Name Mr Key
ID# 7777777

ECO 5375
Eco. & Bus. Forecasting

Prof. Tom Fomby
Fall 2011

MID-TERM EXAM

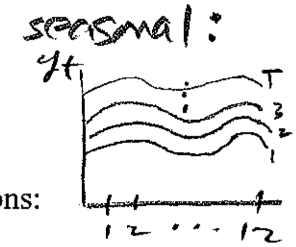
Instructions: Fill in your name and student ID above. You have 1 hour and 20 minutes to complete this exam. This exam is worth a total of 96 points. I am going to give you three bonus points if you can answer the Aggie joke below. The points for the "regular" questions are broken out as follows:

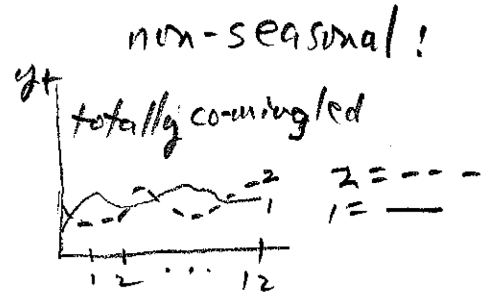
Q1 = Each part counts 2 points for a total of 26 points.
Q2 = (2, 4, 2) = 8 points.
Q3 = (1, 1, 1, 1, 1) = 5 points.
Q4 = (2, 2) = 4 points.
Q5 = (2, 2) = 4 points.
Q6 = 6 points.
Q7 = (4, 12, 2, 10, 4) = 32 points.
Q8 = (2, 4, 5) = 11 points.

Bonus Question worth 3 points:

Aggies think that the English Channel is the TV channel that shows
BR ITISH movies.

+3





1. Short Answer Questions:

   a. Buys-Ballot Plot of Season-by-Year. What distinguishes a Buys-Ballot plot of seasonal data from a Buys-Ballot plot of non-seasonal data?

   Seasonal data in B-B plots have similar undulations (pikes and troughs) from one year to the next while non-seasonal data in B-B plots appear to have no systematic pattern across years.

   b. What are the major components of the "Additive Time Series Decomposition"?

   $y = T + C + S + I$

   T = trend, C = cycle, S = season, I = irregular

   c. One of the important assumptions of the **Stable Seasonal Pattern Model** is that the monthly proportions making up the yearly totals are <u>stable</u>. (i.e. not changing over time)

   d. What are the null and alternative hypotheses of Friedman's non-parametric rank sum test? If the test statistic had a probability value of 0.43, what would your conclusion be?

   H_0: Data are not seasonal
   H_1: Data are seasonal

   Since the probability value is greater than 0.05 we accept the null hypothesis of <u>no</u> seasonality.

   e. Briefly define what we mean by a stationary time series. Why is the concept important for the Box-Jenkins methodology? All B-J models assume that the data being analyzed are stationary. A time series that has a constant mean, constant variance, and constant covariance is <u>stationary</u>.

   f. Consider the AR(1) model $y_t = \emptyset_0 + \emptyset_1 y_{t-1} + a_t$. What values of $\emptyset_1$ guarantee that the model is stationary?

   $-1 < \phi_1 < 1$

   g. What do you get when you apply the operator Δ_{12} to y_t? That is $\Delta_{12} y_t =$ <u>$y_t - y_{t-12}$</u>.

   h. True or (False). The Random Walk without drift model has a constant mean, a constant variance, and a constant covariance.

   i. (True) or False. The Augmented Dickey Fuller tests for a unit root are broken into three cases: The zero mean case, the non-zero mean case, and the trend case.

   j. (True) or False. A good Box-Jenkins model minimizes its goodness-of-fit measures, has white noise residuals, has statistically significant coefficients, and when considering the overfitting models, the overfitting parameters are both statistically insignificant.

k. Consider the following %logtest output. What does this output tell us?

That a log transformation of the data is <u>not</u> needed!

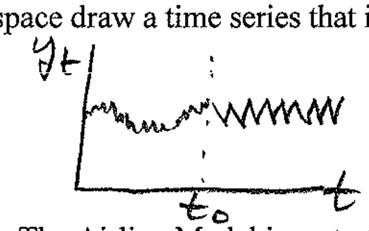
(2)

TRANS	LOGLIK	RMSE	AIC	SBC
NONE	-126.412	0.75214	264.823	280.394
LOG	-140.353	0.78846	292.705	308.276

← smaller values

l. In the below space draw a time series that is not stationary in the covariance function.

(2)



The first half is positively correlated while the second half is negatively correlated.

(2)

m. (True) or False. The Airline Model is sort of a benchmark model that Box-Jenkins model builders use when they are modeling seasonal time series data.

2. In the SSP model, when forecasting the next year's total we can use the linear regression on the yearly totals that we have observed so far.
Suppose that we use linear regression and get the following linear regression based on 4 year totals:

$$\hat{y}_i^T = 30 + 4 * i$$

$\hat{y}_5^T = 30 + 4(5) = 30 + 20 = 50$

(2)

a. What is the 5-year total forecast? $\hat{y}_5^T =$ 50.

b. Considering the year total forecasting equation in question 2 above, suppose that we are working with quarterly data and we have the following quarterly average proportions computed over the first four years of the data:

$$\bar{P}_1 = 0.3, \bar{P}_2 = 0.2, \bar{P}_3 = 0.35, \text{ and } \bar{P}_4 = 0.15$$

Qtr. 1 = 50(.3) = 15

Qtr. 2 = 50(.2) = 10

Qtr. 3 = 50(.35) = 17.5

Qtr. 4 = 50 − 15 − 10 − 17.5 = 50(.15) = 7.5

(4)

The quarterly forecasts for next year (the fifth year) are

Qtr 1 = 15, Qtr 2 = 10, Qtr 3 = 17.5, Qtr 4 = 7.5

c. Given the information in question 3 above, we can say that
The "strongest" quarter during the year is Qtr. 3.
The "weakest" quarter during the year is Qtr. 4.

(2)

3. Consider graphs contained in Output #1 that you have been given.

(1) a. In Graph A the appropriate way to make the data stationary for the Box-Jenkins approach is to take first difference of data: Δy_t

(1) b. In Graph B the appropriate way to make the data stationary for the Box-Jenkins approach is to leave alone. Don't difference data.

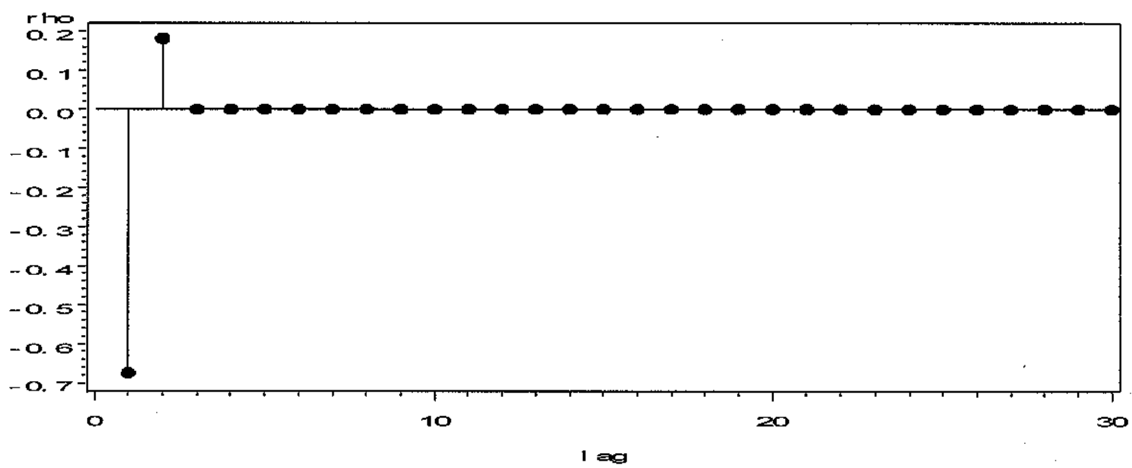
(1) c. In Graph C the appropriate way to make the data stationary for the Box-Jenkins approach is to model data as deterministic trend, $y_t = a + bt$

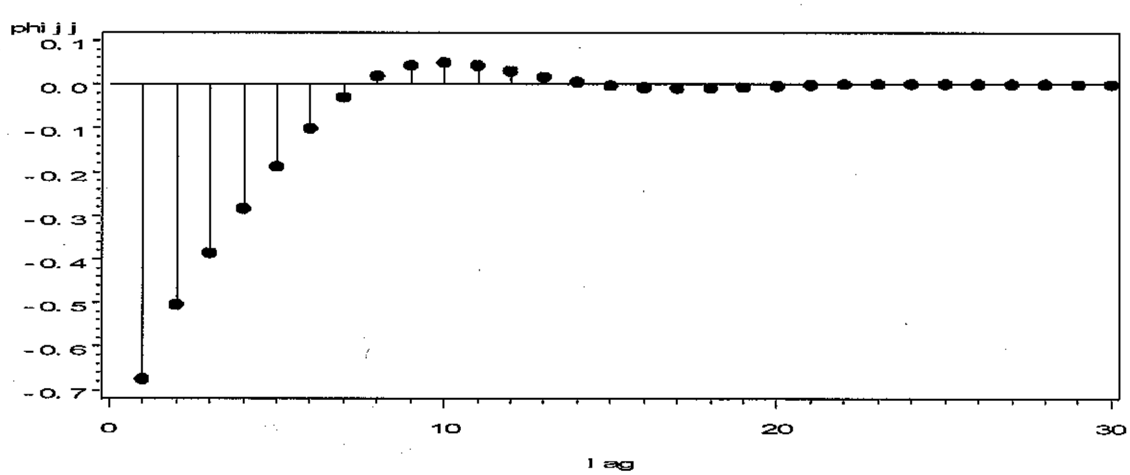
(1) d. In Graph D the appropriate way to make the data stationary for the Box-Jenkins approach is to take first difference of data; Δy_t

(1) e. In Graph E the appropriate way to make the data stationary for the Box-Jenkins approach is to take first difference of logged data: $\Delta \log(y_t)$

4. Consider the population ACF and PACF below. Given these graphs, your tentative identification of p and q in the Box-Jenkins frame work should be P = 0 and Q = 2.

