A Time Series Analysis of 252 Successive Days of the Russell Index.

1 Introduction and Summary .

I was provided by a student in the class with 252 daily closing values of the Russell Index of low capitalization high technology stocks. Some background on this series would be provided if I wanted to get a good grade – maybe a paragraph or two.

An initial analysis of the series revealed apparent nonstationarity which seemed to be removed by taking one difference. The sample ACF/PACF plots suggested that either an AR(1) or an MA(1) model for the differenced series may be appropriate. The two models were fit and the diagnostics were examined. These revealed no significant lack of fit for either model. Subsequent forecasting with the two models showed only slight differences. We conclude that the models are equally valid and either can be used. From the point of view of simplicity and aesthetics, we would prefer the AR(1) model for the differenced series (implying an ARIMA(1,1,0) model for the original series).

2 Detailed Results.

A time series plot of the original series is shown in Figure 1, and the sample ACF of the original series in Figure 2. As is expected of such financial data, it appears nonstationary. As was discussed in the context of an analysis of some Dow-Jones data in lecture, it would be expected that a first difference might be successful in eliminating the nonstationarity. This is also the case for the IBM stock price series (Series B in Box, Jenkins, and Reinsel) (see Table 6.4,p. 196). The time series plot and Sample ACF for the first differenced series are shown in Figures 3 and 4, respectively, and these do strongly suggest that the differenced series is stationary.

An examination of the sample ACF and sample PACF (the latter is shown in Figure 5) for the differenced series reveals significant values in both at lag 1, but no other significant values at other low lags. There are some significant values for both the ACF and PACF at lag 16, and the PACF exhibits a barely significant value at lag 9. However, we have no reason to suspect that there would be large values of the ACF or PACF at such lags and small values otherwise, so we tend to suspect this is the result of chance fluctuations rather than truly significant autocorrelations. Given this judgement, these plots suggest that either an MA(1) model (from the ACF plot) or an AR(1) model (from the PACF plot) might be appropriate.

Both models were fit using the arima.mle function in Splus. These were fit to the differenced series for interpretability reasons. Note that it says in the "Warning" for the help on arima.mle that

> Unlike the ar function, the mean of the series will not be estimated by arima.mle unless you use the xreg argument. arima.mle assumes a zero mean series.

The sample mean of the observed series was subtracted off and the centered series was passed to the arima.mle function in both cases.

2.1 AR(1) Model.

An AR(1) model was fit using the Splus arima.mle function and standard diagnostics computed using the arima.diag function. The estimated AR coefficient was

$$\hat{\phi} = 0.2106175. \tag{1}$$

The diagnostic plot is shown if Figure 6. Focusing mainly on the bottom panel of the diagnostic plots (the portmanteau tests) we see that there are no significant values, indicating no serious inadequacy of the model.

2.2 MA(1) Model.

n MA(1) model was fit using the Splus arima.mle function and standard diagnostics computed using the arima.diag function. The estimated MA coefficient was

$$\hat{\theta} = -0.2038795. \tag{2}$$

The diagnostic plot is shown if Figure 7. Focusing again on the portmanteau tests in the bottom panel of the diagnostic plots we see that there are no evidence of inadequacy of the model.

2.3 Comparison of Forecasts.

To see if there was any essential difference in the models in terms of our eventual objective (forecasting), we computed forecasts and forecast intervals out to a lead time of 20 days beyond the end of the observed series. To accomplish this fairly easily, we went back and fit ARIMA(p,1,q) models to the series where one of p or q is 1 and the other is 0. In order the incorporate the observed nonzero mean of the differences, we included a linear regression (a constant mean after taking differences corresponds to a linear trend whose slope is the mean of the differneces). The fitting is slightly different with differences in the further decimal places, but we judged it insignificant. One can view the results in the listing of Splus code in the next section.

The results of the forecasting exercise are shown in Firuger 8. We see that the point forecasts are identical to within the accuracy of the plot, and one of the models has slightly wider forecast intervals. Printing out the prediction standard deviations revealed that it was the ARIMA(1,1,0) model with the wider forecast intervals.

