

Name Mr Key
ID# 7777777

ECO 5375
Eco. & Bus. Forecasting

Prof. Tom Fomby
Summer I, 2015

MID-TERM EXAM

Instructions: Write in your name and student ID above. You have 1 hour and 30 minutes to complete this exam. This exam is worth a total of 97 points. The points for the separate questions are broken out as follows:

Questions 1 – 17, 19 – 25, 27-28, and 35 are worth 2 points each.

Q18 = 5 points

Q26 = (2, 2, 2) = 6 points

Q29 = 4 points

Q30 = 4 points

Q31 = 4 points

Q32 = 4 points

Q33 = (2, 6) = 8 points

Q34 = (4, 4) = 8 points

SHORT ANSWER AND MULTIPLE CHOICE:

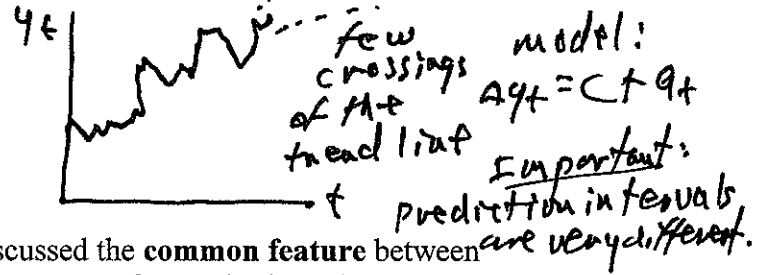
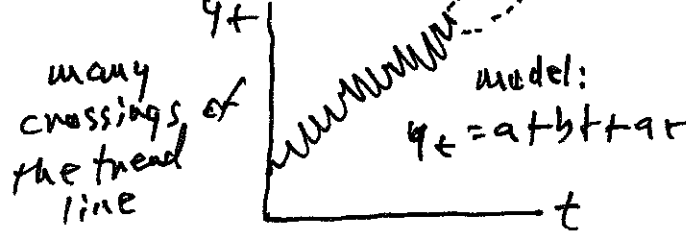
1. What are the 4 components of the Additive Time Series Decomposition and why is it important for us to know about them? $Y = T + S + C + I$ Reason: You cannot build an accurate forecasting model unless you know what components are contained in the time series

$Y =$ series to be decomposed
 $T =$ trend $C =$ cycle
 $S =$ season $I =$ irregular

2. In two separate plots below, plot a deterministically trending time series in one plot and a stochastic trend time series in another plot. Clearly label your plots. Why is it important that we distinguish between these two types of trend? How would you model the time series in the deterministic time trend case? The stochastic trend case?

deterministically trending

stochastic trending



3. In Chapter 2 of your textbook the authors discussed the common feature between gold and silver prices. What is the nature of the common feature in the gold and silver prices and why is it important to note common features in this set of time series?

If time series like silver and gold prices show strong comovement tendencies then modeling the two time series together may provide a bivariate model that forecasts these series more accurately than if we built separate model for each of the time series.

4. SAS programs have two basic steps. They are the Data step and the Proc step.

5. The basic punctuation following each executable statement in SAS is

- a. Quotation mark (“”)
- b. Period (“.”)
- c. Semicolon (“;”)
- d. Dash (“-”)

6. To put comments in a SAS program

- a. Enclose the comments between quotation marks as in: “content”
- b. Enclose the comments between /* and */ as in /*content*/
- c. Enclose the comments between (and) as in (content)
- d. Enclose the comments between # and # as in #content#

7. If a time series y_t exhibits an exponential growth pattern, it can be transformed to stationarity by the transformation

- 2
- a. Δy_t
 - b. $\Delta \ln(y_t)$
 - c. $\exp(y_t)$
 - d. $\tan^{-1}(y_t)$

8. To evaluate whether a proposed forecasting model is a very good one relative to its competitors, veteran forecasters traditionally examine two things:

- 2
- a. Volatility and Statistical Significance of forecasts
 - b. Standard Errors of Forecasts and Widths of Turning Points
 - c. Accuracies of Point Forecasts and Interval Forecasts
 - d. Standard Error of Regression and Autocorrelation of Errors

9. For forecasting experiments we usually partition our time series into two parts:

2

These parts are in-sample data set and out-of-sample data set.

10. When choosing an exponential forecasting methods (at least the way Fomby teaches it) the only two things we need to know is whether or not there is trend in the data and whether or not there is seasonality in the data.

11. Which of the following is a useful tool for detecting whether or not there is seasonality in the time series data we are examining (1 pt.):

- 2
- a. Cross Correlation Plot
 - b. Value at Risk Plot
 - c. Autocorrelation Plot
 - d. Buys-Ballot Plot

12. In chapter 2 of your textbook the authors discussed **Key Features** of time series. In particular they looked at **one or more** of the following:

- 2
- a. Strong versus weak months in US industrial production
 - b. Changing seasonality in US industrial production
 - c. Changing trend in US industrial production
 - d. Changing cycle in US industrial production

13. One of the important assumptions of the Stable Seasonal Pattern (SSP) Time Series model is:

- 2
- a. The number of weak months equal the number of strong months
 - b. The trend in the time series being modeled is constant
 - c. The cycle in the time series being modeled is constant
 - d. The monthly proportions within the year are stable over time.

14. When we used the SSP model for analyzing Fomby Inc. sales we obtained next year's total by (1 pt.)
- Fitting a time trend to the yearly totals
 - Fitting an exponential trend to the yearly totals
 - Using the mean of the differences in yearly totals
 - Using the mean of the differences in the yearly totals relevant to where we are in the business cycle

Consider the following Deterministic Trend/Deterministic Season model for quarterly data:

$$\hat{y}_t = 5 + 10t - 3D_{t2} + 4D_{t3} + 6D_{t4}$$

15. If we had also added a seasonal dummy variable for the first quarter we would have fallen into the so-called dummy variable trap.

16. The trend line for all second quarter data across all years is given by the formula

$$y_t = 5 - 3 + 10t = 2 + 10t$$

The trend line for all fourth quarter data across all years is given by the formula

$$y_t = 5 + 6 + 10t = 11 + 10t$$

17. Let $y_t = \ln(RGDP_t)$ where $RGDP_t$ is annual real gross domestic product in the US. Consider the time series model $\Delta y_t = \mu_1 + \theta D_t + \varepsilon_t$, where $D_t = 1$ when the US economy is in a recession and 0 otherwise. Suppose we estimated this model using Proc Autoreg and got $\hat{\mu}_1 = 0.04$ and $\hat{\theta} = -0.02$. In a sentence or two explain to me the meaning of this model.

This is a two-regime time series model. According to this model the average growth during an expansion is 4% while in a recession it is 2%.

18. Consider the Optimal Inventory Model analyzed in Exercise 3: $R = D_L + \sigma_L \Phi^{-1}(p)$. Define these terms

R = Reorder point
 D_L = Lead Demand
 σ_L = Standard Deviation of Lead Demand
 p = service level (probability of not stocking out)
 $\sigma_L \Phi^{-1}(p)$ = safety stock

19. Briefly explain to me the purpose of this equation: $p = \Phi \left[\sqrt{2 \ln \left(\frac{1}{\sqrt{2\pi}} \frac{M}{H} \right)} \right]$.

This formula allows us to determine an optimal service level p , based upon the trade off between the cost of stocking out (M) and the cost of holding excess inventory (H).

DW: H_0 : no autocorrelation of order 1 in errors of linear regression
 H_1 : there is autocorrelation present in the errors

20. When we first estimate the DTDS model using PROC REG (recall the Plano Sales Tax Revenue data), we examine the DW statistic. What is the purpose of examining this statistic and what is it testing?