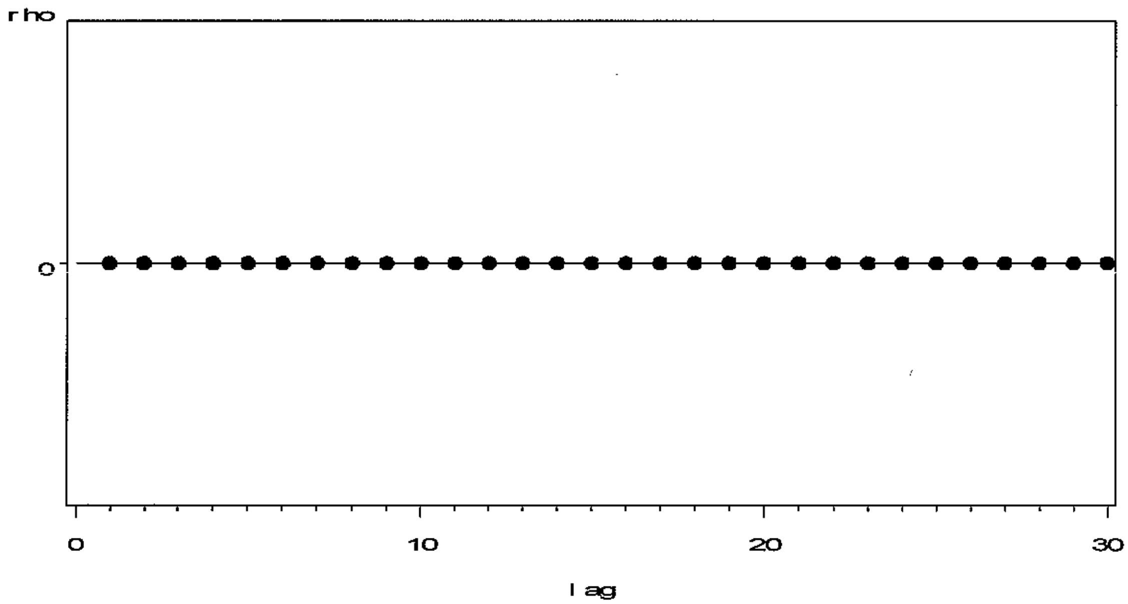
(2)

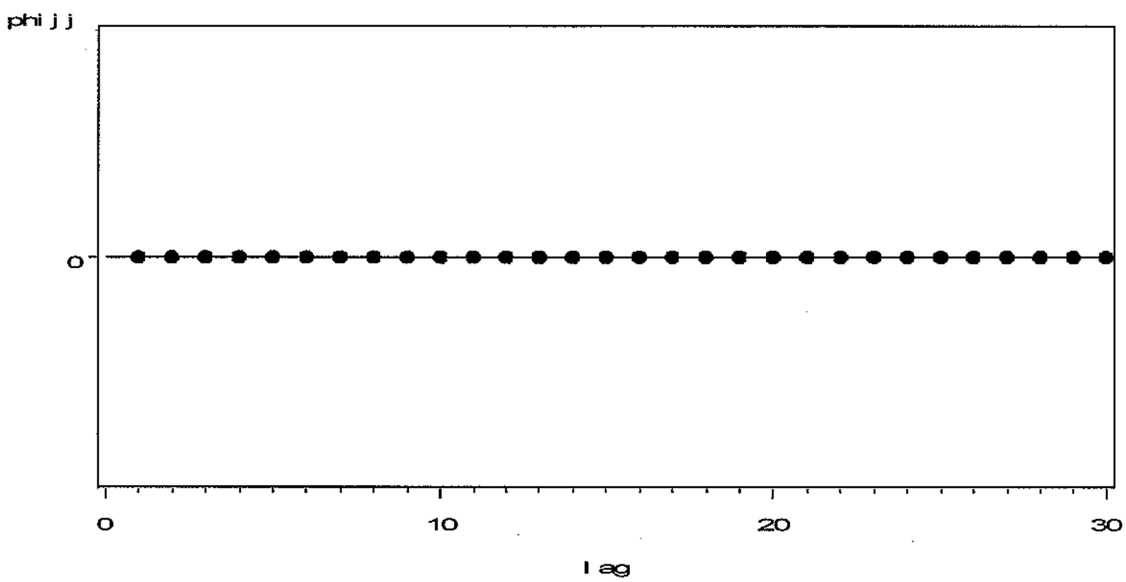




5. Consider the population ACF and PACF below. Given these graphs, your tentative identification of p and q in the Box-Jenkins frame work should be P = 0 and Q = 0.

(2)





6. Fill in the following Box-Jenkins Pattern Table:

Autocorrelation and Partial Autocorrelation Functions of Nonseasonal Models

(6)

Model	ACF	PACF
AR(p)	tails off	p spikes then cuts off
MA(q)	q spikes then cuts off	tails off
ARMA(p,q)	tails off	tails off

7. Consider the Computer Output #2 that you have been given. Use it to answer the following questions.

a. Take a look at the ACF and PACF of provided in Output # 2. What pattern do you see in these graphs and what is your tentative identification of p = __0__ and q = __1__. Be sure to clearly explain your reasoning.

(4)

ACF has one spike in it and then cuts off while the PACF tails off. ⇒ MA(1) model

b. Fill in the following P-Q box assuming that the data are observed **monthly**. Be sure and explain meanings of the entries in the Box.

(12)

← best model (circled: P = 0, Q = 1)

P \ Q	0	1	2
0	351.6999 354.3051 44.39 (0.0069)	285.6983 290.9087 18.90 (0.7072)	287.571 295.3865 19.61 (0.5951)
1	323.8188 329.0292 38.97 (0.0200)	287.5705 295.386 19.61 (0.6073)	
2	314.6469 322.4624 39.21 (0.0133)		

Legend: AIC
SBC
Q_{24}
(p-value)

AIC and SBC are goodness-of-fit measures. The smaller these measures, the better the model. The residuals of a model are white noise if the associated p-value of the Q-statistic is greater than 0.05.

(2) c. ~~b.~~ Given what you have put in the P-Q box and knowing what you do about the sample ACF and sample PACF what is your tentative choice of ARMA model?

P = 0. Q = 1. Explain your reasoning. This model has smallest AIC and SBC and residuals are white noise.

(10) d. ~~c.~~ Use Computer Output # ~~1~~ 3 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 t-statistic. What conclusion do you draw from the overfitting exercise? Explain your answer.

Since both of the overfitting coefficients are statistically insignificant (i.e. p>0.05) we choose the original (MA(1)) model.

Model 1: p = 0, q = 2
Overfitting coefficient = -0.04405,
Overfitting t-statistic = -0.44.
P-value of overfitting t-statistic = 0.6629

Model 2: p = 1, q = 1
Overfitting coefficient = 0.03726,
Overfitting t-statistic = 0.35.
P-value of overfitting t-statistic = 0.7289

(4) e. ~~d.~~ In the below space, write out the final model that you have chosen for the Y time series with accompanying standard errors, goodness-of-fit measures, and a test statistic for white noise residuals with accompanying p-value. (You can use either the intercept-form or the deviation-from-mean form.)

intercept form → $y_t = \underset{(0.18524)}{49.98908} + \hat{a}_t - \underset{(0.03345)}{(-0.95652)}\hat{a}_{t-1}$

deviation-from-mean form → $y_t - 49.98908 = \hat{a}_t - (-0.95652)\hat{a}_{t-1}$

AIC = 285.6988, SBC = 290.9087, Q_{12} = 8.90 (p = 0.7072)

8. Consider the Computer Output # 3 that you have been given and use it to answer the following questions:

(2) a. Which case of the Augmented Dickey-Fuller test is appropriate for analyzing the data plotted in Computer Output # 3?

Case = 3. (or "trend" case)

(4) b. The null hypothesis of the test is

H_0: data needs to be differenced

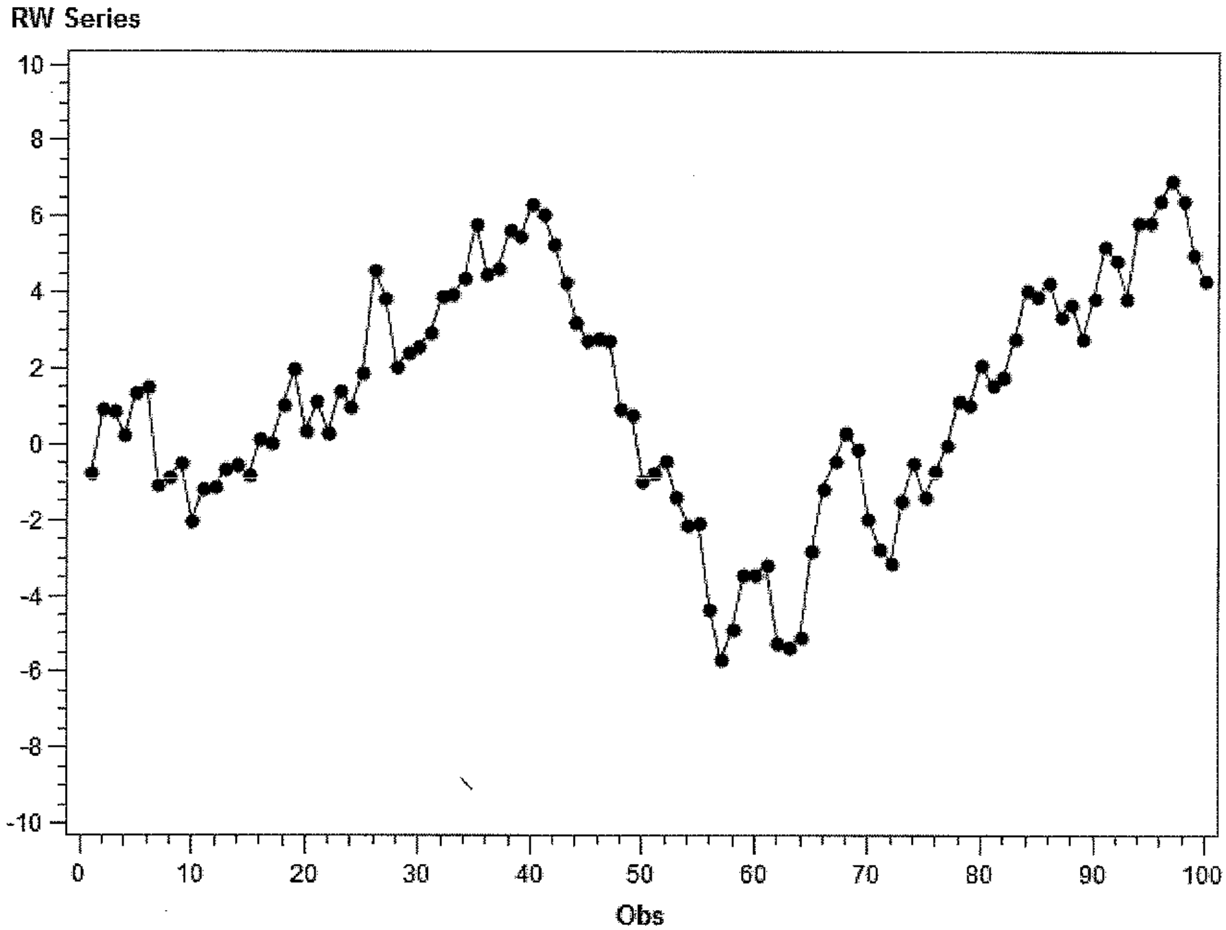
The alternative hypothesis of the test is

H_1: data should be modeled as a deterministic time series.

