M349R (Unique 54230)

**Instructor:** Gustavo Cepparo

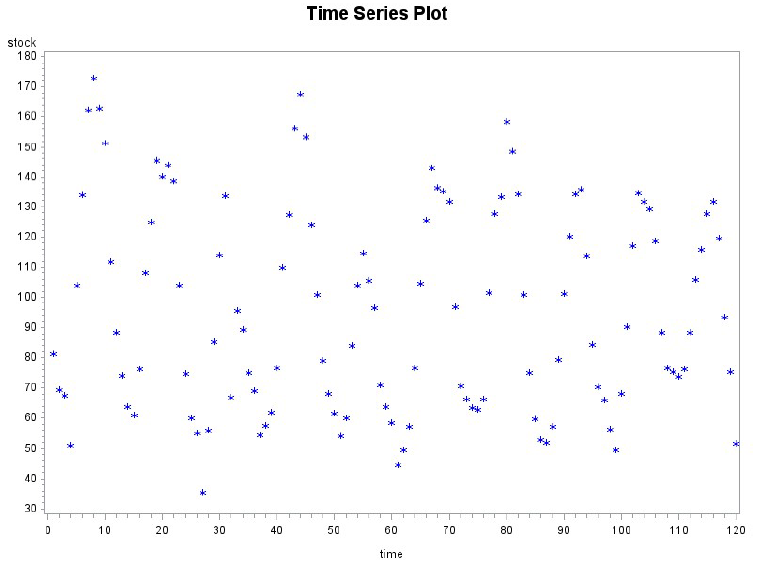
Project 6

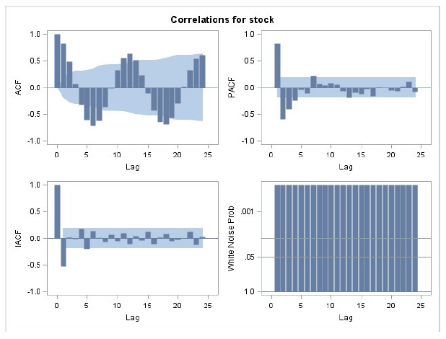
Juan Acosta

JA45384

Fall 2018

**Problem 1 Use Stock dataset (50 points)**

1. Graph the data with a time series plot and describe the time series plot. (5 pts)
2. What can we learn from the ACF and Test for White Noise? (5 pts)

It is clearly seasonal and there is no need for transformation but clear patterns indicate that we need to consider Arima models. It appears to be cutting off in the ACF and dying down on the PACF therefore an MA model seems appropriate with some differencing

1. Provide a differencing analysis. (10 pts)

Analyzing the results from:

**proc** **arima** data=stock;

identify var=stock;

identify var=stock(**1**);

identify var=stock(**12**);

identify var=stock(**1**, **12**);

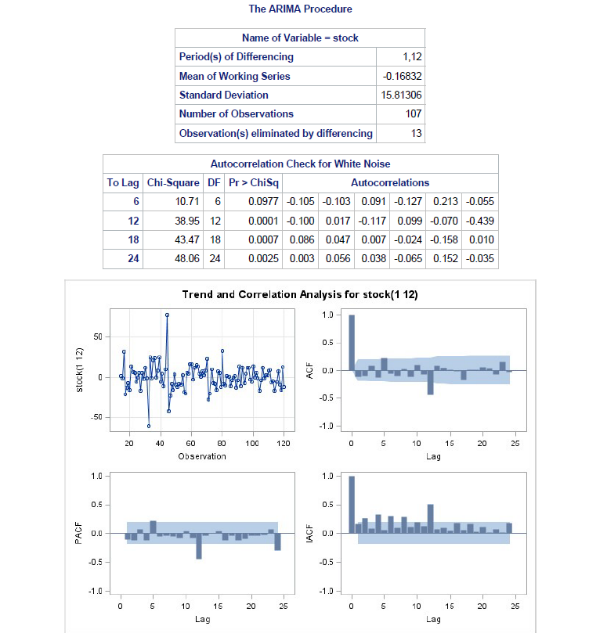
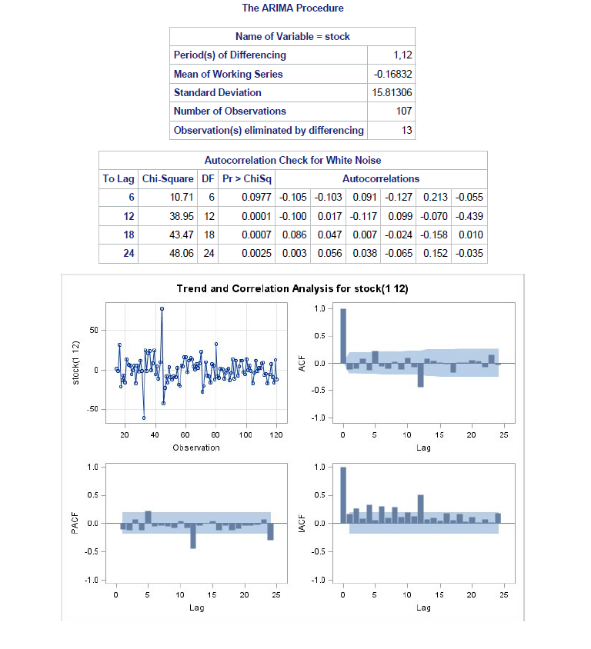
identify var=stock(**6**);

