Name Mr. Key
ID# 7777777

ECO 5375
Eco. & Bus. Forecasting

Prof. Tom Fomby
Fall 2016

MID-TERM EXAM I

Instructions: Write in your name and student ID in the blanks above. You have 1 hour and 20 minutes to complete this exam. This exam is worth a total of 106 points. The points for the separate question are broken out as follows:

Q1 = (10, 2, 2, 2, 2) = 18 points
Q2 = (4, 4, 4) = 12 points
Q3 = (2, 2, 2, 2, 2, 2, 2, 2, 2, 4) = 22 points
Q4 = 4 points
Q5 =4 points
Q6 = 4 points
Q7 = (2, 2, 2) = 6 points
Q8 = 4 points
Q9 = 4 points
Q10 = (12, 2) = 14 points
Q11 = (4, 4, 2) = 10 points
Q12 = 4 points

1. Consider the following quarterly sales data (in millions) of the XYZ corporation:

YEAR	Qtr 1	Qtr 2	Qtr 3	Qtr 4	year total	
						39
1	18	43	22	30	113	89
2	21	46	24	34	125	46
	39	89	46	64		64
3	22.45	51.22	26.47	36.84		238

$p_1 = \frac{39}{238} = 0.1639, \; p_2 = \frac{89}{238} = 0.3739, \; p_3 = \frac{46}{238} = 0.1932, \; p_4 = \frac{64}{238} = 0.2689$

Total for year 3 estimate $= 125 + (125 - 113) = 137$

Qtr 1_yr3 estimate $= 137(0.1639) = 22.45$ Q2-y3 $= 137(0.3739) = 51.22$

a) Using the SSP model approach predict the sales of the XYZ corporation for quarters 1 through 4 for the third year. Show in detail how you got your answers.

Qtr 3-yr3 estimate $= 137(0.1932) = 26.4684$

(10) Qtr 4-yr3 estimate $= 137(0.2689) = 36.8393$

(Note: see next page insert)

b) What are the major assumptions of the SSP model? That the seasonal proportions are stable across years and that the change

(2) in year totals are "smooth", in the sense that they are easy to estimate.

c) In the above quarterly data which quarter is the strongest quarter? Which quarter is the weakest quarter? Support your answer with specific information. The strongest quarter is the ~~third~~ second quarter. It has the highest

(2) proportion ($p_2 = 0.3739$). The weakest quarter is the first quarter. It has the lowest proportion ($p_1 = 0.1639$).

d) Briefly explain the logic of Friedman's two-way ANOVA test for seasonality in time series data. What is the null hypothesis of the test? What is the alternative hypothesis of the test? Each year's data is ranked from highest-to-lowest (1, 2, ⋯, s) and then the ranks of the data are examined

(2) across years. If the ranks vary randomly across years, the Friedman test accepts the null hypothesis that the data are not seasonal. On the other hand, if the ranks by season

are not randomly distributed across years the Friedman test accepts the alternative hypothesis that the data are seasonal.

Question 1

Note: Instead of using the "change" estimator, you could have used the "growth" estimator of the third year total.

$$\frac{125-113}{113} = 0.106$$

In this case the quarterly estimates for year three would be

$$Q1 = 0.1639\overbrace{(125)(1.106)}^{138.25} = 22.62$$

$$Q2 = 0.3739(125)(1.106) = 51.69$$

$$Q3 = 0.1932(125)(1.106) = 26.71$$

$$Q4 = 0.2689(125)(1.106) = 37.18$$

These numbers are not very different from the numbers derived by using the "change" estimator.

You could have used either of these estimates and you would be OK.

e) Suppose that we ran the Friedman test on the above data and the test gave rise to a p-value of p = 0.4. What would you conclude? Explain your answer.

(2) You would accept the null hypothesis that the data is not seasonal since the p-value is greater than 0.05.