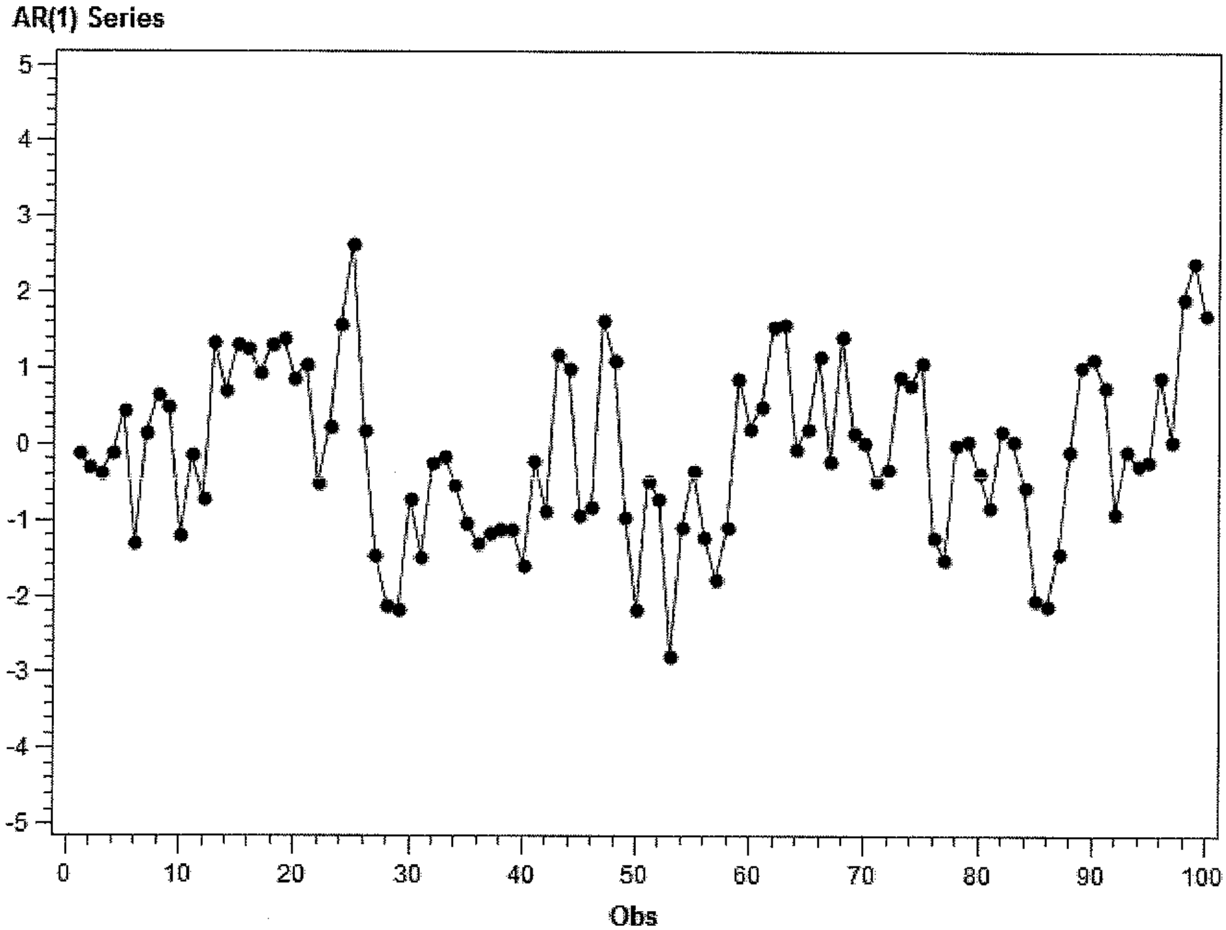
c. The Dickey-Fuller t-statistic = -8.444240
Its probability value = 0.0000.
The number of augmenting terms = 0.
The information criterion used to select these terms = SIC.
The outcome of the test is Reject H_0, Accept H_1.
Therefore model the data as a deterministic trend.

(5)