The examination of the DW statistic allows us to decide whether we need to use OLS (Proc Reg) or GLS (Proc Autoreg), in testing the various parts of the DTDS model (including covariance structure and the presence or absence of seasonality in the data).

21. True or False. If some time series data, say x_t , data needs to be differenced to be made stationary, then the identify statement in the SAS Procedure ARIMA should read "identify var = x;"

E + should be "identify var = x(1);"

22. Consider the following AR(1) Box-Jenkins model: $y_t = 20 + 0.8y_{t-1} + a_t$. This model implies that the mean of the y_t series is 100. Suppose the last observation you have on the series is 95. The one-period ahead forecast for this model should be

96.
$$\mu = \frac{20}{1-0.8} = \frac{20}{.2} = 100 \quad \hat{y}_{t+1} = 20 + 0.8(95) = 96$$

23. True or False. The "Damping" and "Cutting Off" patterns of the ACF and PACF of the stationary form of a time series provide a way to identify the orders of pure Box-Jenkins processes.

24. If the ACF has 2 spikes in it and then cuts off and if the PACF tails off, the ARMA model that is appropriate for the data is ARMA(0, 2).

25. Consider the model: $y_t = y_{t-1} + a_t$. The stationary form of this model is

$Ay_t = a_t$

OTHER QUESTIONS:

26. Suppose you have applied ordinary least squares model to a DTDS model that produces the following table.

OLS regression to get the DW statistic

The REG Procedure	
Model: MODEL1	
Dependent Variable: SALES	
Durbin-Watson D	0.432
Pr < DW	<.0001
Pr > DW	1.0000
Number of Observations	312
1st Order Autocorrelation	0.783

(a) What are the null and alternative hypotheses of this test?

②

H_0 : No autocorrelation of order 1 in errors of regression
 H_1 : Autocorrelation is present in the errors of regression

(b) Given the above result, what is the conclusion of this test?

②

As $Pr < DW$ is $< .0001$ we reject H_0 and accept H_1 that positive autocorrelation is present in the errors. We should use GLS for estimation and inference.

(c) In further testing of the DTDS model that you are investigating, would you be inclined to use Ordinary Least Squares Output (Proc Reg) or Generalized Least Squares Output (Proc Autoreg)? Explain your answer.

②

Use GLS because the errors of the regression are autocorrelated.

27. Suppose you have been given the output displayed in Computer Output # 1. Does it appear that the DTDS model has autocorrelated errors? Explain your reasoning.

②

Yes, because the AR1, AR4, and AR12 coefficients are statistically significant at the 5% level.

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	1	108160	28166	3.84	0.0002
t	1	2197	380.0081	5.78	<.0001
t2	1	-2.3581	1.1592	-2.03	0.0428
d2	1	-13603	5248	-2.59	0.0100
d3	1	26528	6700	3.96	<.0001
d4	1	-19559	7308	-2.68	0.0079
d5	1	-27417	7317	-3.75	0.0002
d6	1	23676	7454	3.18	0.0016
d7	1	49498	7512	6.59	<.0001
d8	1	68258	7465	9.14	<.0001
d9	1	-38426	7336	-5.24	<.0001
d10	1	-22235	7331	-3.03	0.0026
d11	1	-27777	6744	-4.12	<.0001

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > t
d12	1	11731	5325	2.20	0.0284
AR1	1	-0.5884	0.0444	-13.24	<.0001
AR4	1	-0.1216	0.0429	-2.83	0.0049
AR12	1	-0.2234	0.0415	-5.38	<.0001

28. Given Computer Output # 1, does it appear that the trend in the DTDS model has curvature? Explain your answer.

2

Yes. The GLS determined coefficient for the t² variable is statistically significant at the 5% level.

29. Given Computer Output # 2 (which was produced by PROC AUTOREG) that accompanies Computer Output # 1, does it look like there is significant seasonality in the data? Explain your answer. What are the null and alternative hypotheses of the test result reported in this output?

4

Yes. The null hypothesis is that the coefficients on the seasonal dummies are all zero (i.e. no seasonality) while the alternative hypothesis is that one or more of these coefficients are statistically significant (and hence that there is seasonality in the data).

Since the p-value of the test statistic is less than 0.05 we conclude that there is seasonality in the data.

		Test			
Source	DF	Mean Square	F Value	Pr > F	
Numerator	11	29317586854	24.37	<.0001	
Denominator	298	1202946247			

are statistically significant (and hence that there is seasonality in the data).

30. Consider Computer Output # 3 below. Suppose that the data so analyzed has the following output. Which months of the year are weak? Which months are strong? Which is the weakest month? Which is the strongest month? Thoroughly explain your answer on the next page.

4

strongest month = Aug. (biggest positive)
weakest month = Sept (most negative)

Computer Output # 3

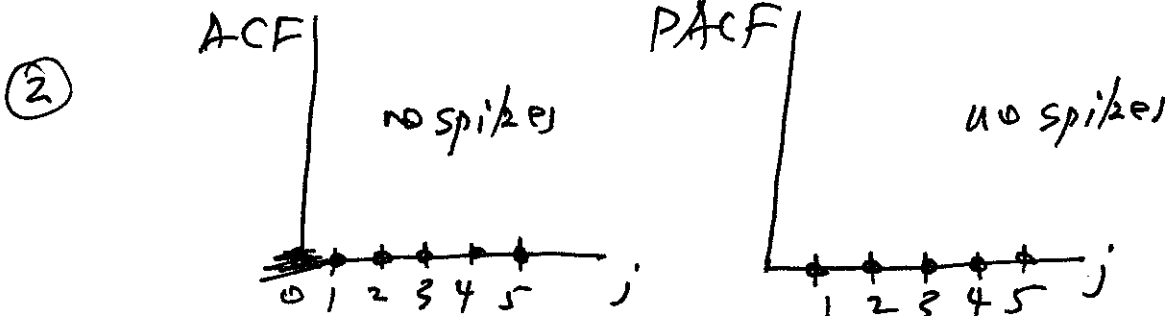
Obs	sum	d1a	d2a	d3a	d4a	d5a	d6a
1	1.97E-15	-0.0231	-0.146	0.21651	-0.1997	-0.2707	0.19076
				3			4
d7a	d8a	d9a	d10a	d11a	d12a		
0.42398	0.59343	-0.37016	-0.22391	-0.27397	0.082865		
2	1	1			5		

"weak months" = {Jan, Feb, Apr, May, Sept., Oct., Nov.}
"strong months" = {Mar, Jun, Jul, Aug, Dec}

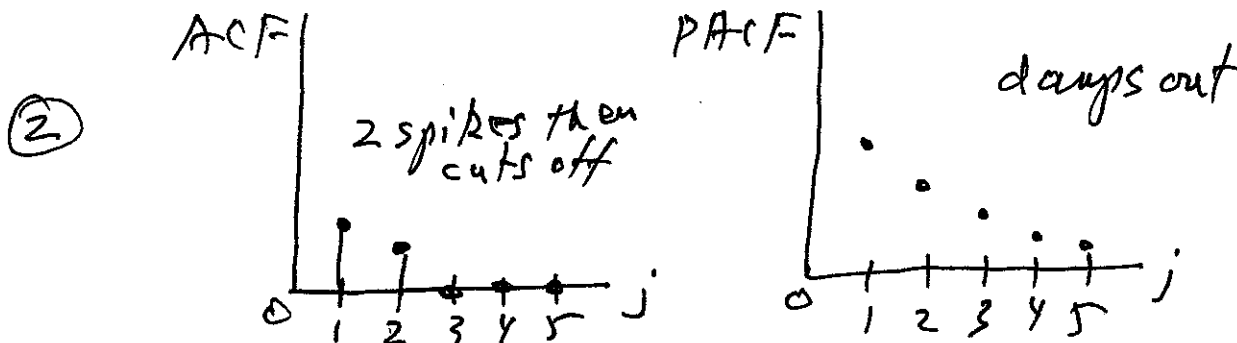
Answer for Question 30:

We look at the magnitudes of the positive and negative coefficients and rank order them accordingly.