2. Consider the following SAS code that we used to model the Plano Sales Tax Revenue data with a Deterministic Trend/Deterministic Season (DTDS) model:

```
proc reg data = plano;
   model rev = t t2 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12/DWPROB;
run;

proc autoreg data = plano;
  model rev = t t2 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12/ nlag = 12
                   method=ml backstep slstay=0.05;
run;

proc autoreg data = plano;
  model rev = t d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12/ nlag = 12
                   method=ml backstep slstay=0.05;
    test d2, d3, d4, d5, d6, d7, d8, d9, d10, d11, d12;
run;
```

a) Describe the purpose of the "proc reg" statement above. What was the conclusion?

(4) This step uses OLS to get the OLS residuals of the model to test for autocorrelated errors using the Durbin-Watson statistic. We found the OLS residuals to have significant autocorrelation so we knew we could not conduct statistical inference using OLS (proc reg). Instead we needed to move to GLS estimation of the model (Proc Autoreg).

b) Describe the purpose of the first "proc autoreg" statement above. What was the conclusion? Using the backstep method we determined that the correct autocorrelation error structure for the model and then used Proc Autoreg (GLS) to determine if there is curvature in the trend of the data. Since the coefficient on the t^2 term was not statistically significant, we concluded that only the linear trend term, t, was needed to model the trend in the data.

(4)

c) Describe the purpose of the second "proc autoreg" statement above. What was the conclusion? Here, in the correct context of GLS, we tested the joint significance of the seasonal dummies using a joint F-test. We found the F-statistic to have a very small p-value leading to the conclusion that there is strong seasonality in the data. Since the residuals of the model were white noise we felt that the model of the data vis-a-vis the DTDS model was complete.

(4)

3. As we know, the DTDS model is written as

$$y_t = \alpha + \beta t + \theta t^2 + \gamma_1 D_{t1} + \gamma_2 D_{t2} + \gamma_3 D_{t3} + \cdots + \gamma_{12} D_{t,12} + \varepsilon_t$$

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \cdots + \rho_r \varepsilon_{t-r} + a_t$$

(2) a) The seasonal part of the model is represented by $\gamma_1 D_{t1} + \gamma_2 D_{t2} + \cdots + \gamma_{12} D_{t,12}$

(2) b) The trend part of the model is represented by $\beta t + \theta t^2$.

(2) c) The cyclical part of the model is represented by $\rho_1 \varepsilon_{t-1} + \cdots + \rho_r \varepsilon_{t-r}$

(2) d) The irregular part of the model is represented by a_t.

e) In our work with the Plano Tax Revenue data we wound up dropping the D_{t1} variable from our model? Why? We wanted to avoid the "dummy variable" trap of having perfectly collinear explanatory variables in the model. We did so by dropping one of the dummy variables from the equation. This is called the "relative-to-January" parameterization.

(2)

f) Suppose you wanted to test for the lack of curvature in your data. What would your null hypothesis be? What would your alternative hypothesis be? $H_0: \theta = 0$, $H_1: \theta \neq 0$.

(2)

g) Suppose you wanted to test for the lack of trend altogether in your data. What would your null hypothesis be? What would your alternative hypothesis be?

$H_0: \beta = \theta = 0$ $H_1: \beta \neq 0, \theta \neq 0$ or both

(2)

h) Suppose you wanted to test for the lack of seasonality in your data. What would your null hypothesis be? What would your alternative hypothesis be?

$H_0: \gamma_2 = \gamma_3 = \cdots = \gamma_{12} = 0$ H_1: at least one of the γ's is non-zero.

(2)

i) Suppose you wanted to test for the lack of cyclical effects in your data. What would your null hypothesis be? What would your alternative hypothesis be?

