is

$$ew_t = 0.117 - 0.101 mo_t + a_t + 0.155 a_{t-1} + .063 a_{t-2} + .072 a_{t-3} + .083 a_{t-4},$$

where the MA coefficients are all significant at the 5% level and the estimated residual variance is 0.7197. Model checking using **tsdiag** fails to indicate model inadequacy, and we use the fitted model to make inference.

Based on the fitted model, the t-ratio of the Monday effect is -1.794 with p-value 0.073. Therefore, the Monday effect is negative and statistically significant at the 10% level. Similar conclusion is obtained if an AR(4) model is entertained for the residuals.

R commands:

```
> da=read.table("d-dellew0105-wk.txt")
> dell=da[,4]*100
> ew=da[,5]*100
> mo=da[,6]
> tu=da[,7]
```

```
> we=da[,8]
> th=da[,9]
> m1=arima(ew,xreg=mo,order=c(0,0,0))
> m1
arima(x = ew, order = c(0, 0, 0), xreg = mo)

Coefficients:
      intercept       mo
         0.1159  -0.0981
s.e.     0.0271   0.0625

sigma^2 estimated as 0.75:  log likelihood = -1601.5,  aic = 3209
> acf(m1$residuals)
> Box.test(m1$residuals,lag=10,type='Ljung')
X-squared = 88.2008, df = 10, p-value = 1.221e-14

> pacf(m1$residuals)
> m1=arima(ew,xreg=mo,order=c(0,0,4))
> m1
arima(x = ew, order = c(0, 0, 4), xreg = mo)

Coefficients:
         ma1     ma2     ma3     ma4  intercept       mo
      0.1551  0.0625  0.0716  0.0832     0.1170  -0.1008
s.e.  0.0282  0.0288  0.0281  0.0267     0.0345   0.0562

sigma^2 estimated as 0.7197:  log likelihood = -1575.62,  aic = 3165.24
> tsdiag(m1)
> t=-.1008/.0562
> pv=2*pnorm(t)
> t
[1] -1.793594
> pv
[1] 0.07287795
> Box.test(m1$residuals,lag=10,type='Ljung')
X-squared = 17.6811, df = 10, p-value = 0.06059

> m2=arima(ew,xreg=mo,order=c(4,0,0))
> m2
arima(x = ew, order = c(4, 0, 0), xreg = mo)

Coefficients:
         ar1     ar2     ar3     ar4  intercept       mo
      0.1544  0.0417  0.0531  0.0746     0.1169  -0.0998
s.e.  0.0281  0.0288  0.0287  0.0285     0.0369   0.0570
```

```
sigma^2 estimated as 0.7184:  log likelihood = -1574.56,  aic = 3163.13
> Box.test(m2$residuals,lag=10,type='Ljung')
X-squared = 15.3101, df = 10, p-value = 0.1212

> tsdiag(m2)
> t=-.0998/.057
> pv=2*pnorm(t)
> t
[1] -1.750877
> pv
[1] 0.07996707
```

4. There are several ways to solve this problem. Since there are five trading days in a week, we can use four indicator variables to study the weekday effect. Consider the model

$$r_t = \beta_0 + \beta_1 mo_t + \beta_2 tu_t + \beta_3 we_t + \beta_4 th_t + e_t.$$

If t is a Friday, then $mo_t = tu_t = we_t = th_t = 0$ and the model reduces to

$$r_t = \beta_0 + e_t.$$

Thus, the intercept term β_0 denotes the Friday effect, and β_1 is the difference between Monday and Friday, β_2 is the difference between Tuesday and Friday, etc.

For the Dell stock returns, the fitted model is

$$r_t = -0.0988 + 0.1556 mo_t + 0.1573 tu_t + .2969 we_t + .2541 th_t + e_t.$$

Compared with their standard errors, all coefficients are not significant at the 10% level. However, the residual ACF shows some serial correlations with Q(10) = 30.09 and p-value 0.00083. The ACF also indicates an MA(2) model for the residuals.