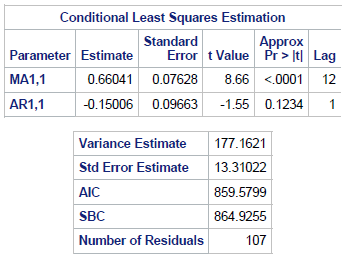
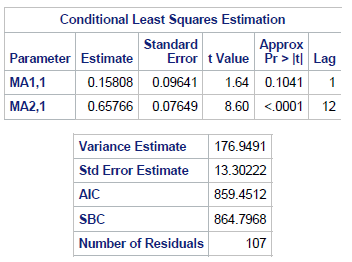
identify var=stock(**1**, **6**);

identify var=stock(**1**, **6**, **12**);

**run**;

We conclude that a differencing of (1, 12) provides the best fit.

1. Check the models Arima(0 , 1, 1)(0, 1, 1) and Arima(1, 1, 0)(0, 1, 1) check the fit of both models and compare AIC, L’Jung Box and Standard error. (10 pts)

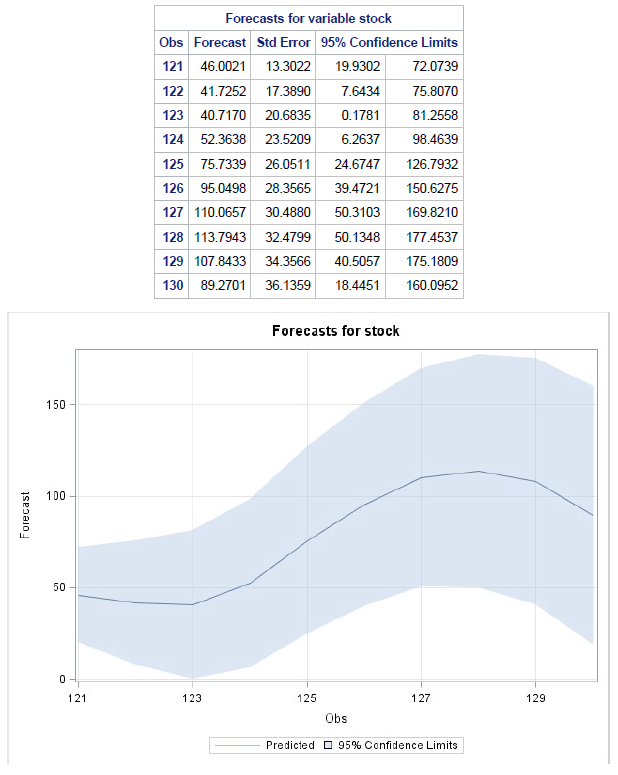


The model Arima(0 , 1, 1)(0, 1, 1) has slightly lower values for AIC and Standard error and the fit is better in terms of the ACF, PACF and white noise