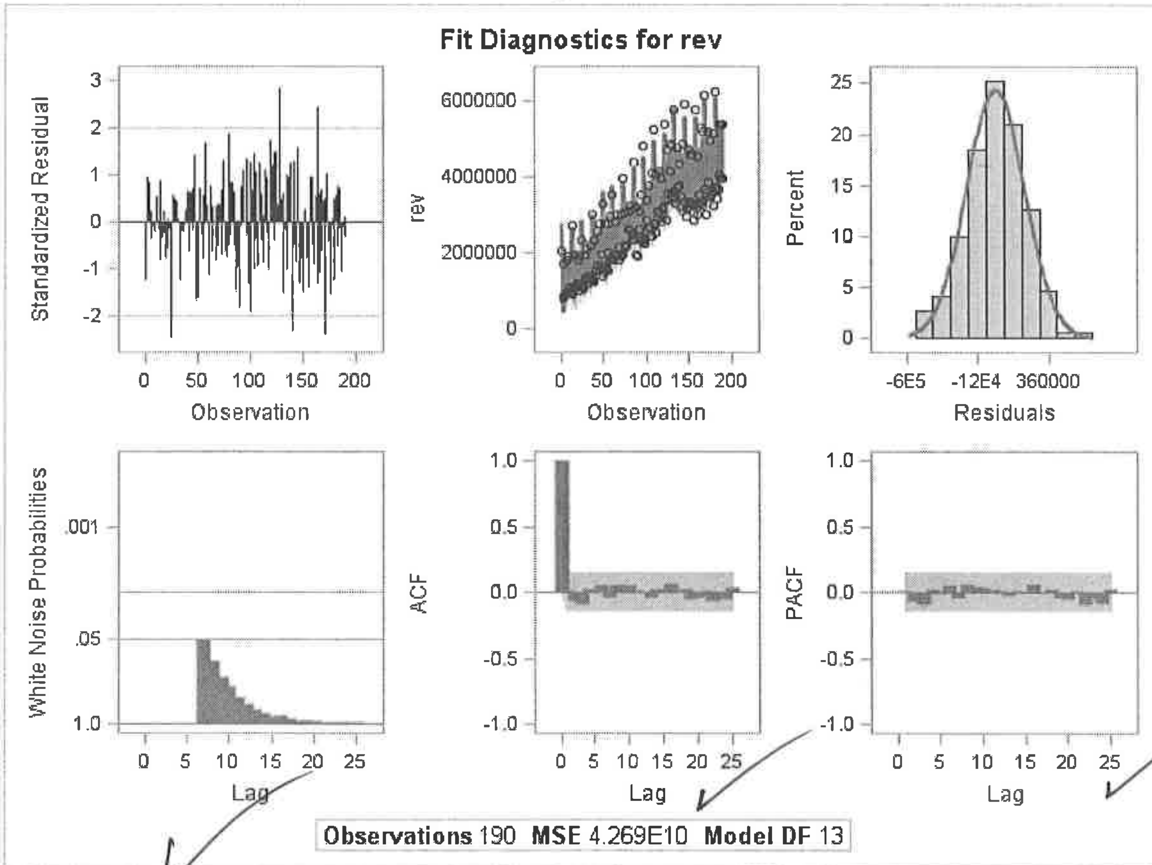
$H_0: \rho_1 = \rho_2 = \cdots = \rho_r = 0$

(2) H_1: At least one of the ρ's is non-zero.

the Box-Pierce-Ljung Q statistics all have probability values that are greater than 0.05 at the various lag lengths.

j) What are the implications of the following diagrams produced by the second proc autoreg statement above? Fully explain what you see in the relevant graphs.

(4) Three of the below graphs indicate that the GLS residuals of the fitted model are white noise indicating that no additional variables need to be added to the model. The sample autocorrelations and sample partial autocorrelations are all insignificantly different from zero. In addition,



4. Briefly describe to me how we went about building a "best" UCM model for the Plano Sales Tax Revenue data.

(4) We looked at an increasingly complex set of UCM time series models ranging from the Basic structural model (BSM) (stochastic level, stochastic slope, and stochastic dummy season) to increasing complex variants of the BSM. We choose the model that minimized the AIC and BIC information measures and, at the same time, had white noise residuals.

5. Briefly explain to me why the concepts of Stationarity and Invertibility are important considerations in build Box-Jenkins time series models. Stationarity of a time series allows one realization of the data to provide consistent estimation of the population mean, variance and covariance function of the data. The invertibility condition disallows a time series representation that weighs far past data more than near past data in determining the current variation in the time series.