The refined model is

$$r_t = -0.098 + 0.147 mo_t + 0.164 tu_t + 0.294 we_t + 0.250 th_t + e_t,$$

$$e_t = a_t - 0.0026 a_{t-1} - 0.1181 a_{t-2}.$$

The residual checking shows that the fitted model is adequate. Again, all weekday coefficients are insigificant at the 5% level (compared with their standard error). Thus, there is no weekday effect.

```
> xx=cbind(mo,tu,we,th)
> m5=arima(dell,xreg=xx,order=c(0,0,0))
> m5
arima(x = dell, order = c(0, 0, 0), xreg = xx)
```

```
Coefficients:
      intercept      mo      tu      we      th
        -0.0988  0.1557  0.1573  0.2969  0.2541
s.e.     0.1587  0.2280  0.2231  0.2234  0.2245

sigma^2 estimated as 6.348:  log likelihood = -2942.79,  aic = 5897.58
> acf(m5$residuals)
> Box.test(m5$residuals,lag=10,type='Ljung')
X-squared = 30.0871, df = 10, p-value = 0.000829

> m5=arima(dell,xreg=xx,order=c(0,0,2))
> m5
arima(x = dell, order = c(0, 0, 2), xreg = xx)

Coefficients:
          ma1      ma2  intercept      mo      tu      we      th
      -0.0026  -0.1181    -0.0983  0.1473  0.1644  0.2942  0.2498
s.e.   0.0282   0.0287     0.1584  0.2276  0.2345  0.2363  0.2241

sigma^2 estimated as 6.265:  log likelihood = -2934.53,  aic = 5885.05
> tsdiag(m5)
> Box.test(m5$residuals,lag=10,type='Ljung')
X-squared = 12.7031, df = 10, p-value = 0.2407
```

5. For the log earning series, the ACF shows strong serial correlations. The ACF of the first differenced series shows strong seasonal correlations at lags 4, 8, 12, 16, etc. Thus, we also take the seasonal difference. ACF of $(1-B)(1-B^4)x_t$ shows a single significant serial correlation at lag 1, where x_t is the log earnings. Therefore, we entertain the model

$$(1-B)(1-B^4)x_t = (1-\theta B)a_t.$$

The fitted model is

$$(1-B)(1-B^4)x_t = a_t - 0.3634a_{t-1},$$

where the estimated variance of a_t is 0.004. Model checking using **tsdiag** and the Ljung-Box statistics of residuals indicate that the fitted model is adequate. Note that the estimated coefficient is significant at the 5% level.

The 1-step to 4-step ahead forecasts are $-0.2401, -0.4011, -0.5930, -0.2004$, respectively, and their standard errors are $0.0635, 0.0754, 0.0854, 0.0945$. Taking the expoential transformation of the point forecasts, we have 0.79, 0.67, 0.55, and 0.82 as the 1-step to 4-step ahead earning forecasts.

```
> da=read.table("q-pgeps.txt")
```

```
> pg=log(da[,4])
> plot(pg,type='l')
> acf(pg)
> acf(diff(pg))
> acf(diff(diff(pg),4))
> m6=arima(pg,order=c(0,1,1),seasonal=list(order=c(0,1,0),period=4))
> m6
arima(x = pg, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 0), period = 4))

Coefficients:
          ma1
      -0.3634
s.e.   0.1782

sigma^2 estimated as 0.004031:  log likelihood = 68.16,  aic = -132.33
> tsdiag(m6)
> Box.test(m6$residuals,lag=12,type='Ljung')
X-squared = 14.4913, df = 12, p-value = 0.2704

> predict(m6,4)
$pred
Start = 57
[1] -0.2401441 -0.4010745 -0.5929655 -0.2004038
$se
[1] 0.06349137 0.07526582 0.08543262 0.09451202

> p1=predict(m6,4)
> exp(p1$pred)
Start = 57
[1] 0.7865145 0.6696002 0.5526859 0.8184002
```