1. Write the model on d in terms of the backshift operator, and then without using the backshift operator. (10 pts)

(1 – β12)(1-β)Yt\* - (1-θ12)(1-θ)at\*

Ŷt  =  Yt-12 + Yt-1 – Yt-13 - θ1at-1 – Θ1at-12 + θ1Θ1at-13

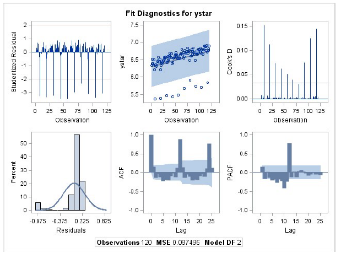
1. Forecast 10 periods. (10 pts)

**Problem 2 Use Sales dataset (50 points)**

Follow the four steps of arima modeling. Forecast 6 periods

Sales did not have constant variance therefore, I applied the logarithmic function

1. From looking at the plot, ACF and PACF graphs I computed the following differencing:

**proc** **arima** data=sales;

identify var=ystar;

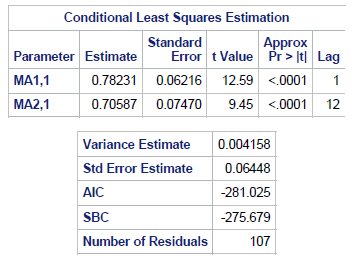
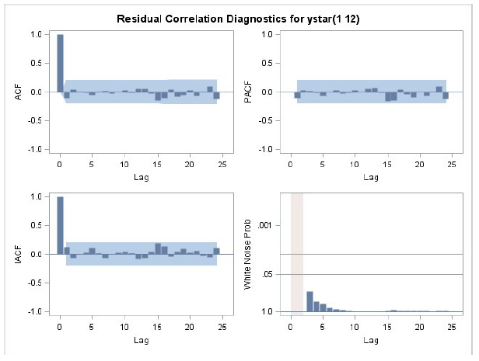
identify var=ystar(**1**);

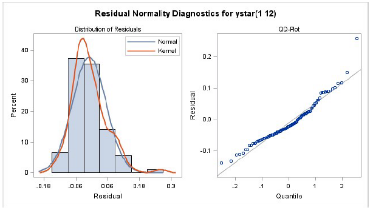
identify var=ystar(**12**);

identify var=ystar(**1**, **12**);

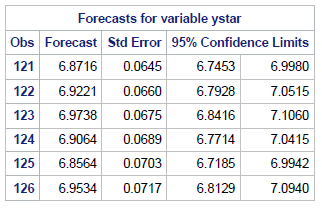
**run**;

and concluded that Log(sales)(1, 12) was preferred

1. Tentative model since it cuts off on the ACF and dies down on the PACF is an MA at 1 and 12. MU was not statistically different from 0 and retried with noconstant.
2. 
3. Residual analysis indicates model is adequate. White noise everywhere. Normality on the residuals.

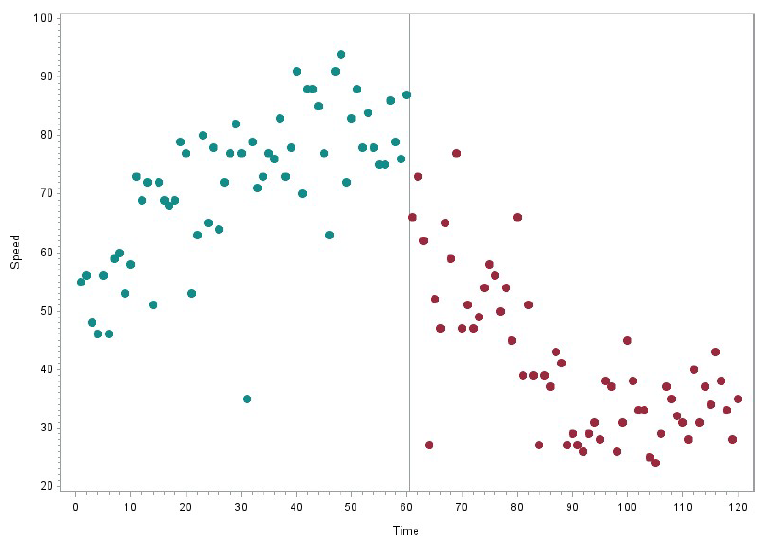


|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Obs | Forecast y | Std Error | 95% CI | |
| 121 | 964.4905167 | 1.06662558 | 850.0541 | 1094.442 |
| 122 | 1014.4481 | 1.06822672 | 891.406 | 1154.589 |
| 123 | 1068.27449 | 1.06983026 | 935.9855 | 1219.261 |
| 124 | 998.645639 | 1.07132907 | 872.5326 | 1143.101 |
| 125 | 949.9411165 | 1.07282998 | 827.5752 | 1090.291 |
| 126 | 1046.702474 | 1.074333 | 909.5045 | 1204.717 |