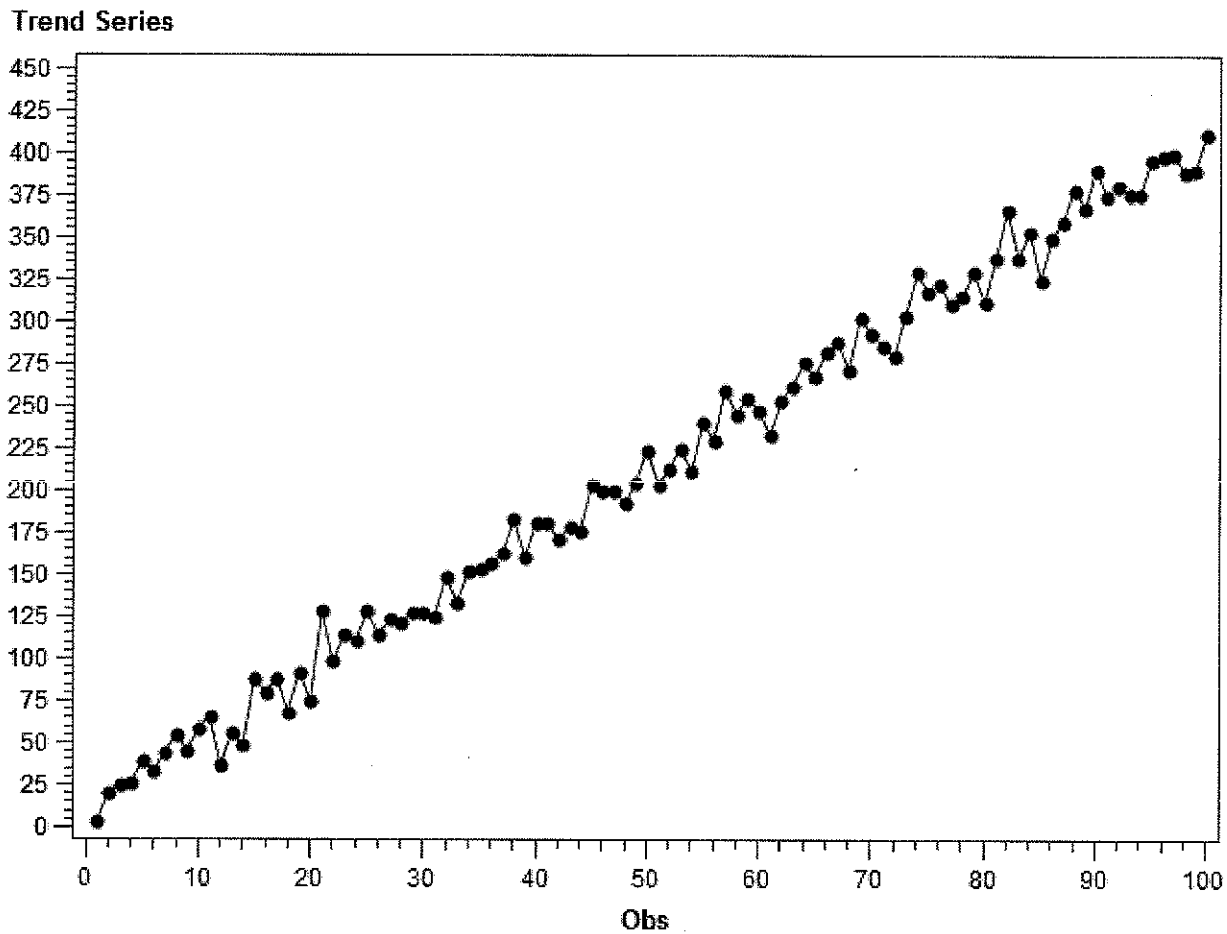
Output #1



GRAPH A

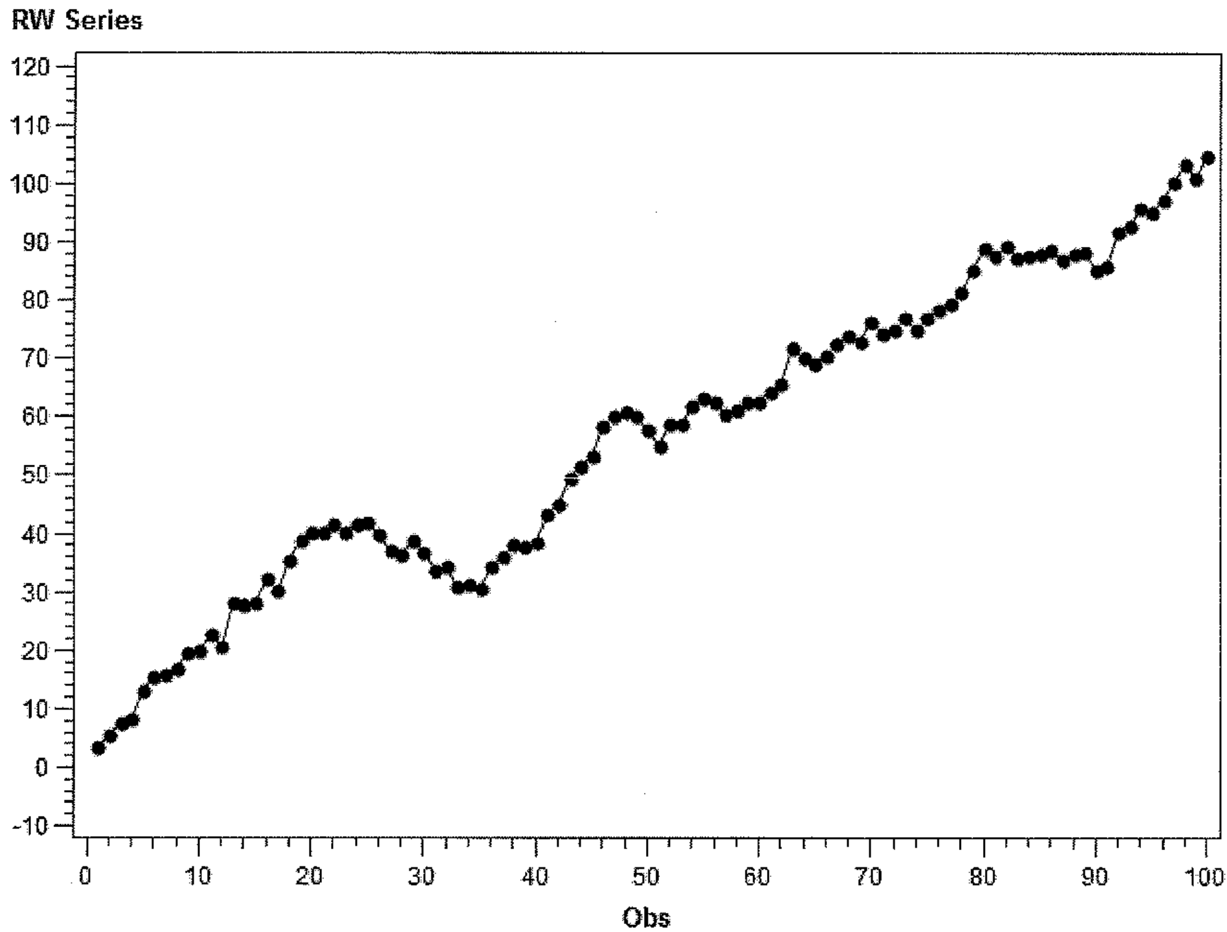


GRAPH B



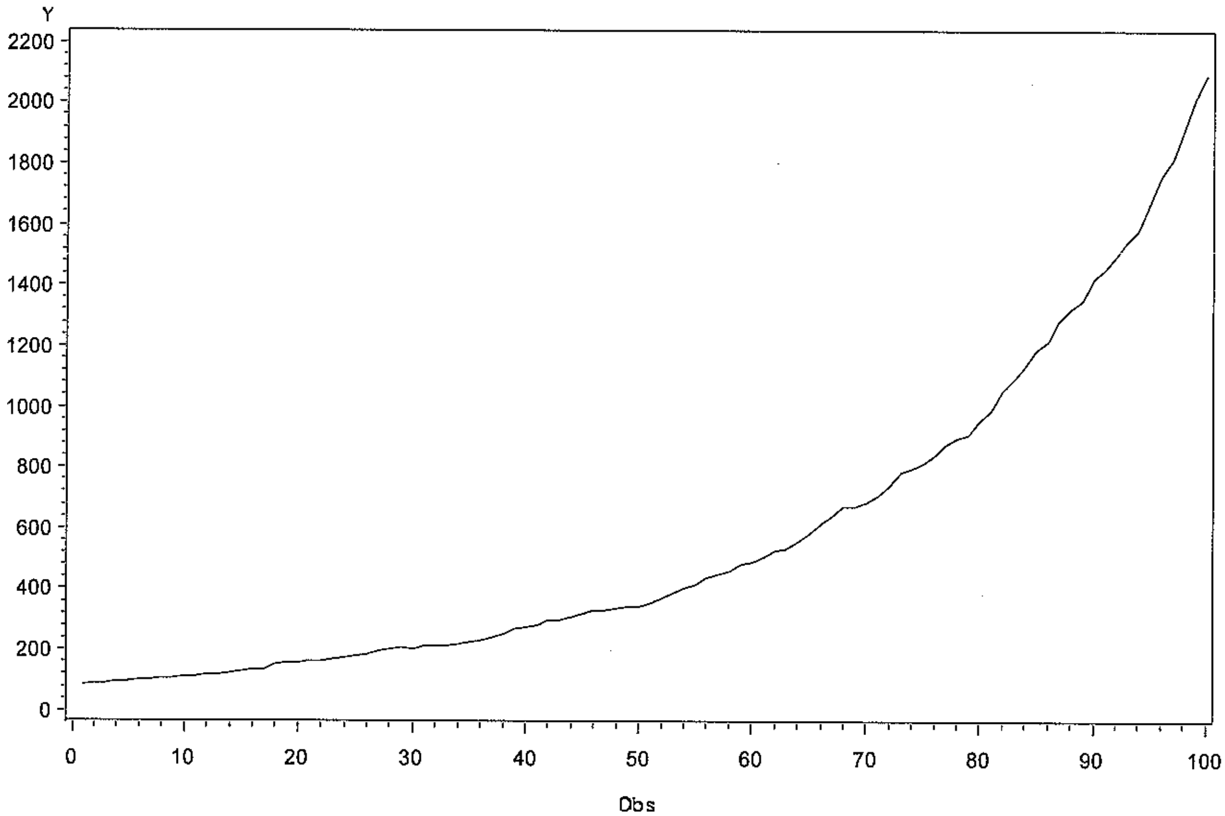
Trend Series
450
425
400
375
350
325
300
275
250
225
200
175
150
125
100
75
50
25
0
0 10 20 30 40 50 60 70 80 90 100
Obs

GRAPH C


RW Series
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
0
10
20
30
40
50
60
70
80
90
100
Obs

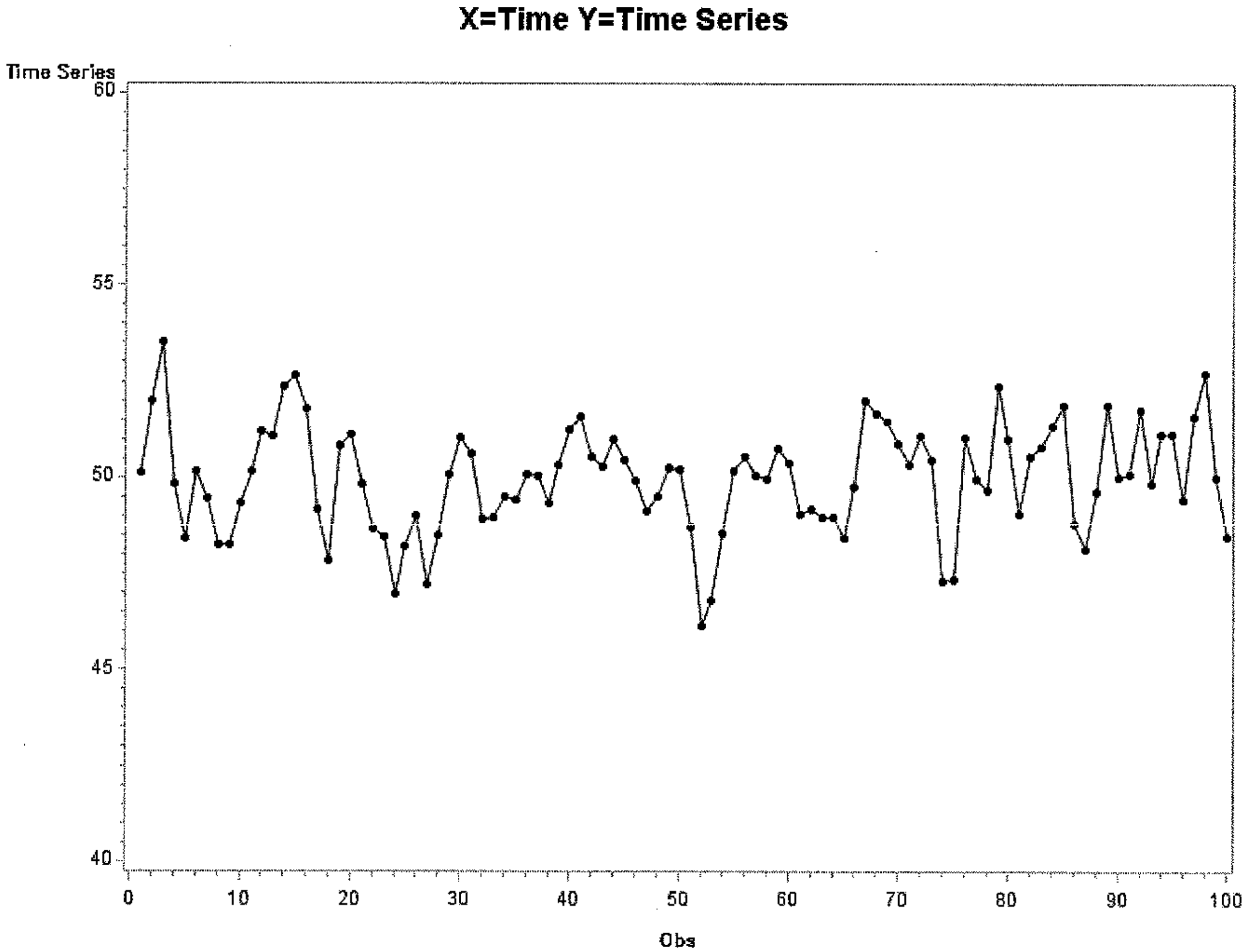
GRAPH D

Plot of Y
100 Observations



GRAPH E

Output #2


X=Time Y=Time Series
Time Series
60
55
50
45
40
0
10
20
30
40
50
60
70
80
90
100
Obs

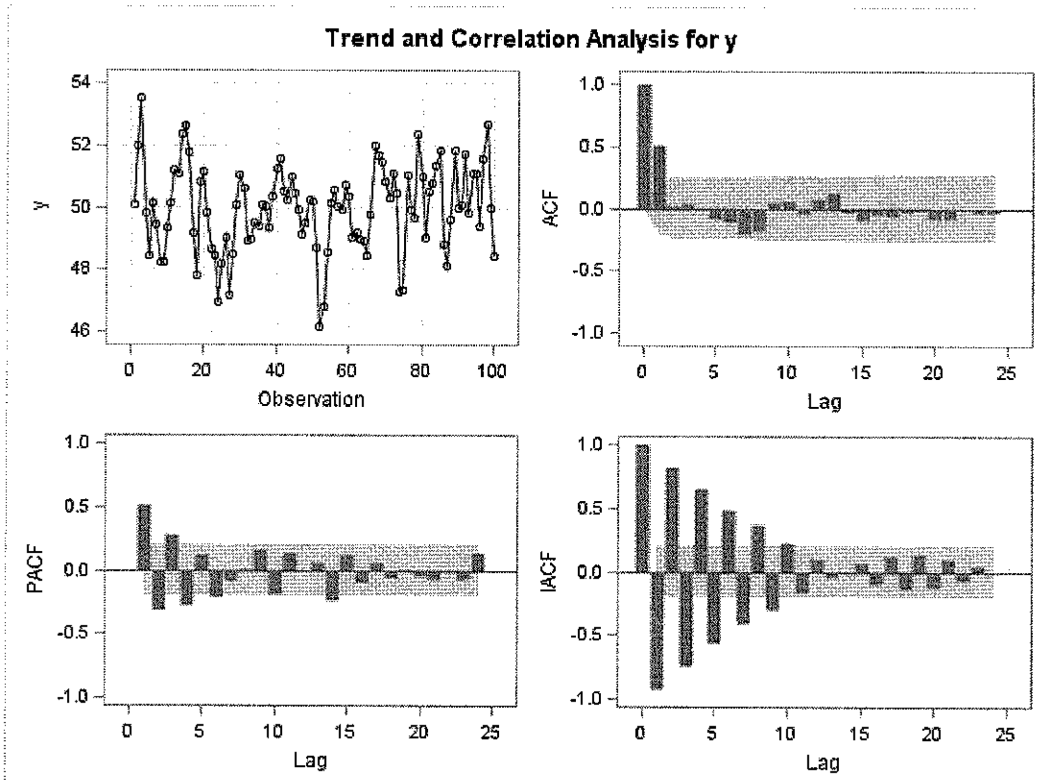
Time Series for Computer Output 1

The ARIMA Procedure

Name of Variable = y	
Mean of Working Series	49.96876
Standard Deviation	1.390358
Number of Observations	100