31. (a) In the below space I want you to draw the ACF and PACF of the ARMA(0,0) model. On one graph you should have "ACF" on the y-axis and J on the x-axis. On the other graph you should have "PACF" on the y-axis and J on the x-axis.



(b) In a similar manner draw me "plausible" ACF and PACF functions for the MA(2) (equivalently ARMA(0,2)) model in the below space.



Now consider the SAS Computer Output # 4 that is provided as an insert to this exam. We are analyzing the monthly sales of Goofy Burgers. Use this output to answer the following 4 questions. We wish to determine the best Box-Jenkins model for this data.

32. Using the sample ACF and sample PACF that is provided by Computer Output # 4, give me a tentative identification of the p, and q values for the series. Explain your answer.

② $p = 0, q = 1.$

② Explanation: The ACF has one spike and then cuts off while the PACF damps out. This is indicative of a MA(1) process

33. Using **Computer Output # 4**, fill in the following P-Q box. Be sure to tell me what the entries of the cells of your box are. Which model is indicated to be the best model in the P-Q box? Explain your reasoning.

(2)

Reasoning: The MA(1) model is the best. It has the lowest AIC and SBC measures and its residuals are white noise.

Legend:
AIC
SBC
Q₂₄
(p-value)

(6)

		Q		
		0	1	2
P	0	717.0552 720.3535 67.13 (1.0001)	<u>601.7556</u> <u>608.3522</u> 15.81 (0.8631)	603.7219 613.6169 15.83 (0.8243)
	1	672.7702 679.3758 58.35 (1.0001)	603.7278 613.6228 15.83 (0.8244)	
	2	640.9584 650.8533 33.37 (0.0569)		

34. Use **Computer Output #4** to conduct an overfitting exercise on the model you chose from the P-Q box. Below, report the overfitting coefficient of each overfitting model and its corresponding p-value. What conclusion do you draw from the overfitting exercise? Explain your answer.

Overfitting Model 1 is ARMA(0, 2).

The overfitting coefficient estimate is 0.01527.

The p-value of the overfitting coefficient estimate is 0.8302.

Therefore the overfitting coefficient from this model is statistically (significant/insignificant). Circle one alternative.

Overfitting Model 2 is ARMA(1, 1).

The overfitting coefficient estimate is -0.01325.

The p-value of the overfitting coefficient estimate is 0.8676.

Therefore the overfitting coefficient from this model is statistically (significant/insignificant). Circle one alternative.

35. My conclusion is that the best model for Goofy Burgers is

MA(1)

because besides the best P-Q box numbers, the two overfitting parameters above are statistically insignificant and we make the MA(1) model the final choice.

For your information only:

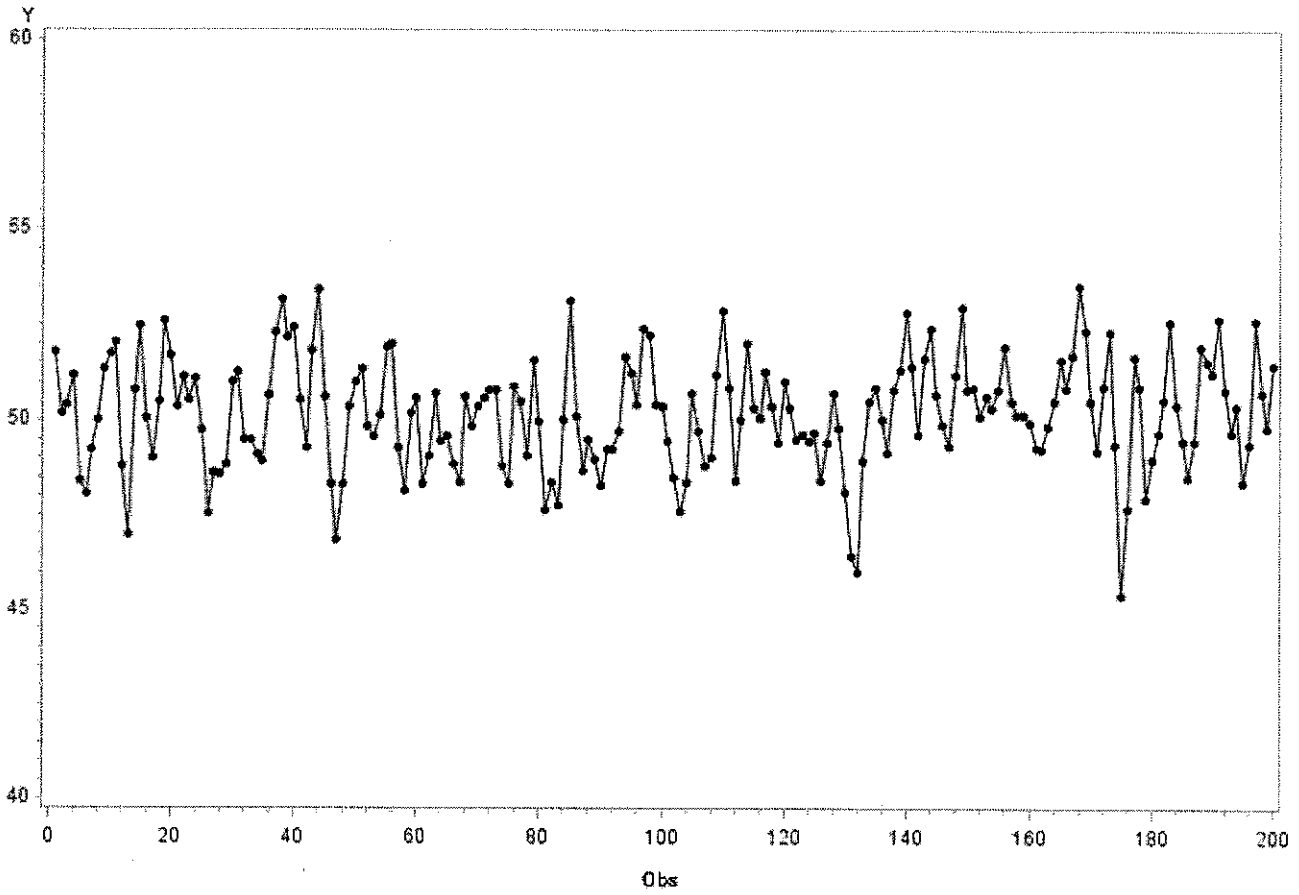
intercept-form: $\hat{y}_t = 50.21 + a_t - (-0.90561)a_{t-1}$
(0.144) (0.03096)