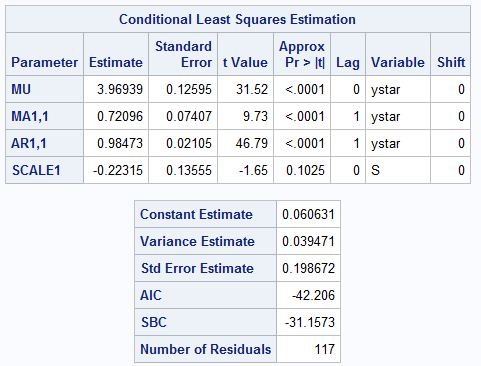
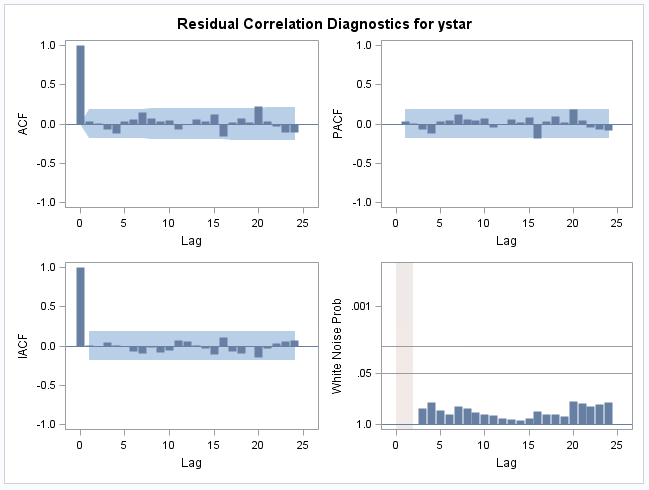
Forecast:

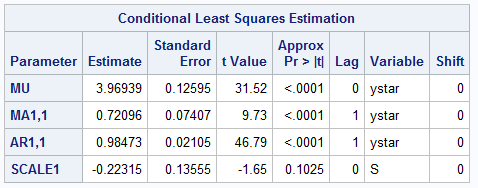
**Problem 3**

The data are the daily scores achieved by a patient with mental problems on a test of perceptual speed. The patient began receiving a powerful tranquilizer on the sixty-first day and continued receiving the drug for the rest of the sample period. It is expected that this drug will reduce perceptual speed.

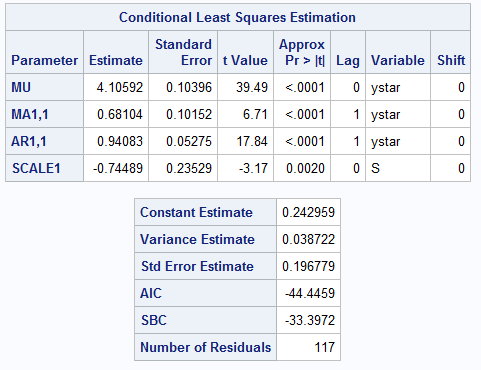
Produce a time plot of the data showing where the intervention occurred. (5 pts)

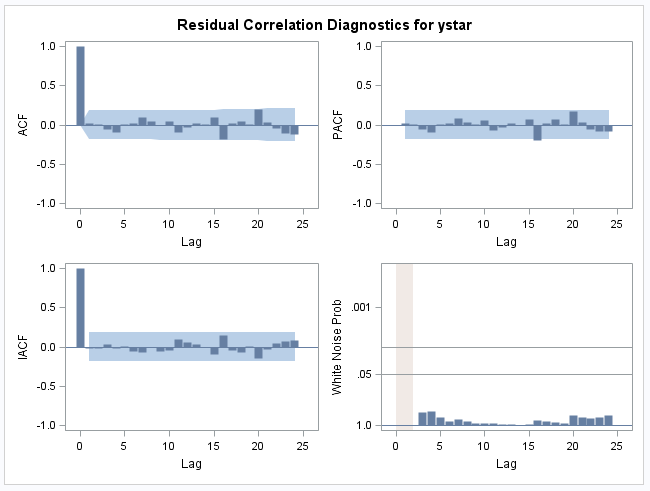
1. Fit an intervention model with a step function intervention to the series. Write down the model including the ARIMA model for the errors. (15 pts)