2.4 Conclusions.

We conclude that from a practical point of view, there is virtually no reason to prefer one of the two models which we fit. They are both deemed adequate by the standard diagnostic test and the forecasts and forecast intervals are not significantly different. As it is easier to fit and predict with AR rather than MA models, one might prefer the ARIMA(1,1,0) model.

3 Listing of S-Plus Code.

The Splus code used in this analysis is given below. This was created by using the unix script function (and the exit function to terminate the script) and then editing the typescript file which is output by the script function. We have added some comments into the flow to explain some details and results.

```
Script started on Sun May 02 06:26:10 1999
stat.rice.edu% Splus3.4
S-PLUS : Copyright (c) 1988, 1996 MathSoft, Inc.
S : Copyright AT&T.
Version 3.4 Release 1 for Sun SPARC, SunOS 5.3 : 1996
Working data will be in .Data
> russell_scan("../russell_399.txt")
> length(russell)
[1] 252
> X11()
> tsplot(russell)
> #clearly looks nonstationary
> #THIS APPEARS IN FIGURE 1.
> russell.acf_acf(russell)
> #correlations very high for long lags -- sign of nonstationarity.
> #THIS APPEARS IN FIGURE 2.
> #try differencing
> drussell_diff(russell)
> tsplot(drussell)
> #looks much better.  SEE FIGURE 3.
> drussell.acf_acf(drussell)
```

```
> #one significant acf @ lag 1.  SEE FIGURE 4.
> drussell.pacf_acf(drussell,type="partial")
> #one significant pacf @ lag 1  SEE FIGURE 5.
> #try fitting both AR(1) and MA(1), see which is better
> q()
stat.rice.edu% exit
stat.rice.edu%
script done on Sun May 02 06:46:22 1999
```

```
Script started on Sun May 02 07:36:19 1999
stat.rice.edu% Splus3.4
S-PLUS : Copyright (c) 1988, 1996 MathSoft, Inc.
S : Copyright AT&T.
Version 3.4 Release 1 for Sun SPARC, SunOS 5.3 : 1996
Working data will be in .Data
> mean(drussell)
[1] -0.3047809
> #clearly can't assume a 0 mean series.
> drussell0_drussell-mean(drussell)
> fit.ar1.0_arima.mle(drussell0,model=list(ar=0))
> diag.ar1.0_arima.diag(fit.ar1.0)
> #SEE FIGURE 6.
> fit.ar1.0$model
$ar:
[1] 0.2106175
```

```
$ndiff:
[1] 0


$order:
[1] 1 0 0


> fit.ma1.0_arima.mle(drussell0,model=list(ma=0))
> fit.ma1.0$model
$ma:
[1] -0.2038795


$ndiff:
[1] 0


$order:
[1] 0 0 1


> diag.ar1.0_arima.diag(fit.ar1.0)
> diag.ma1.0_arima.diag(fit.ma1.0)
> #SEE FIGURE 7.
> q()
stat.rice.edu% exit
stat.rice.edu%
script done on Sun May 02 07:55:00 1999



Script started on Wed May 05 19:15:03 1999
```

```
markov.stat.rice.edu% Splus3.4
S-PLUS : Copyright (c) 1988, 1996 MathSoft, Inc.
S : Copyright AT&T.
Version 3.4 Release 1 for Sun SPARC, SunOS 5.3 : 1996
Working data will be in .Data
> length(russell)
[1] 252
> #fitting ARIMA(p,1,q) models for forecasting
> fit.arima110_arima.mle(russell,model=list(order=c(1,1,0)),
+ xreg=(1:252))
> fit.arima110$model
$order:
[1] 1 1 0

$ar:
[1] 0.2106179

$ndiff:
[1] 1

> fit.arima110$reg.coef
[1] -0.3042274
> #note that the fitted AR coeff and reg.coef are very close to the
> # sample mean and the fitted AR model above.
> fit.arima011_arima.mle(russell,model=list(order=c(0,1,1)),
+ xreg=(1:252))
> fit.arima011$model
```

```
$order:
[1] 0 1 1

$ndiff:
[1] 1

$ma:
[1] -0.2038836

> fit.arima011$reg.coef
[1] -0.2988433
> fc110_arima.forecast(russell,model=fit.arima110$model,n=20,xreg=(1:272),
+ reg.coef=fit.arima110$reg.coef)
> fc011_arima.forecast(russell,model=fit.arima011$model,n=20,xreg=(1:272),
+ reg.coef=fit.arima011$reg.coef)
> X11()
> tsplot(fc110$mean,fc110$mean+2*fc110$std.err,fc110$mean-2*fc110$std.err,
+ fc011$mean,fc011$mean+2*fc011$std.err,fc011$mean-2*fc011$std.err)
> title(main="Forecasts and Intervals for both models")
> cbind(fc110$std.err,fc011$std.err)
           [,1]      [,2]
 [1,]  4.909526  4.905599
 [2,]  7.709044  7.677436
 [3,]  9.868715  9.686132
 [4,] 11.658519 11.344570
 [5,] 13.212518 12.789738
 [6,] 14.602928 14.087424
```

```
 [7,] 15.872167 15.275262
 [8,] 17.047199 16.377172
 [9,] 18.146310 17.409477
[10,] 19.182548 18.383906
[11,] 20.165609 19.309224
[12,] 21.102924 20.192183
[13,] 22.000341 21.038117
[14,] 22.862560 21.851327
[15,] 23.693422 22.635340
[16,] 24.496119 23.393092
[17,] 25.273335 24.127057
[18,] 26.027353 24.839344
[19,] 26.760133 25.531767
[20,] 27.473375 26.205901
> #Looks like the ARIMA(1,1,0) model has slightly larger prediction
> # standard deviations.
> q()
markov.stat.rice.edu% exit
markov.stat.rice.edu%
script done on Wed May 05 19:31:02 1999
```