Deviation from mean form:
 $\hat{y}_t - 50.21 = a_t - (-0.90561)a_{t-1}$

Computer Output #4

Sales of Goofy Burgers

X=Time Y=Y

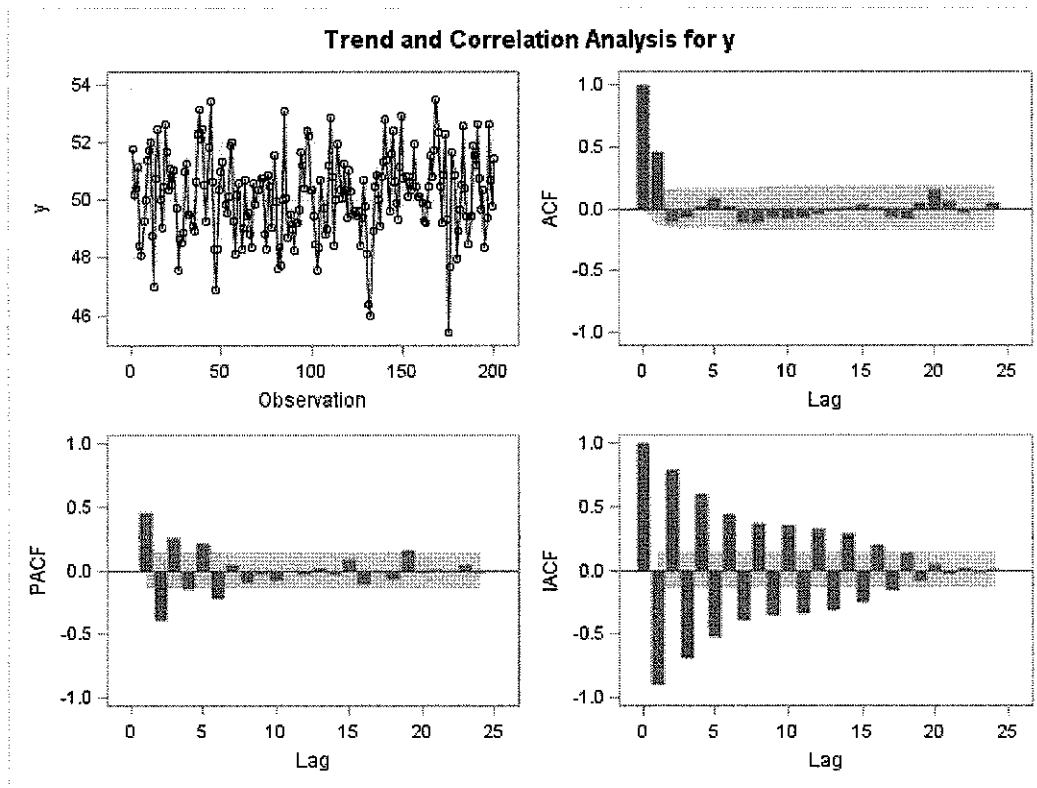


Building the P-Q Box for the Goofy Burgers Time Series

The ARIMA Procedure

Name of Variable = y	
Mean of Working Series	50.16262
Standard Deviation	1.445853
Number of Observations	200

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	47.01	6	<.0001	0.454	-0.105	-0.077	0.020	0.089	0.018
12	56.43	12	<.0001	-0.117	-0.105	-0.070	-0.084	-0.072	-0.051
18	59.20	18	<.0001	-0.017	0.003	0.030	0.011	-0.068	-0.081
24	67.13	24	<.0001	0.054	0.157	0.067	-0.033	-0.006	0.046



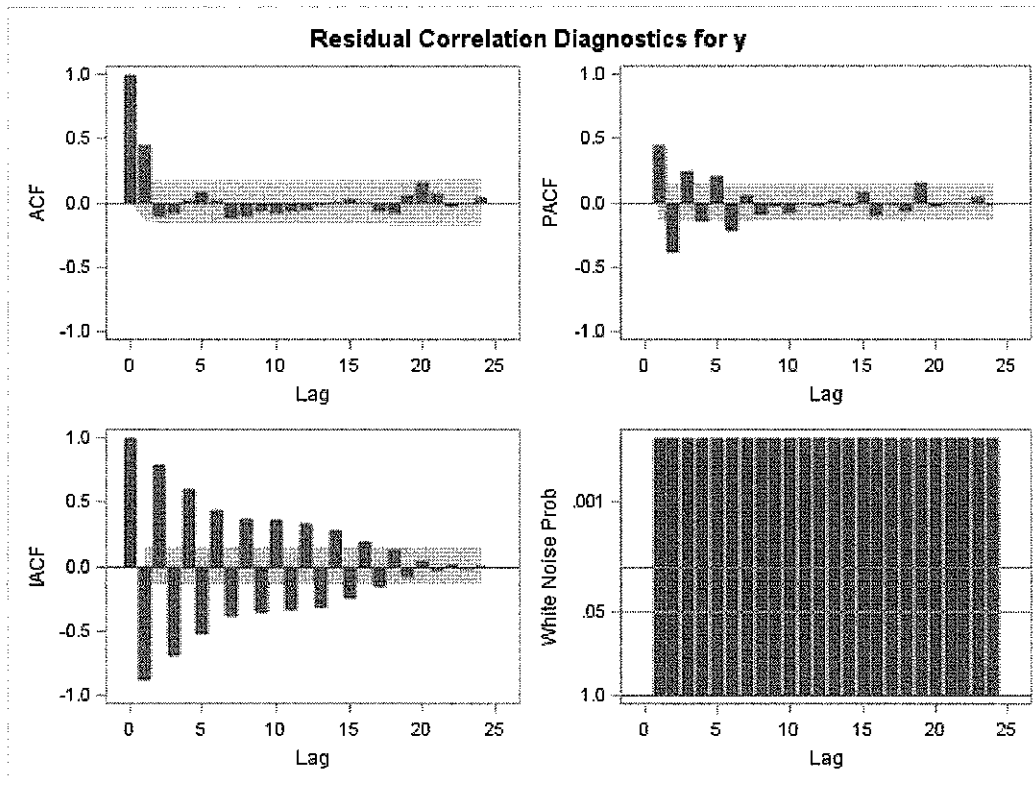
Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	50.16262	0.10249	489.42	<.0001	0

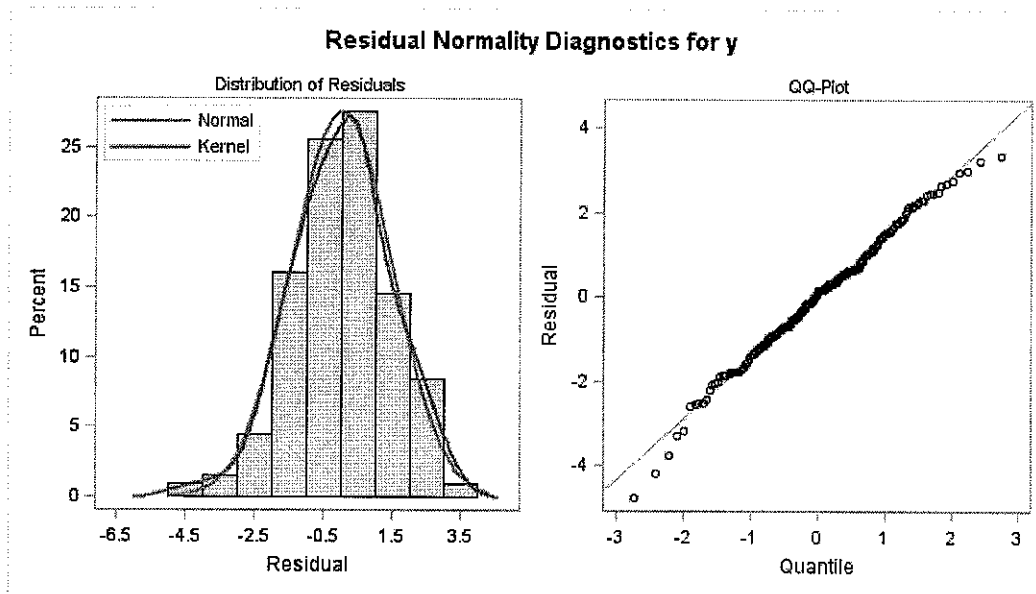
Constant Estimate	50.16262
Variance Estimate	2.100996
Std Error Estimate	1.449481
AIC	717.0552

SBC	720.3535
Number of Residuals	200

* AIC and SBC do not include log determinant.