Used transformation log(y) to standardize variance in data. Fit model Arima(1,1,1) (sorry) to the data pre-intervention (data values 0 to 60). Fit the step function to the whole dataset

1. What does the model say about the effect of the drug? (5 pts)

The model says that the drug has a negative effect on speed. Therefore, it has the desired effect. Nonetheless, it is not “statistically” significant. Therefore, we would like to experiment with a delayed response.

1. Fit a new intervention model with a delayed response to the drug. (15 pts)

In this model we can see that SCALE is now statistically significant and our model is still a good fit

1. Which model fits the data better? Are the forecast from the models very different? (5 pts each)

The delayed effect model seems to be a better fit by means of the AIC and Std Errors. Moreover, the SCALE estimate is statistically significant in the delayed, while it is not in the Step Function.

The forecasts are slightly different as we can see in this chart:

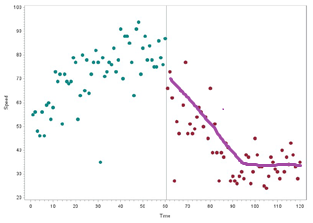
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Delayed function (log(y))** | | | | |  | **Delayed function** | | | | |
| **Obs** | **Forecast** | **Std Error** | **95% Confidence Limits** | |  | **Obs** | **Forecast** | **Std Error** | **95% Confidence Limits** | |
| **121** | 3.4657 | 0.1968 | 3.08 | 3.8514 |  | **121** | 31.99885 | 1.2175 | 21.7584 | 47.0589 |
| **122** | 3.4595 | 0.2033 | 3.061 | 3.858 |  | **122** | 31.80107 | 1.22544 | 21.3489 | 47.3705 |
| **123** | 3.4537 | 0.2089 | 3.0442 | 3.8632 |  | **123** | 31.61716 | 1.23232 | 20.9932 | 47.6175 |
| **124** | 3.4482 | 0.2138 | 3.0292 | 3.8672 |  | **124** | 31.44374 | 1.23837 | 20.6807 | 47.8083 |
| **125** | 3.443 | 0.218 | 3.0158 | 3.8702 |  | **125** | 31.28066 | 1.24359 | 20.4054 | 47.952 |
|  |  |  |  |  |  |  |  |  |  |  |
| **Step Function (log(y))** | | | | |  | **Step Function** | | | | |
| **Obs** | **Forecast** | **Std Error** | **95% Confidence Limits** | |  | **Obs** | **Forecast** | **Std Error** | **95% Confidence Limits** | |
| **121** | 3.5314 | 0.1987 | 3.1421 | 3.9208 |  | **121** | 34.17177 | 1.21982 | 23.1524 | 50.4408 |
| **122** | 3.5347 | 0.2055 | 3.132 | 3.9374 |  | **122** | 34.28473 | 1.22814 | 22.9198 | 51.2851 |
| **123** | 3.538 | 0.2118 | 3.1227 | 3.9532 |  | **123** | 34.39805 | 1.2359 | 22.7076 | 52.1018 |
| **124** | 3.5411 | 0.2179 | 3.1141 | 3.9681 |  | **124** | 34.50485 | 1.24346 | 22.5132 | 52.884 |
| **125** | 3.5443 | 0.2235 | 3.1062 | 3.9824 |  | **125** | 34.61545 | 1.25045 | 22.336 | 53.6456 |

The delayed effect function has lower estimates for the forecasted values, and they keep decreasing, contrary to the step function which are actually increasing in the forecast.