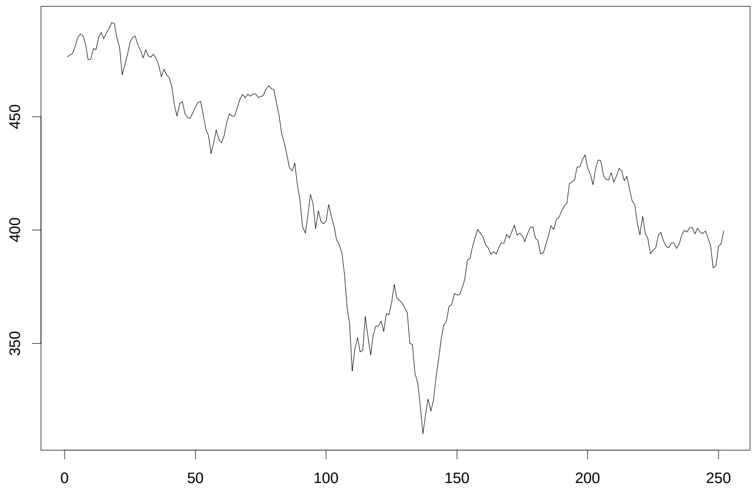

Figure 1: Time Series Plot of Russell Series.

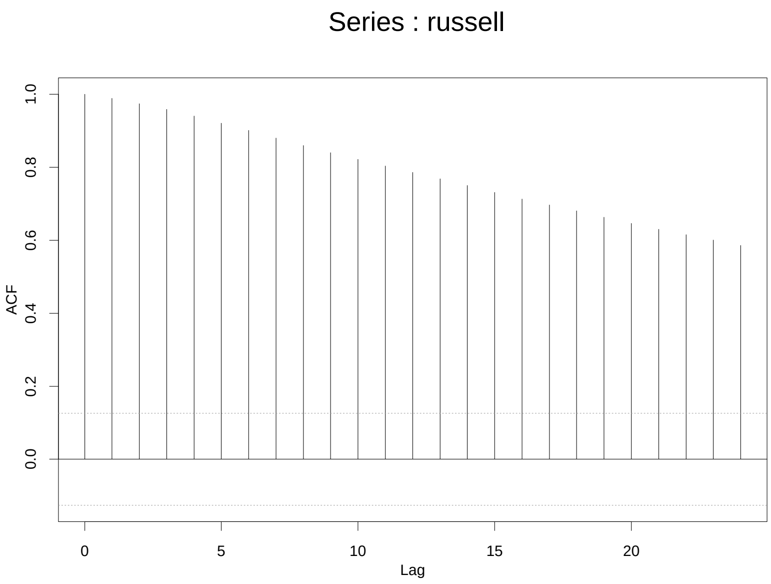

Figure 2: Sample ACF of Russell Series.