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	47.01	6	<.0001	0.454	-0.105	-0.077	0.020	0.089	0.018
12	56.43	12	<.0001	-0.117	-0.105	-0.070	-0.084	-0.072	-0.051
18	59.20	18	<.0001	-0.017	0.003	0.030	0.011	-0.068	-0.081
24	67.13	24	<.0001	0.054	0.157	0.067	-0.033	-0.006	0.046
30	81.62	30	<.0001	0.038	-0.049	-0.117	0.038	0.199	0.053
36	92.33	36	<.0001	-0.117	-0.082	-0.005	0.054	0.008	-0.144





Model for variable y	
Estimated Mean	50.16262

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	50.18020	0.16708	300.34	<.0001	0
AR1,1	0.45547	0.06343	7.18	<.0001	1

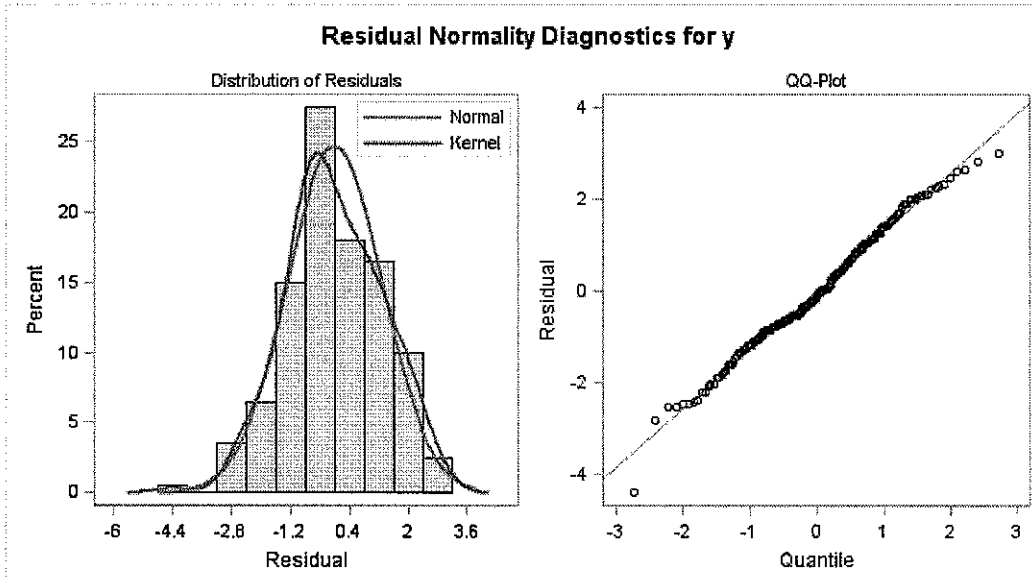
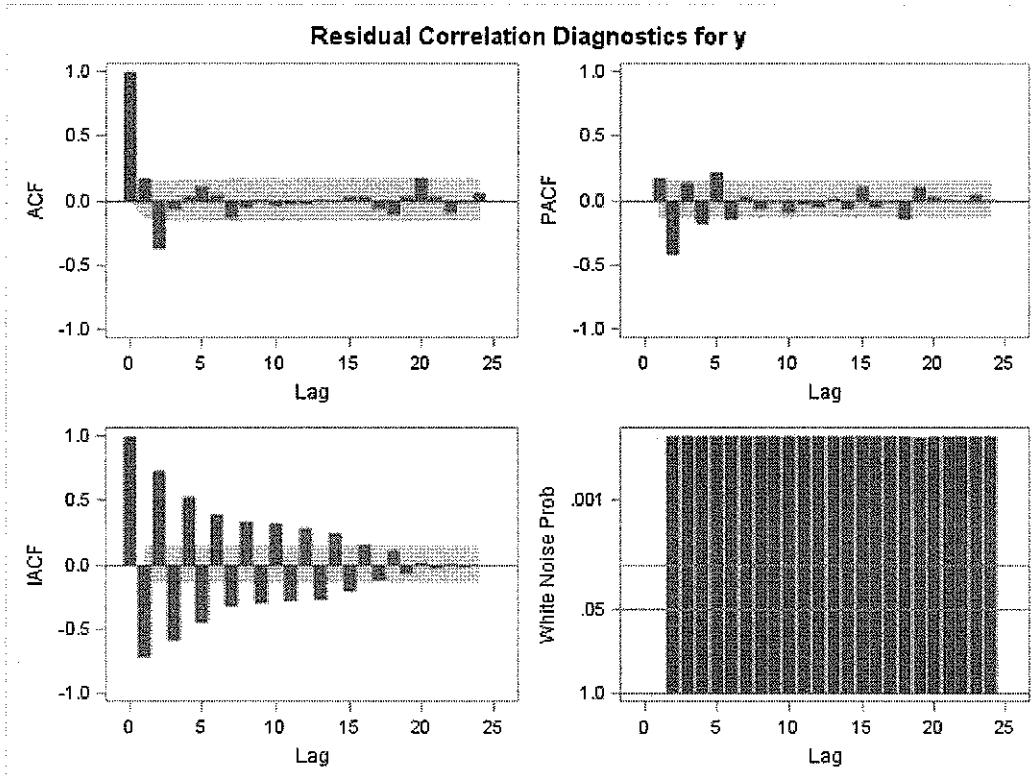
Constant Estimate	27.32461
Variance Estimate	1.675429
Std Error Estimate	1.294384
AIC	672.7792
SBC	679.3758
Number of Residuals	200

* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates		
Parameter	MU	AR1,1
MU	1.000	0.016
AR1,1	0.016	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	39.66	5	<.0001	0.179	-0.377	-0.069	0.028	0.114	0.040
12	44.43	11	<.0001	-0.128	-0.054	0.001	-0.043	-0.031	-0.025
18	48.68	17	<.0001	0.005	-0.002	0.035	0.038	-0.065	-0.111
24	58.35	23	<.0001	0.039	0.171	0.030	-0.089	-0.015	0.056

30	82.02	29	<.0001	0.060	-0.037	-0.175	0.014	0.253	0.032
36	94.45	35	<.0001	-0.163	-0.054	0.014	0.082	0.063	-0.105



Model for variable y	
Estimated Mean	50.1802

Autoregressive Factors	
Factor 1:	1 - 0.45547 B**(1)