1. Construct an ARIMA model ignoring the intervention and compare the forecast with those obtain from your preferred intervention model. How much does the intervention affect the forecast? (10 pts)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Ignoring Intervention** | | | | |  | **Ignoring Intervention** | | | | |
| **Obs** | **Forecast** | **Std Error** | **95% Confidence Limits** | |  | **Obs** | **Forecast** | **Std Error** | **95% Confidence Limits** | |
| **121** | 3.5293 | 0.1975 | 3.1421 | 3.9165 |  | **121** | 34.10009 | 1.21835 | 23.1524 | 50.2244 |
| **122** | 3.5335 | 0.2069 | 3.1279 | 3.939 |  | **122** | 34.24361 | 1.22986 | 22.826 | 51.3672 |
| **123** | 3.5376 | 0.2157 | 3.1148 | 3.9604 |  | **123** | 34.3843 | 1.24073 | 22.5289 | 52.4783 |
| **124** | 3.5417 | 0.224 | 3.1027 | 3.9808 |  | **124** | 34.52556 | 1.25107 | 22.258 | 53.5599 |
| **125** | 3.5458 | 0.2318 | 3.0914 | 4.0002 |  | **125** | 34.66741 | 1.26087 | 22.0079 | 54.6091 |
|  |  |  |  |  |  |  |  |  |  |  |
| **Delayed function (log(y))** | | | | |  | **Delayed function** | | | | |
| **Obs** | **Forecast** | **Std Error** | **95% Confidence Limits** | |  | **Obs** | **Forecast** | **Std Error** | **95% Confidence Limits** | |
| **121** | 3.4657 | 0.1968 | 3.08 | 3.8514 |  | **121** | 31.99885 | 1.2175 | 21.7584 | 47.0589 |
| **122** | 3.4595 | 0.2033 | 3.061 | 3.858 |  | **122** | 31.80107 | 1.22544 | 21.3489 | 47.3705 |
| **123** | 3.4537 | 0.2089 | 3.0442 | 3.8632 |  | **123** | 31.61716 | 1.23232 | 20.9932 | 47.6175 |
| **124** | 3.4482 | 0.2138 | 3.0292 | 3.8672 |  | **124** | 31.44374 | 1.23837 | 20.6807 | 47.8083 |
| **125** | 3.443 | 0.218 | 3.0158 | 3.8702 |  | **125** | 31.28066 | 1.24359 | 20.4054 | 47.952 |
|  |  |  |  |  |  |  |  |  |  |  |
| **Step Function (log(y))** | | | | |  | **Step Function** | | | | |
| **Obs** | **Forecast** | **Std Error** | **95% Confidence Limits** | |  | **Obs** | **Forecast** | **Std Error** | **95% Confidence Limits** | |
| **121** | 3.5314 | 0.1987 | 3.1421 | 3.9208 |  | **121** | 34.17177 | 1.21982 | 23.1524 | 50.4408 |
| **122** | 3.5347 | 0.2055 | 3.132 | 3.9374 |  | **122** | 34.28473 | 1.22814 | 22.9198 | 51.2851 |
| **123** | 3.538 | 0.2118 | 3.1227 | 3.9532 |  | **123** | 34.39805 | 1.2359 | 22.7076 | 52.1018 |
| **124** | 3.5411 | 0.2179 | 3.1141 | 3.9681 |  | **124** | 34.50485 | 1.24346 | 22.5132 | 52.884 |
| **125** | 3.5443 | 0.2235 | 3.1062 | 3.9824 |  | **125** | 34.61545 | 1.25045 | 22.336 | 53.6456 |

From the charts we can see that the intervention really affects the forecasted values. Moreover, the values retain the decreasing trend in the model with accounts for the intervention, while the one that ignores it does not present a decreasing trend (it is in fact increasing).



Finally, it would probably be a better model to take into account the curve of the data 60 onward where we can see that approximately around 90 it flattens out. Therefore, if we adjust the delayed S such that it is increasing rom 61 to 90 and then it becomes 1, we would probably obtain even better forecast estimates.