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	28.72	6	<.0001	0.505	0.019	0.038	-0.009	-0.089	-0.113
12	38.09	12	0.0001	-0.203	-0.175	0.053	0.054	-0.040	0.074
18	42.30	18	0.0010	0.132	-0.034	-0.101	-0.039	-0.060	-0.034
24	44.39	24	0.0069	-0.024	-0.086	-0.081	-0.004	-0.033	-0.026

Trend and Correlation Analysis for y



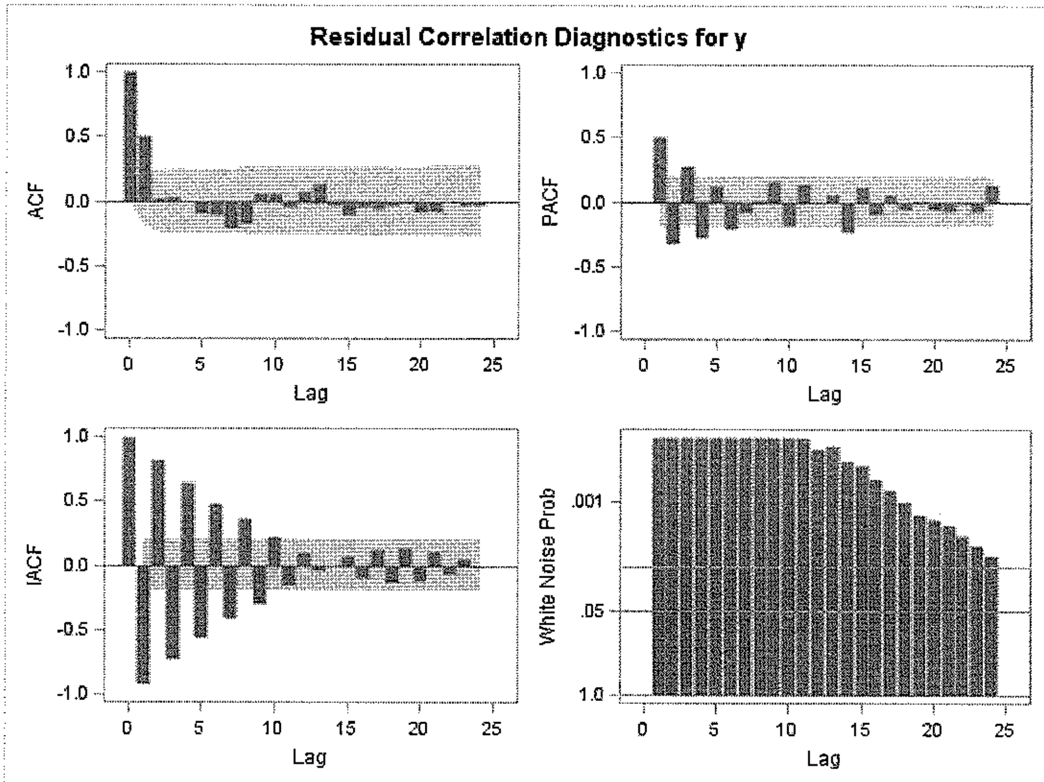
Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	49.96876	0.13974	357.59	<.0001	0

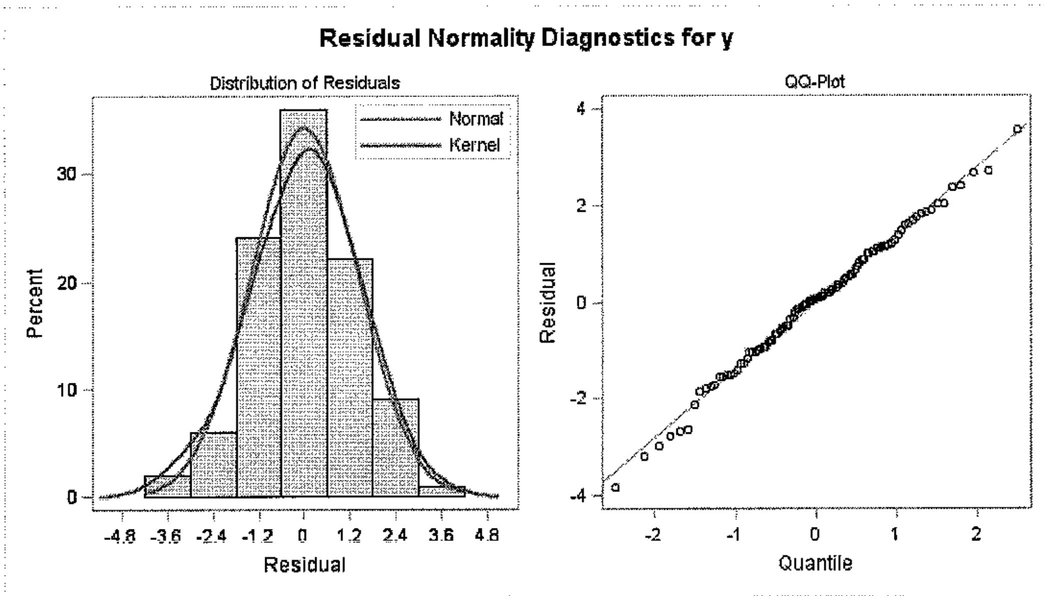
Constant Estimate	49.96876
Variance Estimate	1.952621
Std Error Estimate	1.397362
AIC	351.6999

SBC	354.3051
Number of Residuals	100

*** AIC and SBC do not include log determinant.**

Autocorrelation Check of Residuals									
To Lag	**Chi-Square**	**DF**	**Pr > ChiSq**	**Autocorrelations**					
6	28.72	6	<.0001	0.505	0.019	0.038	-0.009	-0.089	-0.113
12	38.09	12	0.0001	-0.203	-0.175	0.053	0.054	-0.040	0.074
18	42.30	18	0.0010	0.132	-0.034	-0.101	-0.039	-0.060	-0.034
24	44.39	24	0.0069	-0.024	-0.086	-0.081	-0.004	-0.033	-0.026





Model for variable y	
Estimated Mean	49.96876

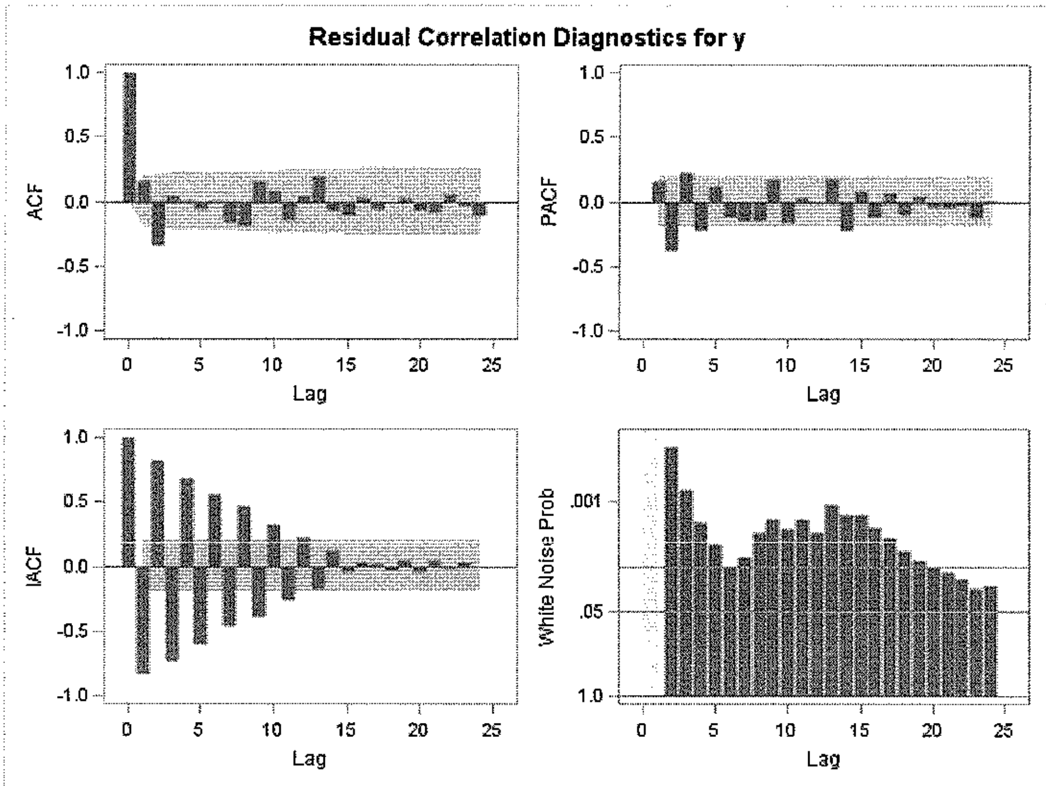
Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	49.95608	0.24369	205.00	<.0001	0
AR1,1	0.51126	0.08752	5.84	<.0001	1