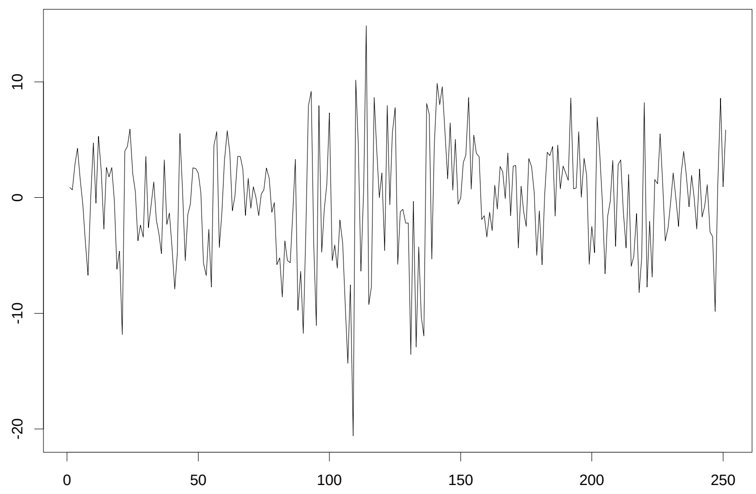

Figure 3: Time Series Plot of First Differenced Russell Series.

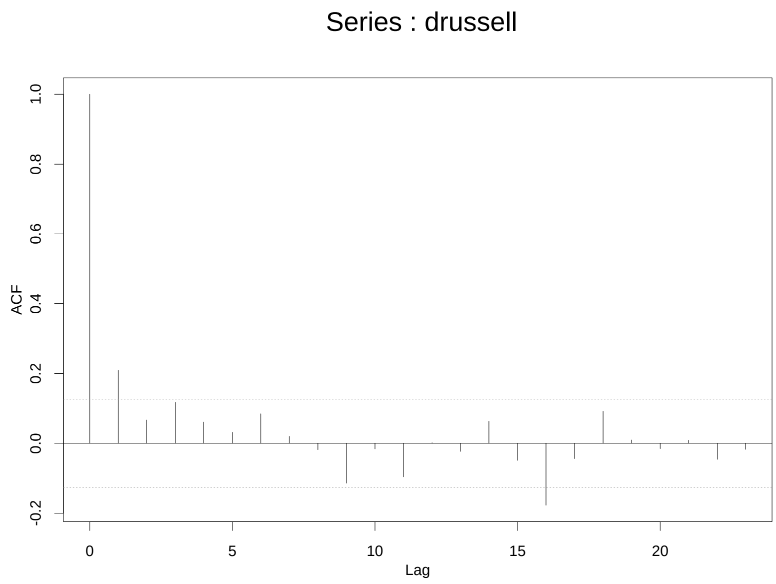

Figure 4: Sample ACF of First Differenced Russell Series.

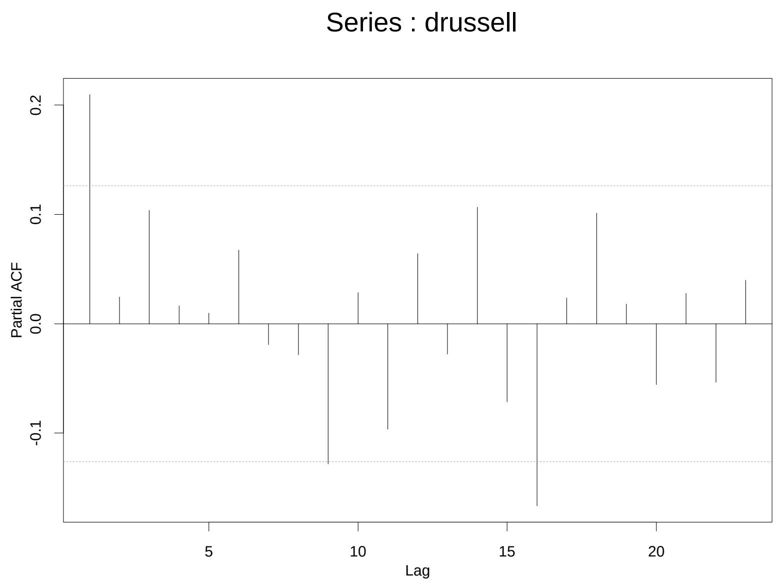

Figure 5: Sample PACF of First Differenced Russell Series.