Conditional Least Squares Estimation

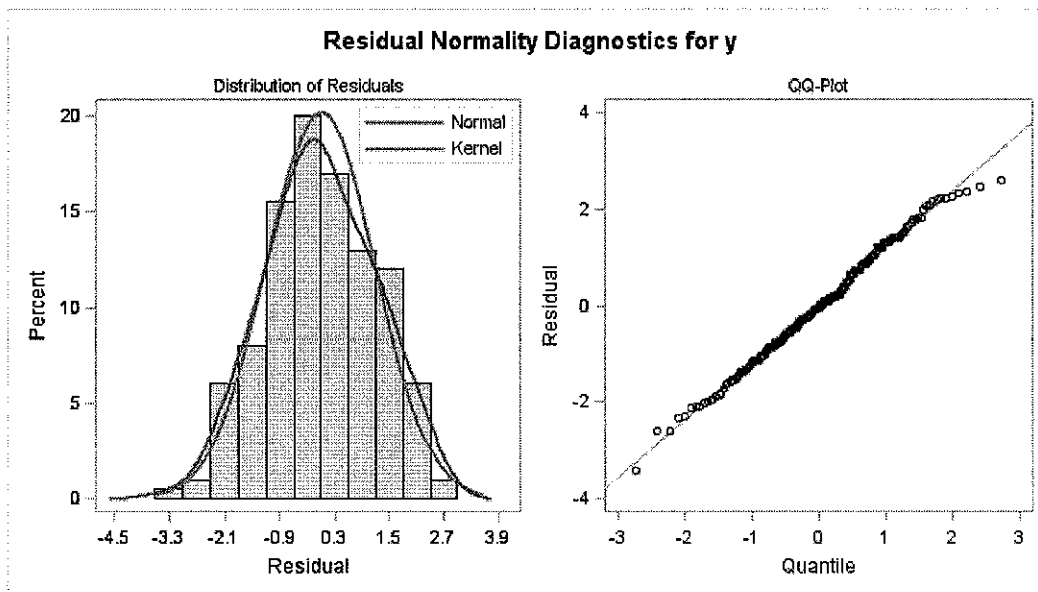
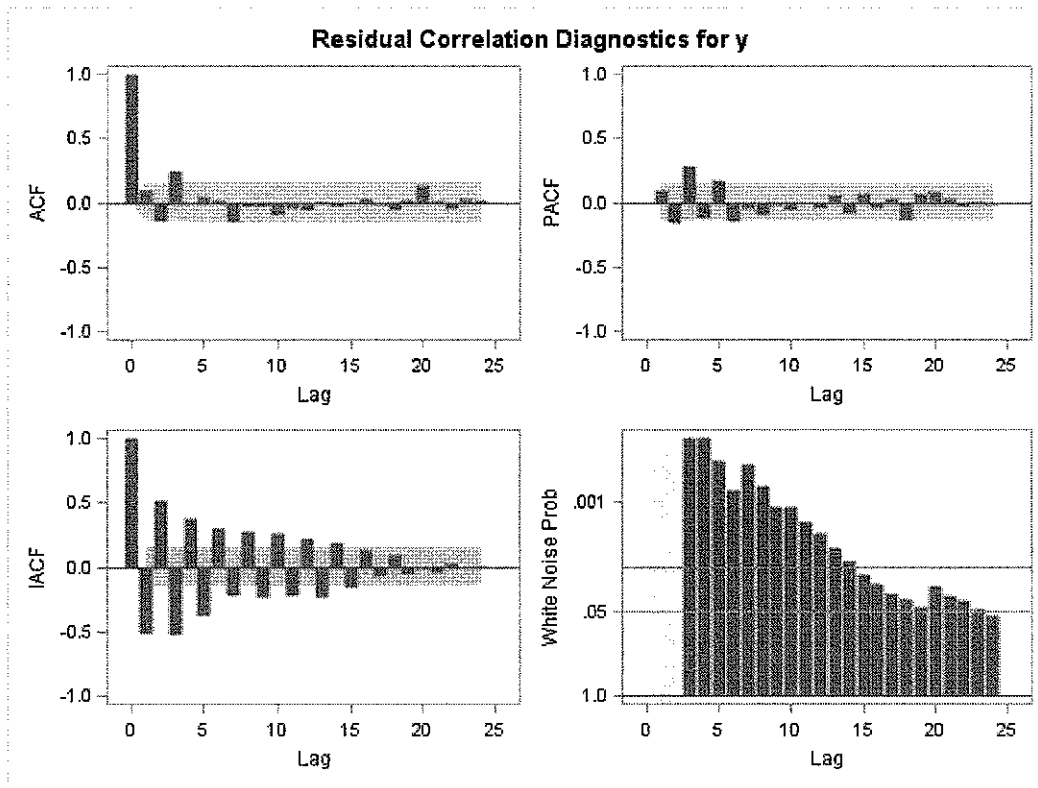
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	50.17248	0.11101	451.98	<.0001	0
AR1,1	0.63609	0.06568	9.68	<.0001	1
AR1,2	-0.39577	0.06569	-6.02	<.0001	2

Constant Estimate	38.11474
Variance Estimate	1.421949
Std Error Estimate	1.192455
AIC	640.9584
SBC	650.8533
Number of Residuals	200

* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates			
Parameter	MU	AR1,1	AR1,2
MU	1.000	0.011	0.004
AR1,1	0.011	1.000	-0.457
AR1,2	0.004	-0.457	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	19.45	4	0.0006	0.101	-0.143	0.249	-0.011	0.040	0.021
12	26.65	10	0.0030	-0.140	-0.037	-0.026	-0.090	-0.042	-0.050
18	28.10	16	0.0308	0.004	-0.035	-0.004	0.033	-0.021	-0.061
24	33.37	22	0.0569	0.020	0.138	0.022	-0.046	0.032	0.021
30	47.19	28	0.0131	0.020	0.028	-0.118	-0.010	0.208	-0.015
36	54.30	34	0.0150	-0.118	0.035	-0.030	-0.011	0.012	-0.114



Model for variable y	
Estimated Mean	50.17248

Autoregressive Factors	
Factor 1:	$1 - 0.63609 B^{**}(1) + 0.39577 B^{**}(2)$

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	50.21403	0.14427	348.05	<.0001	0

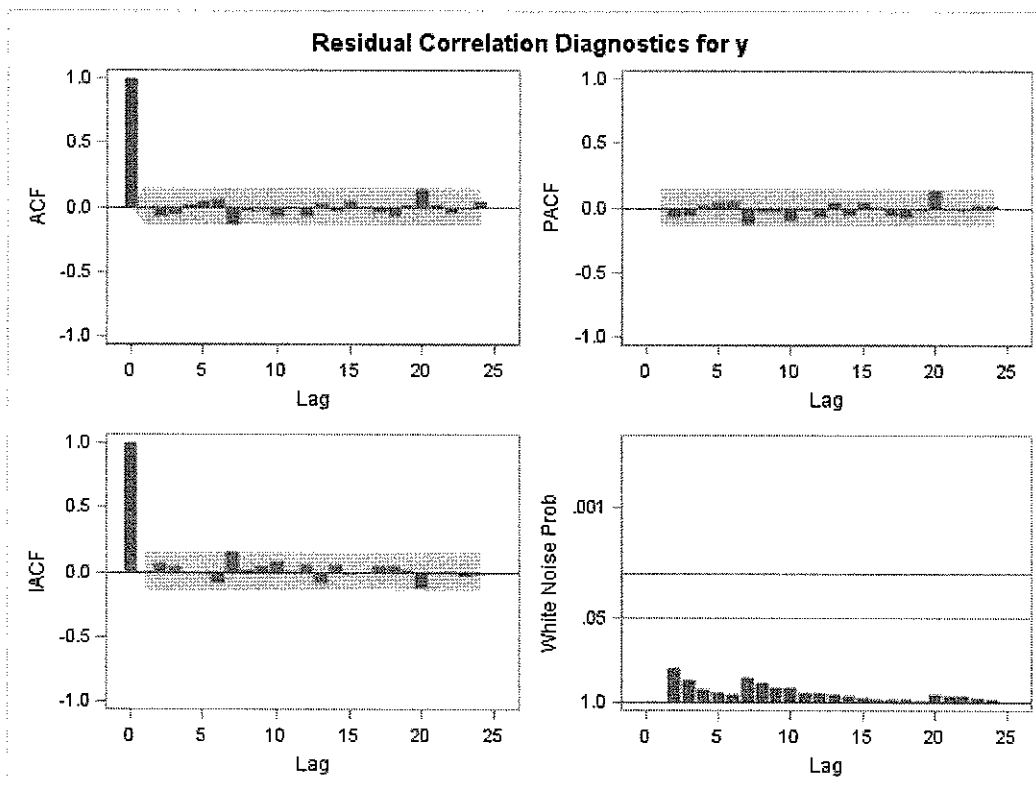
MA1,1	-0.90561	0.03096	-29.25	<.0001	1
--------------	----------	---------	--------	--------	---