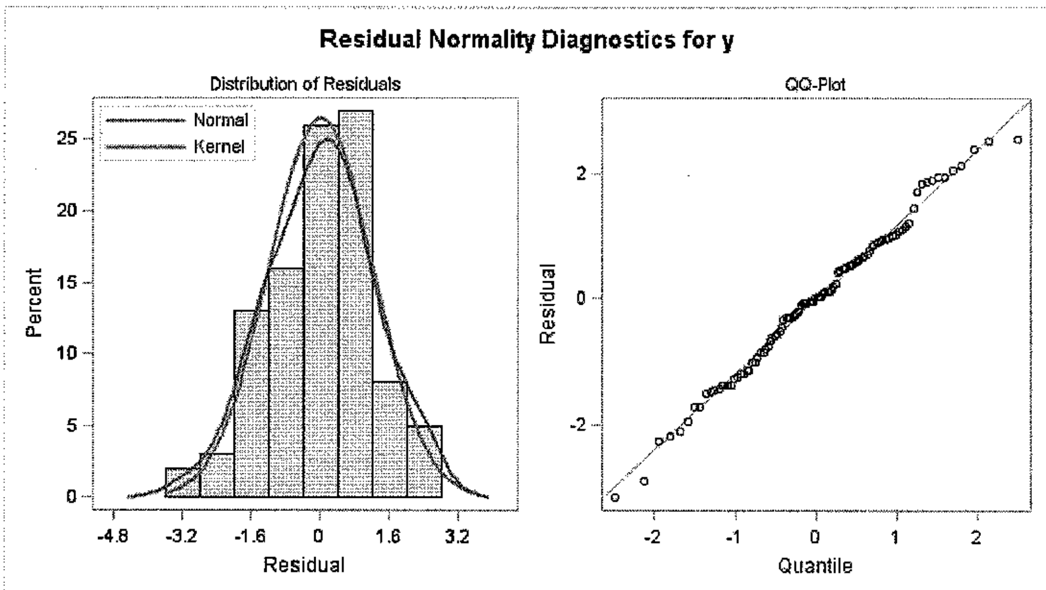
Constant Estimate	24.41574
Variance Estimate	1.463036
Std Error Estimate	1.20956
AIC	323.8188
SBC	329.0292
Number of Residuals	100

*** AIC and SBC do not include log determinant.**

Correlations of Parameter Estimates		
Parameter	MU	AR1,1
MU	1.000	-0.020
AR1,1	-0.020	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	15.05	5	0.0101	0.165	-0.334	0.046	0.010	-0.061	0.006
12	28.24	11	0.0030	-0.154	-0.188	0.163	0.083	-0.151	0.051
18	35.33	17	0.0056	0.194	-0.073	-0.111	0.038	-0.057	-0.003
24	38.97	23	0.0200	0.037	-0.067	-0.075	0.058	-0.035	-0.108





Model for variable y	
Estimated Mean	49.95608

Autoregressive Factors	
Factor 1:	1 - 0.51126 B**(1)

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	49.94664	0.17676	282.56	<.0001	0

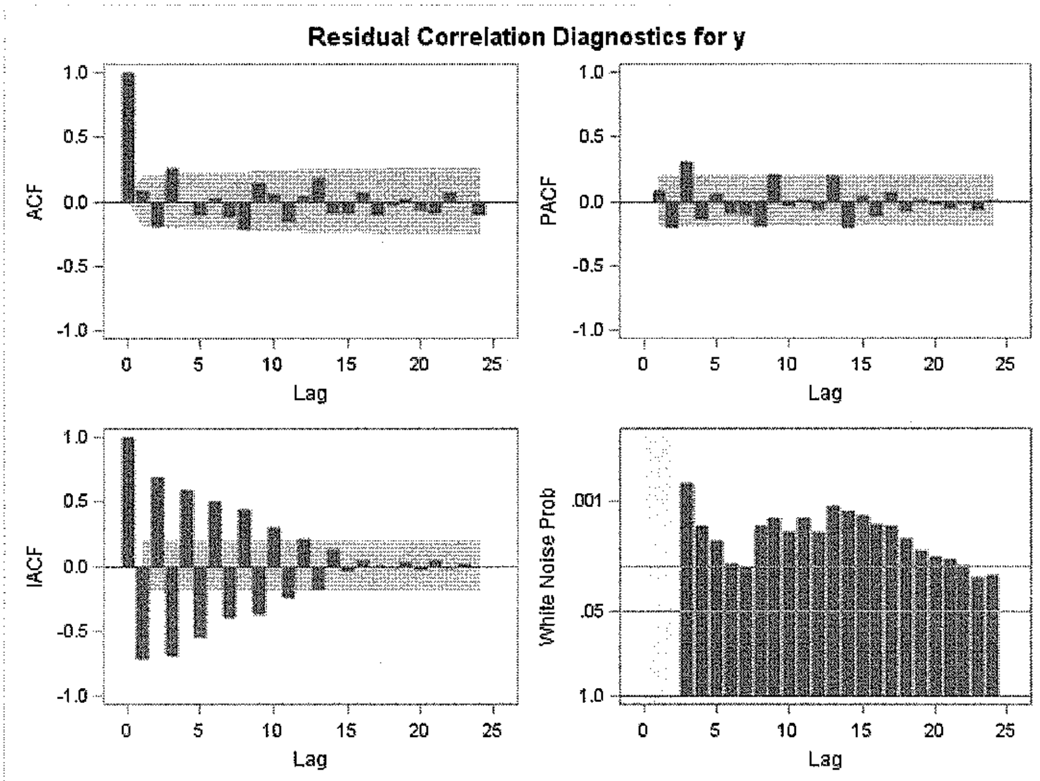
AR1,1	0.67900	0.09679	7.01	<.0001	1
AR1,2	-0.32773	0.09680	-3.39	0.0010	2

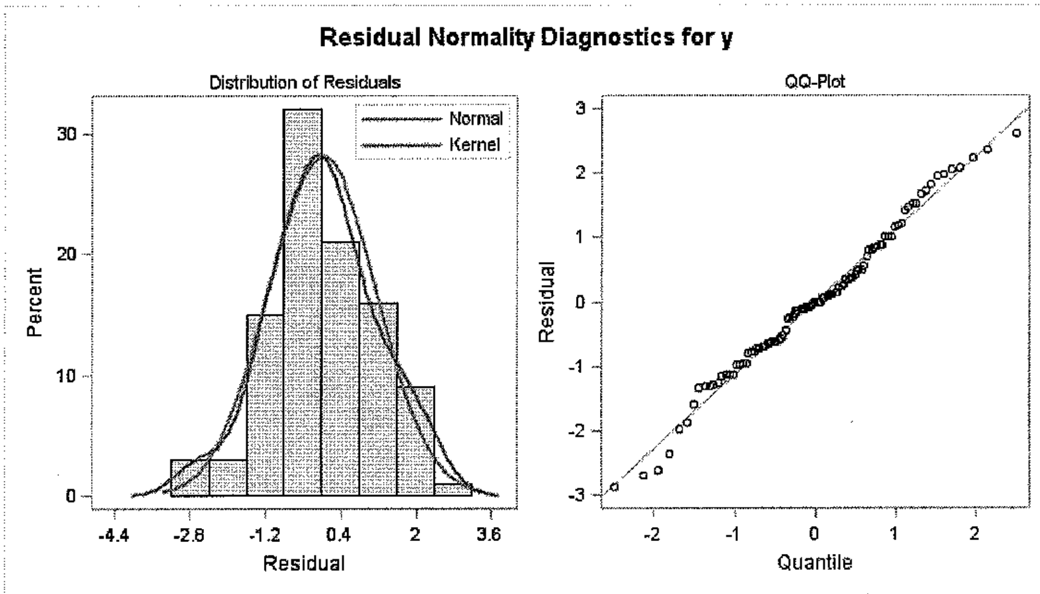
Constant Estimate	32.40194
Variance Estimate	1.321875
Std Error Estimate	1.149728
AIC	314.6469
SBC	322.4624
Number of Residuals	100

*** AIC and SBC do not include log determinant.**

Correlations of Parameter Estimates			
Parameter	**MU**	**AR1,1**	**AR1,2**
MU	1.000	-0.014	-0.015
AR1,1	-0.014	1.000	-0.511
AR1,2	-0.015	-0.511	1.000

Autocorrelation Check of Residuals									
To Lag	**Chi-Square**	**DF**	**Pr > ChiSq**	**Autocorrelations**					
6	13.43	4	0.0094	0.087	-0.194	0.264	-0.011	-0.109	0.032
12	26.59	10	0.0030	-0.121	-0.220	0.149	0.062	-0.163	0.041
18	35.22	16	0.0037	0.190	-0.101	-0.091	0.074	-0.109	-0.027
24	39.21	22	0.0133	0.025	-0.069	-0.096	0.066	-0.005	-0.107





Model for variable y	
Estimated Mean	49.94664

Autoregressive Factors	
Factor 1:	1 - 0.679 B**(1) + 0.32773 B**(2)

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	49.98908	0.18524	269.86	<.0001	0
MA1,1	-0.95652	0.03345	-28.60	<.0001	1

Constant Estimate	49.98908
Variance Estimate	0.999309
Std Error Estimate	0.999654
AIC	285.6983
SBC	290.9087
Number of Residuals	100