**Code**

**Problem 0**

meta$Gastfem <- meta$Gastric\*meta$Sex

fit1 <- lm(Metabol ~ Gastric + Gastfem +0, data=meta)

summary(fit1)

est1 <- 1.9278/(1.9278 - 1.2021)

est1

newmeta <- meta[-c(31, 32), ]

fit2 <- lm(Metabol ~ Gastric + Gastfem +0, data=newmeta)

summary(fit2)

est2 <- 1.5989/(1.5989 - 0.8732)

est2

cov(newmeta$Gastric, newmeta$Gastfem, use = "everything")

vcov(fit2)

**Problem 1**

**data** stock;

SET stock;

time = \_N\_;

**run**;

**proc** **print** data=stock;

**run**;

**proc** **gplot** data=stock;

plot stock \* time;

symbol1 v=star c=blue;

title "Time Series Plot";

**run**;

**quit**;

title;

**proc** **autoreg** data=stock;

model stock = time/dwprob;

**run**;

**proc** **corr** data=stock;

var stock time;

**run**;

**proc** **timeseries** data=stock plots=(series residual histogram corr);

var stock;

**run**;

**proc** **arima** data=stock;

identify var=stock;

identify var=stock(**1**);

identify var=stock(**12**);

identify var=stock(**1**, **12**);

identify var=stock(**6**);

identify var=stock(**1**, **6**);

identify var=stock(**1**, **6**, **12**);

**run**;

**proc** **arima** data=stock;

identify var=stock(**1**, **6**, **12**);

estimate q=(**1**) q=(**6**) q=(**12**)noconstant printall;

**run**;

**proc** **arima** data=stock;

identify var=stock(**1**, **12**);

estimate p=(**1**) q=(**12**)noconstant printall;

**run**;

**proc** **arima** data=stock;

identify var=stock(**1**, **12**);

estimate q=(**1**) q=(**12**)noconstant printall;

forecast lead=**10** out=work.fcast;

**data** fcast2;

set work.fcast1;

forecasty=Exp(forecast);

L95CI=Exp(L95) ;

U95CI=Exp(u95) ;

**proc** **print** data= work.fcast2;

var forecasty L95CI U95CI;

**run**;

**Problem 2**

**data** sales;

SET sales;

time = \_N\_;

ystar=log(sales);

**run**;

**proc** **print** data=sales;

**run**;

**proc** **gplot** data=sales;

plot ystar \* time;

symbol1 v=star c=blue;

title "Time Series Plot";

**run**;

**quit**;

title;

**proc** **autoreg** data=sales;

model ystar = time/dwprob;

**run**;

**proc** **corr** data=sales;

var ystar time;

**run**;

**proc** **timeseries** data=sales plots=(series residual histogram corr);

var ystar;

**run**;

**proc** **arima** data=sales;

identify var=ystar;

identify var=ystar(**1**);

identify var=ystar(**12**);

identify var=ystar(**1**, **12**);

**run**;

**proc** **arima** data=sales;

identify var=ystar(**1**, **12**);

estimate q=(**1**) q=(**12**)noconstant printall;

**run**;

**proc** **arima** data=sales;

identify var=ystar(**1**, **12**);

estimate q=(**1**) q=(**12**)noconstant printall;

forecast lead=**6** out=work.fcast;

**data** fcast2;

set work.fcast1;

forecasty=Exp(forecast);

L95CI=Exp(L95) ;

U95CI=Exp(u95) ;

**proc** **print** data= work.fcast2;

var forecasty L95CI U95CI;

**run**;

**Problem 3**

**data** speed;

SET speed;

time = \_N\_;

**run**;

**proc** **print** data=speed;

**run**;

**data** anno;

length function color $**8**;

retain xsys ysys '2' when 'a';

set speed;

function='symbol';

x=time;

y=y;

size=**1.3**;

text='dot';

if time gt **60** then color='depk';

else color='vibg';

output;

**run**;

**proc** **gplot** data=speed;

plot y\*time / haxis=axis1 vaxis=axis2 href=**60.5** lvref=**20** cvref=grp annotate=anno;

symbol1 interpol=none value=none color=white;

axis1 label=("Time") offset=(**2**,**2**)pct;

axis2 label=(angle=**90** "Speed");

**run**;

**data** prespeed;

input y;

datalines;

55

56

48

46

56

46

59

60

53

58

73

69

72

51

72

69

68

69

79

77

53

63

80

65

78

64

72

77

82

77

35

79

71

73

77

76

83

73

78

91

70

88

88

85

77

63

91

94

72

83

88

78

84

78

75

75

86

79

76

87

66

;

**run**;

**proc** **arima** data=prespeed;

identify var=y;

**run**;

**data** prespeed;

SET prespeed;

ystar = log(y);

**run**;

**proc** **arima** data=prespeed;

identify var=ystar;