(4)

6. Write out the conditions for a time series y_t to be stationary.

(4)

i) $E(y_t) = \mu_y \quad \forall t$ (constant mean)

ii) $Var(y_t) = \sigma_y^2 \quad \forall t$ (constant variance)

iii) $cov(y_t, y_{t-j}) = \gamma_j \quad \forall t$ and any given $j = 1, 2, \ldots$ (constant covariance)

7. Write out the minimum mean square error forecasting equations for the following models:

a) ARMA(0,0)

(2) $\hat{y}_{t+h} = \bar{y} \quad \forall h = 1, 2, \cdots$

b) ARMA(0,1)

(2) $\hat{y}_{t+h} = \bar{y} - \hat{\theta}_1 \cdot (\hat{a}_t)$ for $h = 1$

$\bar{y}$ for $h \geq 2$.

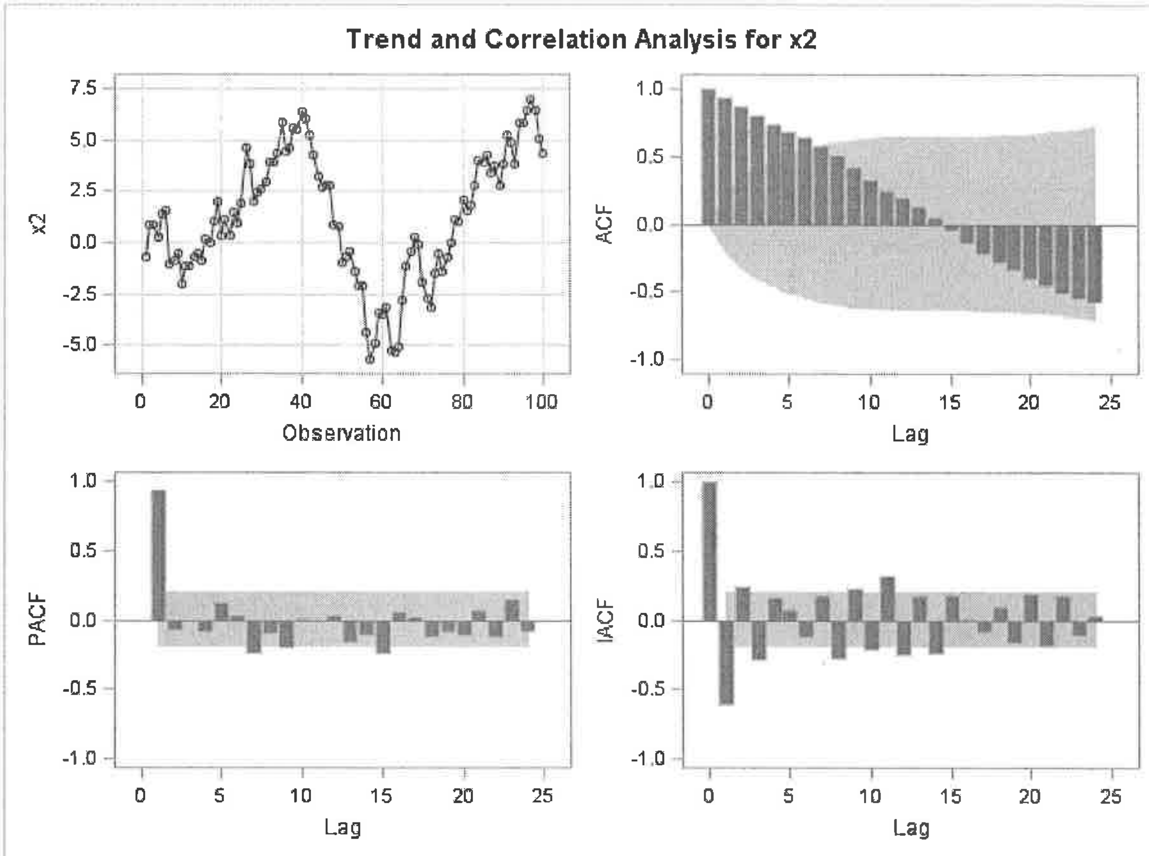
c) ARMA(1,0)

(2) $\hat{y}_{t+h} = \bar{y} + \hat{\phi}_1^h (y_t - \bar{y})$, $h = 1, 2, \cdots$

8. Suppose you had a data set that produced the following plot, sample ACF, and sample PACF. Would you say that the time series is stationary? Why or why not?