*** AIC and SBC do not include log determinant.**

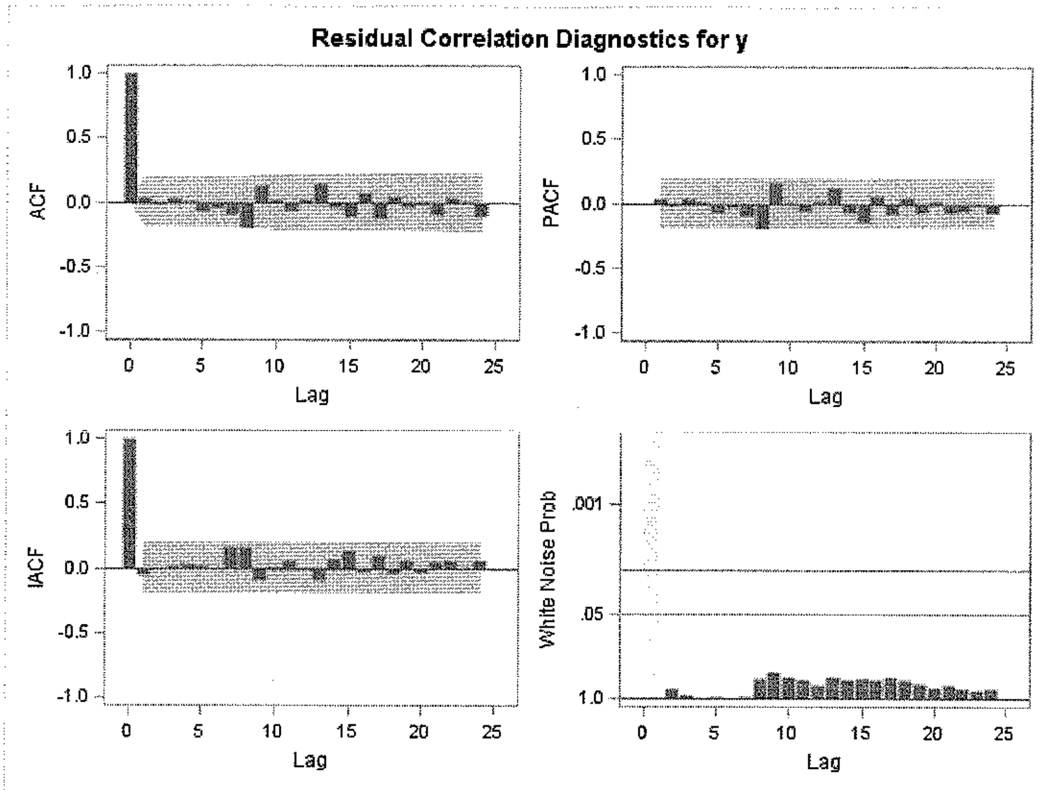
Correlations of Parameter Estimates

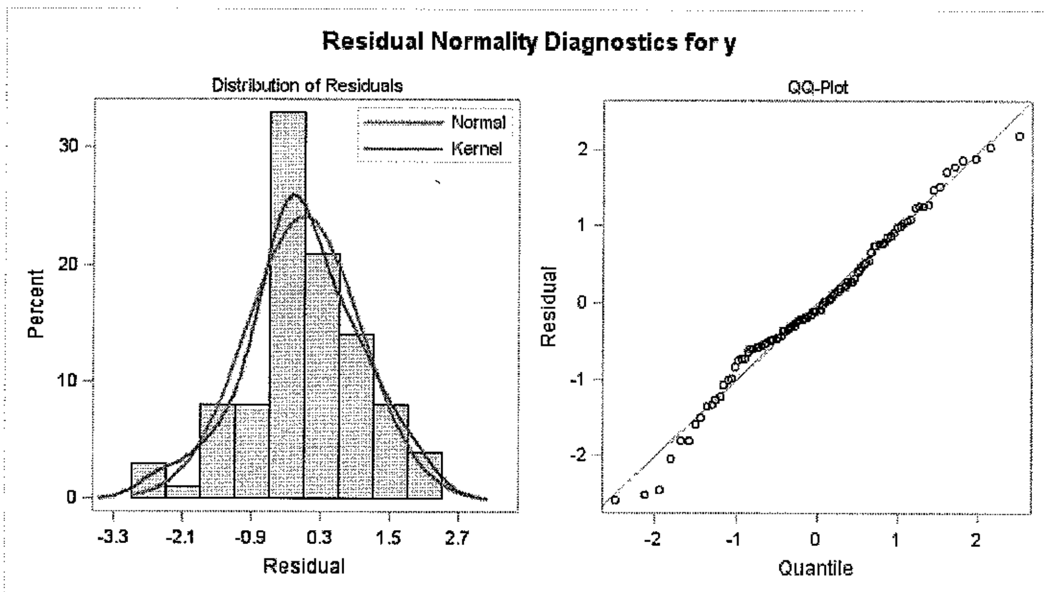
Parameter	MU	MA1,1
MU	1.000	-0.046
MA1,1	-0.046	1.000

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	0.96	5	0.9656	0.034	-0.013	0.032	0.005	-0.072	-0.037
12	9.14	11	0.6088	-0.094	-0.201	0.139	0.018	-0.062	0.027

18	15.98	17	0.5252	0.149	-0.045	-0.110	0.068	-0.121	0.041
24	18.90	23	0.7072	-0.035	-0.020	-0.090	0.030	0.008	-0.107





Model for variable y	
Estimated Mean	49.98908

Moving Average Factors	
Factor 1:	1 + 0.95652 B**(1)

Conditional Least Squares Estimation

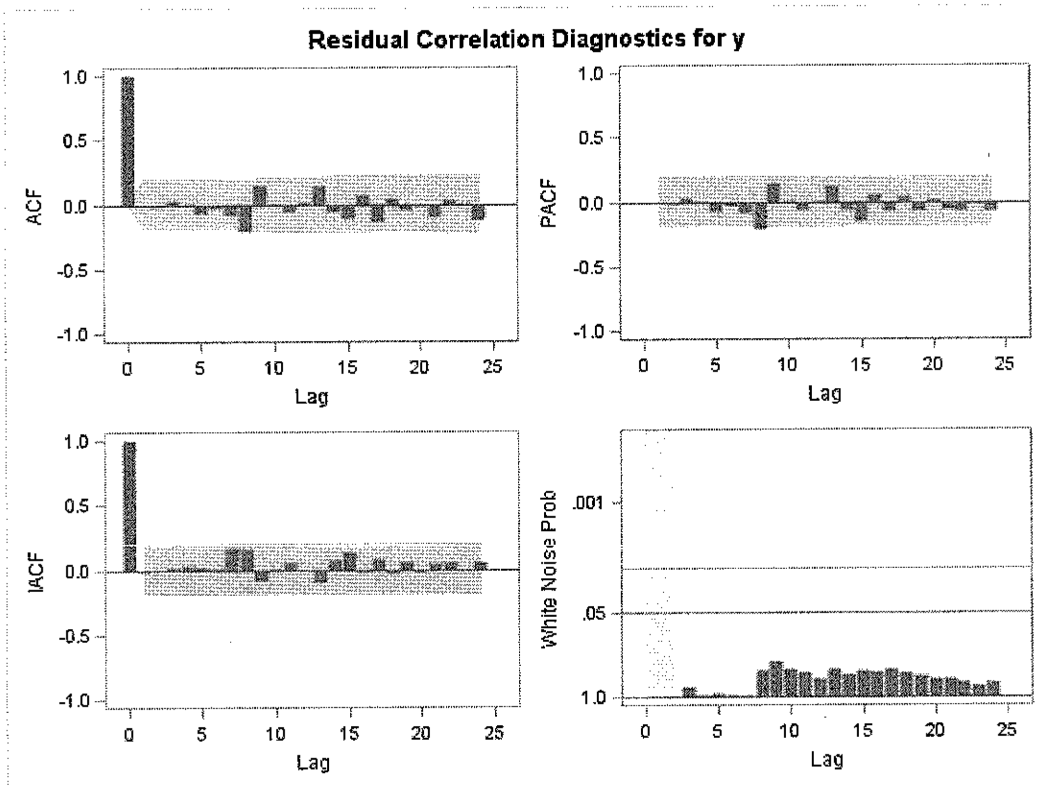
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	49.99108	0.19308	258.91	<.0001	0
MA1,1	-1.00002	0.10219	-9.79	<.0001	1
MA1,2	-0.04405	0.10074	-0.44	0.6629	2

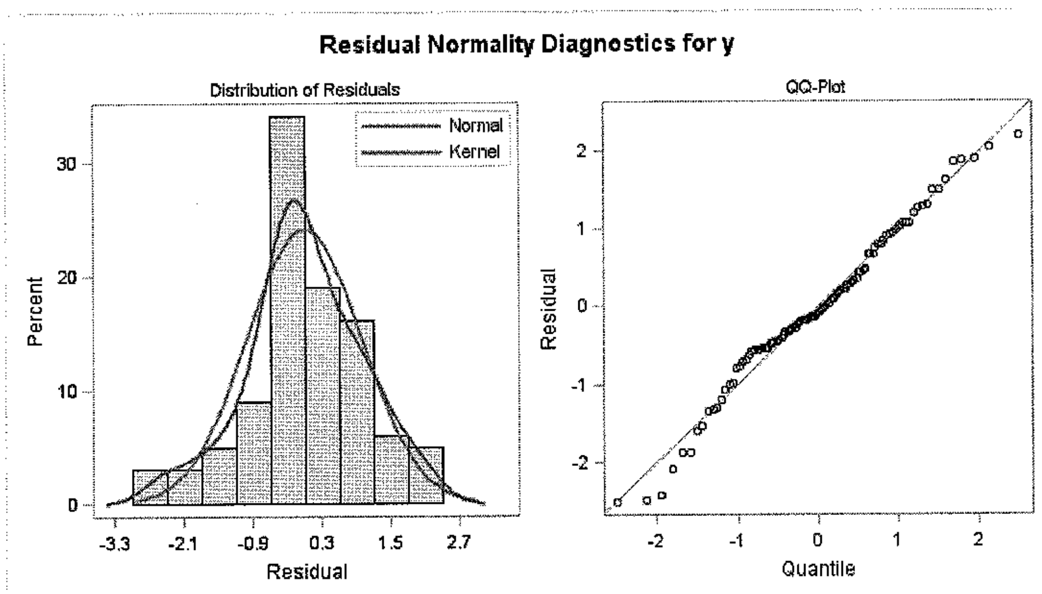
Constant Estimate	49.99108
Variance Estimate	1.008327
Std Error Estimate	1.004155
AIC	287.571
SBC	295.3865
Number of Residuals	100

*** AIC and SBC do not include log determinant.**

Correlations of Parameter Estimates			
Parameter	MU	MA1,1	MA1,2
MU	1.000	-0.021	-0.008
MA1,1	-0.021	1.000	0.947
MA1,2	-0.008	0.947	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	0.79	4	0.9402	-0.008	-0.017	0.035	0.004	-0.069	-0.032
12	9.27	10	0.5068	-0.082	-0.206	0.150	0.012	-0.062	0.021
18	16.63	16	0.4099	0.151	-0.048	-0.109	0.078	-0.126	0.047
24	19.81	22	0.5951	-0.036	-0.015	-0.090	0.031	0.013	-0.116





Model for variable y	
Estimated Mean	49.99108

Moving Average Factors	
Factor 1:	1 + 1.00002 B**(1) + 0.04405 B**(2)

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	49.99101	0.19201	260.35	<.0001	0