**run**;

**proc** **arima** data=prespeed;

identify var=ystar;

estimate p=(**1**) q=(**1**) printall;

**run**;

**data** newspeed;

input y;

datalines;

55

56

48

46

56

46

59

60

53

58

73

69

72

51

72

69

68

69

79

77

53

63

80

65

78

64

72

77

82

77

35

79

71

73

77

76

83

73

78

91

70

88

88

85

77

63

91

94

72

83

88

78

84

78

75

75

86

79

76

87

66

73

62

27

52

47

65

59

77

47

51

47

49

54

58

56

50

54

45

66

39

51

39

27

39

37

43

41

27

29

27

26

29

31

28

38

37

26

31

45

38

33

33

25

24

29

37

35

32

31

28

40

31

37

34

43

38

33

28

35

;

**run**;

**data** newspeed2;

set newspeed;

time=\_n\_;

ystar = log(y);

if time >=**60** then S=**1**;

else S=**0**;

**run**;

**data** future;

input y ystar time S;

datalines;

. . 121 1

. . 122 1

. . 123 1

. . 124 1

. . 125 1

;

**run**;

**data** newspeed3;

update newspeed2 future;

by time S;

**run**;

**proc** **arima** data=newspeed3;

identify var=ystar crosscor=(S);

estimate p=(**1**) q=(**1**) Input=S printall altparm maxit=**30** backlim= -**3** plot;

forecast lead=**5**;

**run**;

**data** newerspeed;

input y S;

datalines;

55 0

56 0

48 0

46 0

56 0

46 0

59 0

60 0

53 0

58 0

73 0

69 0

72 0

51 0

72 0

69 0

68 0

69 0

79 0

77 0

53 0

63 0

80 0

65 0

78 0

64 0

72 0

77 0

82 0

77 0

35 0

79 0

71 0

73 0

77 0

76 0

83 0

73 0

78 0

91 0

70 0

88 0

88 0

85 0

77 0

63 0

91 0

94 0

72 0

83 0

88 0

78 0

84 0

78 0

75 0

75 0

86 0

79 0

76 0

87 0

66 0.016666667

73 0.033333333

62 0.05

27 0.066666667

52 0.083333333

47 0.1

65 0.116666667

59 0.133333333

77 0.15

47 0.166666667

51 0.183333333

47 0.2

49 0.216666667

54 0.233333333

58 0.25

56 0.266666667

50 0.283333333

54 0.3

45 0.316666667

66 0.333333333

39 0.35

51 0.366666667

39 0.383333333

27 0.4

39 0.416666667

37 0.433333333

43 0.45

41 0.466666667

27 0.483333333

29 0.5

27 0.516666667

26 0.533333333

29 0.55

31 0.566666667

28 0.583333333

38 0.6

37 0.616666667

26 0.633333333

31 0.65

45 0.666666667

38 0.683333333

33 0.7

33 0.716666667

25 0.733333333

24 0.75

29 0.766666667

37 0.783333333

35 0.8

32 0.816666667

31 0.833333333

28 0.85

40 0.866666667

31 0.883333333

37 0.9

34 0.916666667

43 0.933333333

38 0.95

33 0.966666667

28 0.983333333

35 1

;

**run**;

**data** newerspeed2;

set newerspeed;

time=\_n\_;

ystar = log(y);

**run**;

**data** future2;

input y ystar time S;

datalines;

. . 121 1

. . 122 1

. . 123 1

. . 124 1

. . 125 1

;

**run**;

**data** newerspeed3;

update newerspeed2 future2;

by time S;

**run**;

**proc** **arima** data=newerspeed3;

identify var=ystar crosscor=(S);

estimate p=(**1**) q=(**1**) Input=S printall altparm maxit=**30** backlim= -**3** plot;

forecast lead=**5**;

**run**;

**data** speed;

SET speed;

ystar = log(y);

**run**;

**proc** **arima** data=speed;

identify var=ystar;

estimate p=(**1**) printall;

**run**;

**proc** **arima** data=speed;

identify var=ystar;

estimate p=(**1**,**2**) q=(**1**) printall;

**run**;

**proc** **arima** data=speed;

identify var=ystar;

estimate p=(**1**) q=(**1**) printall;

forecast lead=**5**;

**run**;