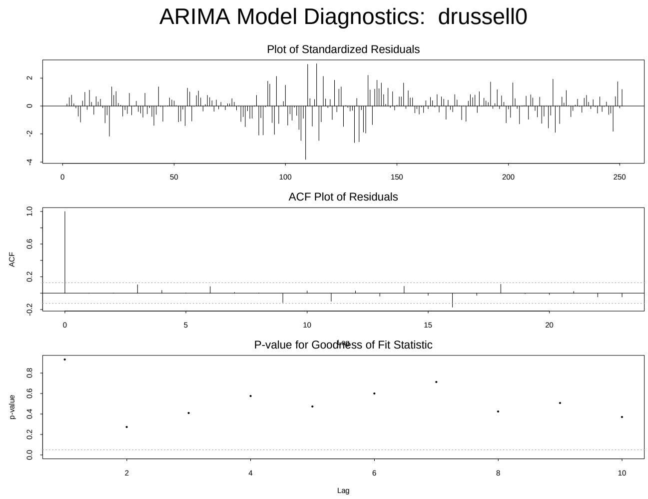

Figure 6: Diagnostic plots for the AR(1) fit to the first differenced Russell series.

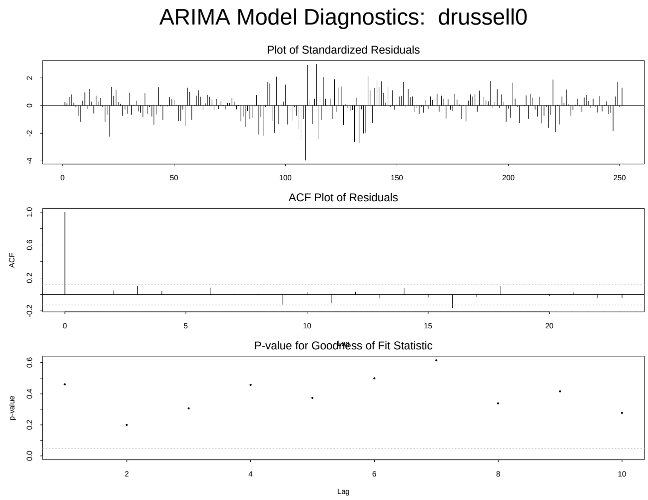

Figure 7: Diagnostic plots for the MA(1) fit to the first differenced Russell series.

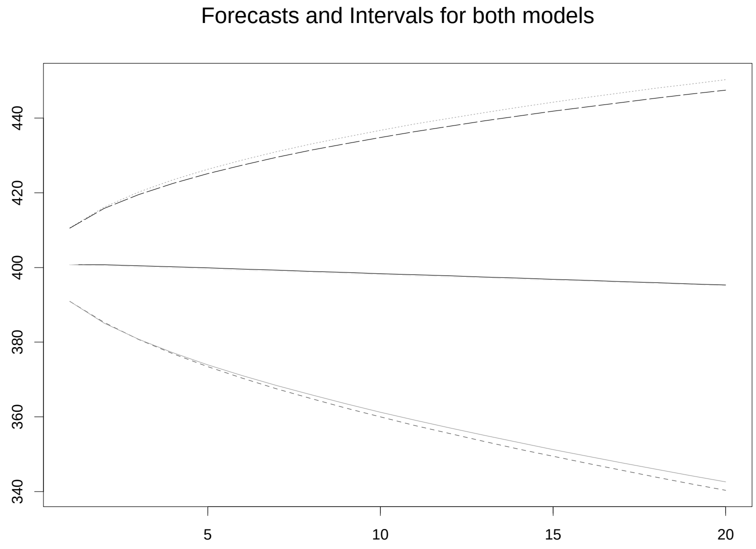

Figure 8: Plot of point forecasts and 95% forecast intervals for both the ARIMA(1,1,0) and ARIMA(0,1,1) models out to a lead time of 20 days.