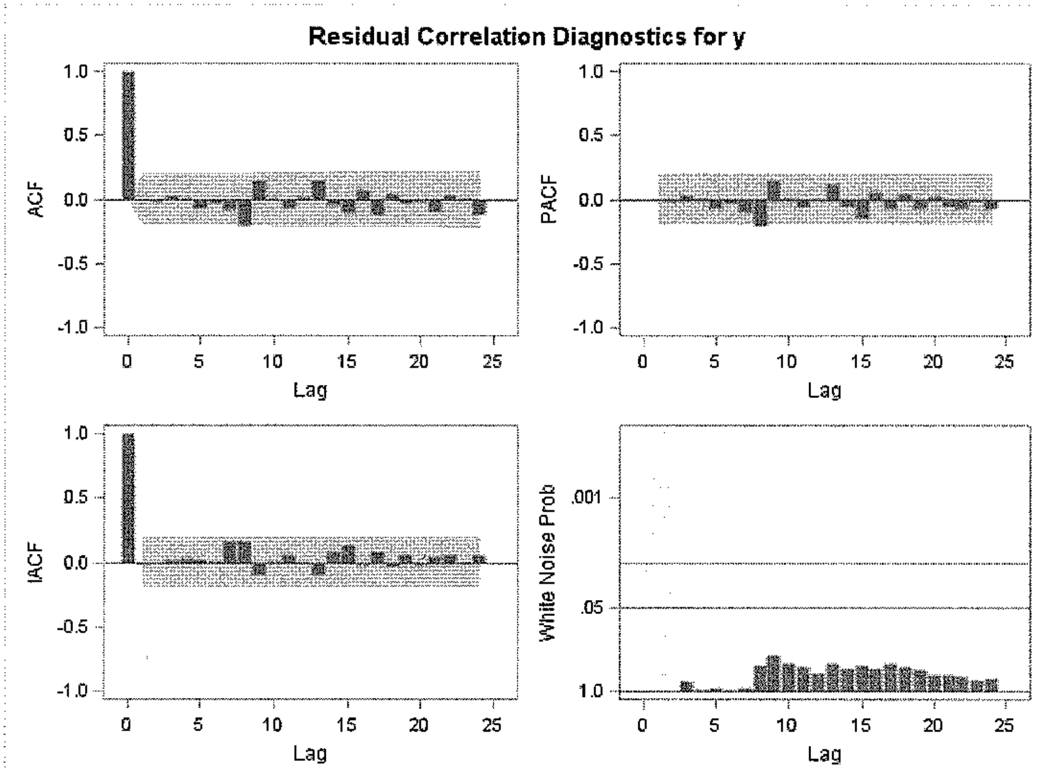
MA1,1	-0.95435	0.03590	-26.59	<.0001	1
AR1,1	0.03726	0.10720	0.35	0.7289	1

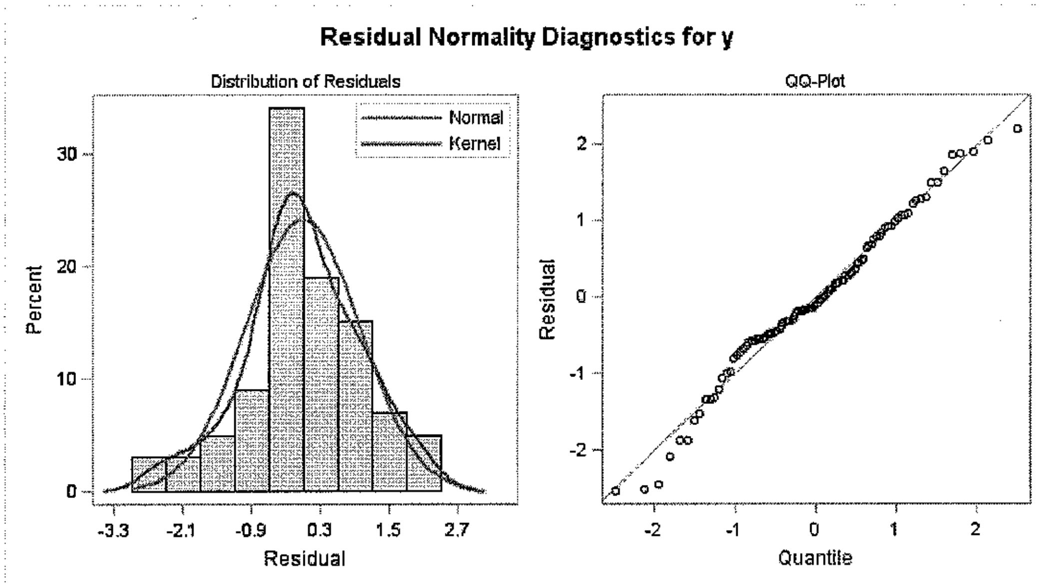
Constant Estimate	48.12834
Variance Estimate	1.008322
Std Error Estimate	1.004152
AIC	287.5705
SBC	295.386
Number of Residuals	100

*** AIC and SBC do not include log determinant.**

Correlations of Parameter Estimates			
Parameter	MU	MA1,1	AR1,1
MU	1.000	-0.041	0.006
MA1,1	-0.041	1.000	0.312
AR1,1	0.006	0.312	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	0.79	4	0.9392	0.000	-0.018	0.035	0.005	-0.069	-0.033
12	9.23	10	0.5106	-0.084	-0.205	0.148	0.014	-0.062	0.022
18	16.49	16	0.4195	0.151	-0.048	-0.109	0.076	-0.124	0.046
24	19.61	22	0.6073	-0.035	-0.016	-0.090	0.031	0.012	-0.114



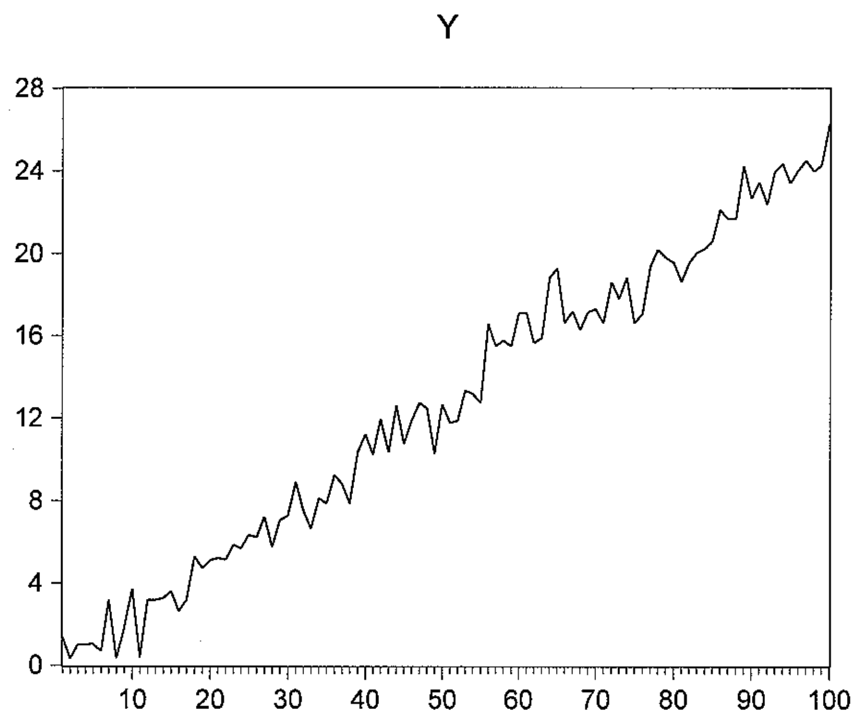


Model for variable y	
Estimated Mean	49.99101

Autoregressive Factors	
Factor 1:	1 - 0.03726 B**(1)

Moving Average Factors	
Factor 1:	1 + 0.95435 B**(1)

Output #3



Augmented Dickey-Fuller Unit Root Test on Y

Null Hypothesis: Y has a unit root
Exogenous: None
Lag Length: 3 (Automatic - based on SIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		3.832490	1.0000
Test critical values:	1% level	-2.589273	
	5% level	-1.944211	
	10% level	-1.614532	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(Y)
Method: Least Squares
Date: 10/21/11 Time: 10:04
Sample (adjusted): 5 100
Included observations: 96 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	0.034356	0.008965	3.832490	0.0002
D(Y(-1))	-0.705677	0.106942	-6.598669	0.0000
D(Y(-2))	-0.464576	0.119686	-3.881612	0.0002
D(Y(-3))	-0.157363	0.105266	-1.494913	0.1384

R-squared	0.314794	Mean dependent var	0.263349
Adjusted R-squared	0.292451	S.D. dependent var	1.346603
S.E. of regression	1.132708	Akaike info criterion	3.127873
Sum squared resid	118.0385	Schwarz criterion	3.234721
Log likelihood	-146.1379	Hannan-Quinn criter.	3.171062
Durbin-Watson stat	2.028775		

Augmented Dickey-Fuller Unit Root Test on Y

Null Hypothesis: Y has a unit root
Exogenous: Constant
Lag Length: 3 (Automatic - based on SIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.124729	0.9429
Test critical values:	1% level	-3.499910	
	5% level	-2.891871	
	10% level	-2.583017	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(Y)
Method: Least Squares
Date: 10/21/11 Time: 10:05
Sample (adjusted): 5 100
Included observations: 96 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	-0.001973	0.015821	-0.124729	0.9010
D(Y(-1))	-0.740583	0.104117	-7.112965	0.0000
D(Y(-2))	-0.526210	0.117813	-4.466506	0.0000
D(Y(-3))	-0.206858	0.103304	-2.002421	0.0482
C	0.648279	0.236233	2.744231	0.0073

R-squared	0.367165	Mean dependent var	0.263349
Adjusted R-squared	0.339348	S.D. dependent var	1.346603
S.E. of regression	1.094525	Akaike info criterion	3.069196
Sum squared resid	109.0167	Schwarz criterion	3.202756
Log likelihood	-142.3214	Hannan-Quinn criter.	3.123183
F-statistic	13.19936	Durbin-Watson stat	2.068511
Prob(F-statistic)	0.000000		

Augmented Dickey-Fuller Unit Root Test on Y

Null Hypothesis: Y has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-8.444240	0.0000
Test critical values:	1% level	-4.053392	
	5% level	-3.455842	
	10% level	-3.153710	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(Y)
Method: Least Squares
Date: 10/21/11 Time: 10:05
Sample (adjusted): 2 100
Included observations: 99 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	-0.850625	0.100734	-8.444240	0.0000
C	0.072588	0.206939	0.350770	0.7265
@TREND(1)	0.216415	0.025709	8.417842	0.0000

R-squared	0.426754	Mean dependent var	0.251891
Adjusted R-squared	0.414811	S.D. dependent var	1.333111
S.E. of regression	1.019798	Akaike info criterion	2.906921
Sum squared resid	99.83887	Schwarz criterion	2.985561
Log likelihood	-140.8926	Hannan-Quinn criter.	2.938739
F-statistic	35.73367	Durbin-Watson stat	2.009839
Prob(F-statistic)	0.000000		