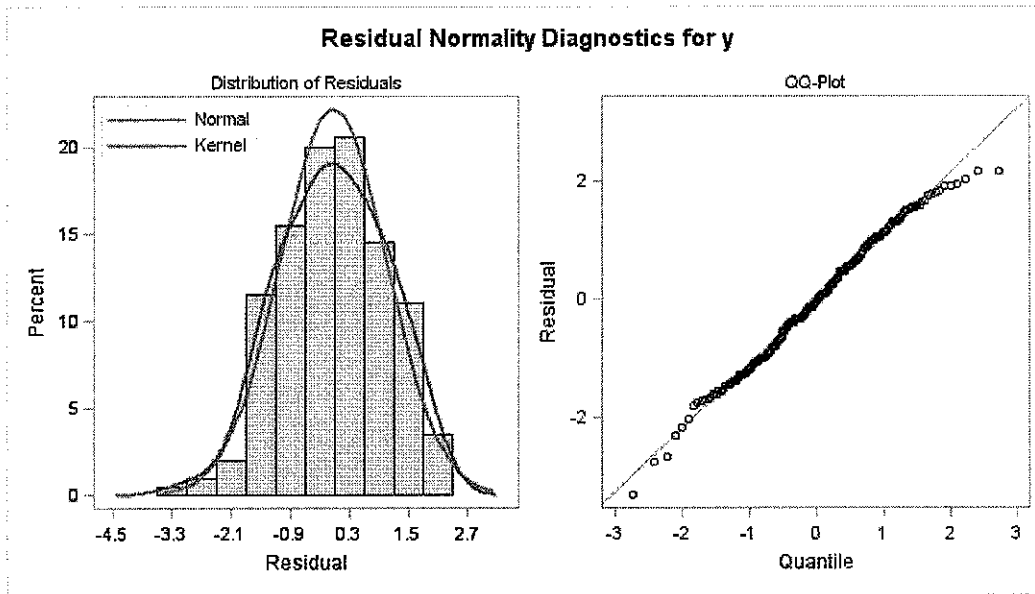
Constant Estimate	50.21403
Variance Estimate	1.174628
Std Error Estimate	1.083802
AIC	601.7556
SBC	608.3522
Number of Residuals	200

* AIC and SBC do not include log determinant.

Parameter	MU	MA1,1
MU	1.000	-0.053
MA1,1	-0.053	1.000

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.86	5	0.7213	-0.010	-0.072	-0.050	0.025	0.049	0.055
12	8.32	11	0.6845	-0.126	-0.033	-0.017	-0.069	-0.002	-0.063
18	10.51	17	0.8810	0.035	-0.034	0.041	0.010	-0.038	-0.066
24	15.81	23	0.8631	0.019	0.139	0.017	-0.039	-0.005	0.044
30	27.36	29	0.5522	0.016	-0.002	-0.119	0.008	0.185	0.019
36	34.43	35	0.4956	-0.119	-0.010	-0.016	0.045	0.040	-0.105





Model for variable y	
Estimated Mean	50.21403

Moving Average Factors	
Factor 1:	$1 + 0.90561 B^{**}(1)$

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	50.21235	0.14255	352.25	<.0001	0
MA1,1	-0.89212	0.07133	-12.51	<.0001	1
MA1,2	0.01527	0.07112	0.21	0.8302	2

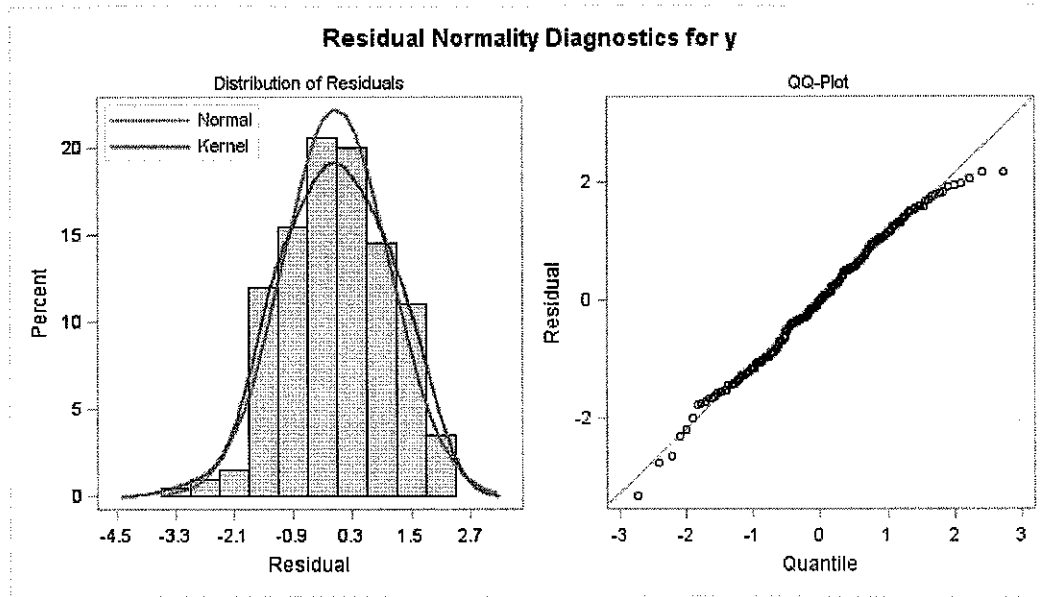
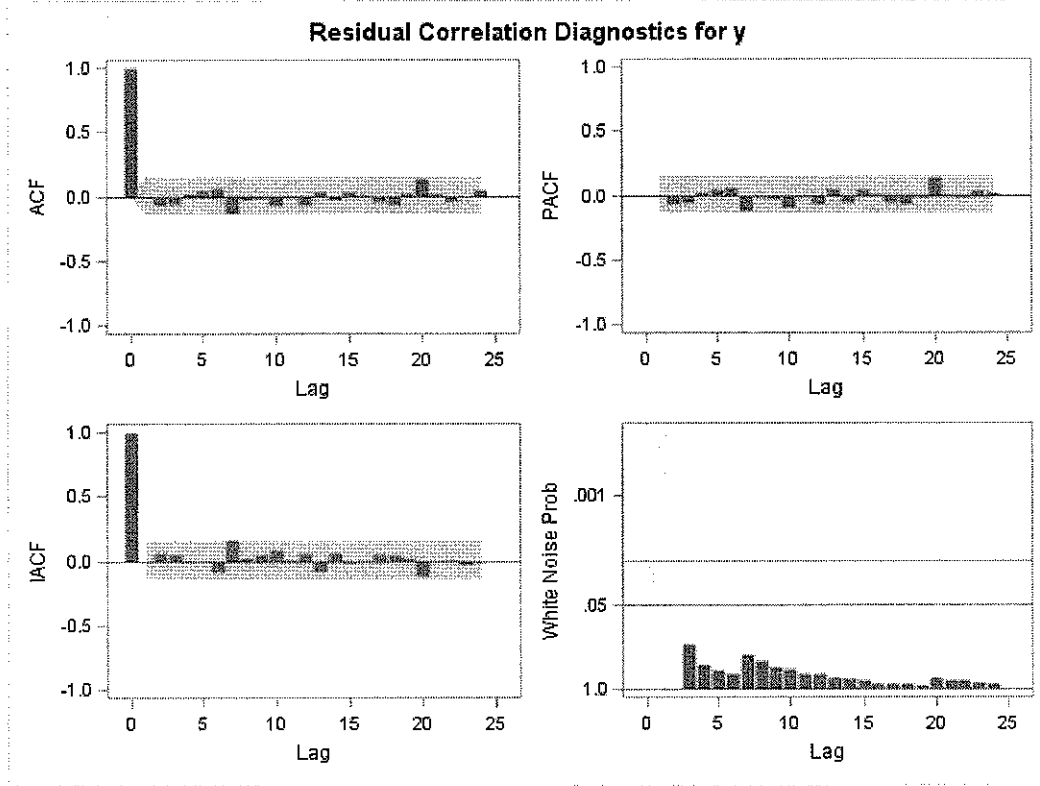
Constant Estimate	50.21235
Variance Estimate	1.180391
Std Error Estimate	1.086458
AIC	603.7219
SBC	613.6169
Number of Residuals	200

* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates			
Parameter	MU	MA1,1	MA1,2
MU	1.000	-0.047	-0.026
MA1,1	-0.047	1.000	0.901
MA1,2	-0.026	0.901	1.000

Autocorrelation Check of Residuals			

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.86	4	0.5821	0.002	-0.070	-0.054	0.027	0.047	0.056
12	8.35	10	0.5945	-0.127	-0.033	-0.020	-0.068	-0.006	-0.062
18	10.48	16	0.8404	0.033	-0.032	0.040	0.011	-0.039	-0.066
24	15.83	22	0.8243	0.020	0.140	0.018	-0.038	-0.005	0.044
30	27.35	28	0.4995	0.016	-0.003	-0.118	0.009	0.185	0.020
36	34.44	34	0.4465	-0.118	-0.013	-0.015	0.045	0.039	-0.107



Model for variable y

Estimated Mean 50.21235

Moving Average Factors	
Factor 1:	1 + 0.89212 B**(1) - 0.01527 B**(2)

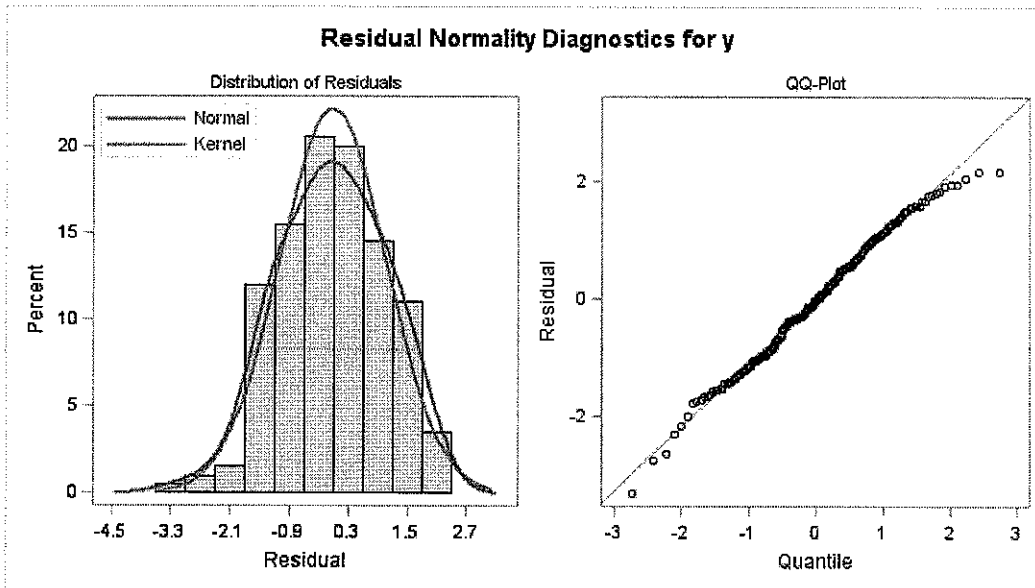
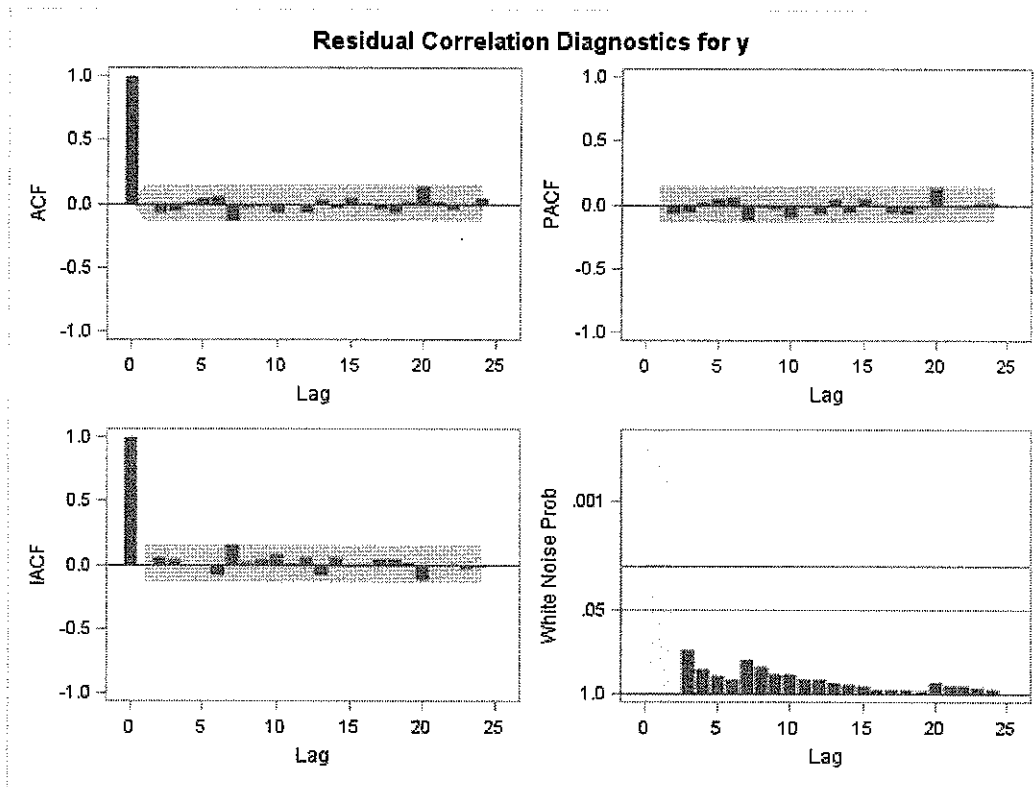
Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	50.21278	0.14303	351.06	<.0001	0
MA1,1	-0.90829	0.03406	-26.67	<.0001	1
AR1,1	-0.01325	0.07935	-0.17	0.8676	1

Constant Estimate	50.87798
Variance Estimate	1.180426
Std Error Estimate	1.086474
AIC	603.7278
SBC	613.6228
Number of Residuals	200

* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates			
Parameter	MU	MA1,1	AR1,1
MU	1.000	-0.036	0.026
MA1,1	-0.036	1.000	0.437
AR1,1	0.026	0.437	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.86	4	0.5817	-0.001	-0.071	-0.053	0.027	0.048	0.056
12	8.34	10	0.5953	-0.127	-0.033	-0.019	-0.068	-0.005	-0.062
18	10.48	16	0.8401	0.033	-0.033	0.040	0.011	-0.038	-0.066
24	15.83	22	0.8244	0.020	0.140	0.018	-0.038	-0.005	0.044
30	27.35	28	0.4990	0.016	-0.003	-0.119	0.009	0.185	0.020
36	34.45	34	0.4464	-0.118	-0.012	-0.016	0.045	0.040	-0.106



Model for variable y	
Estimated Mean	50.21278
Autoregressive Factors	
Factor 1:	1 + 0.01325 B**(1)
Moving Average Factors	
Factor 1:	1 + 0.90829 B**(1)