(4)

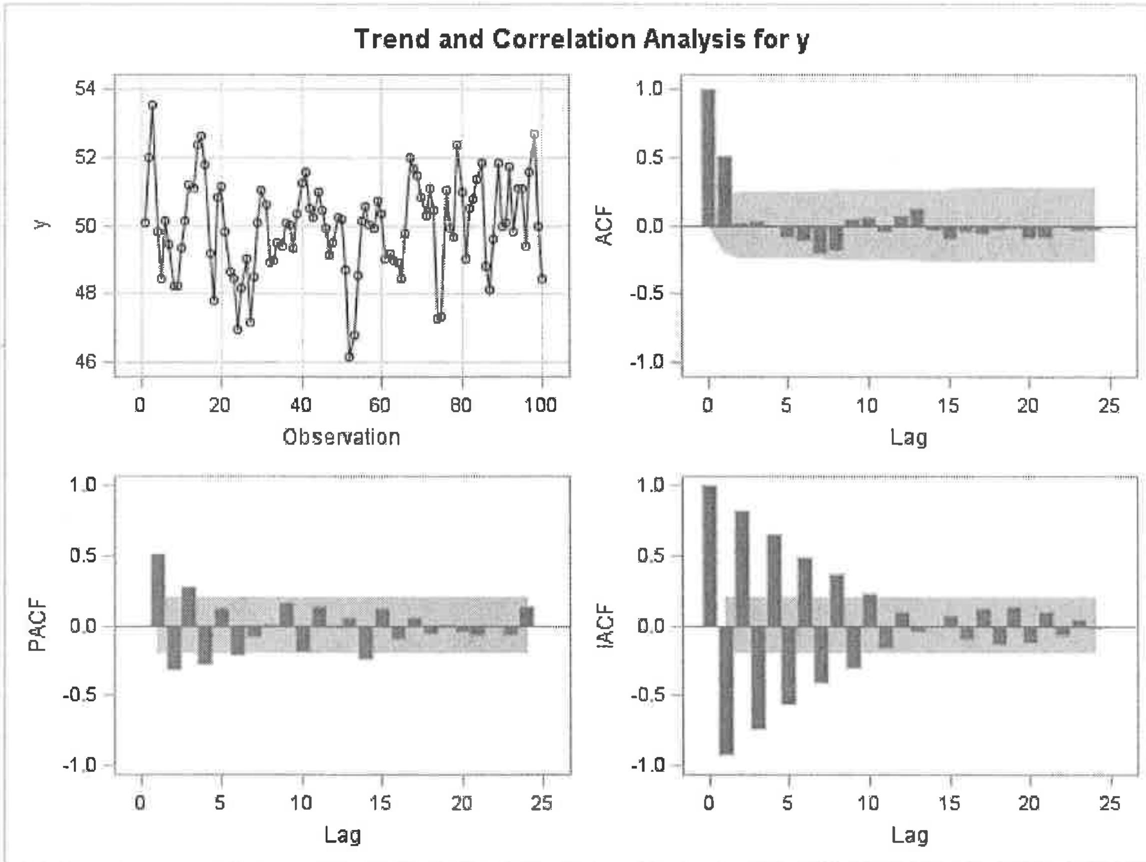
The data appear not to be stationary because (i) the data are slow-turning around a sample mean line that we could put through the data, (ii) the sample autocorrelation function is very slowly damping down to zero and (iii) PACF at lag one is near one.
All of these conditions are emblematic of data that are not stationary and that need to be differenced before analysis.



9. Suppose you were given the below graphical information presented in Computer Output #1. What would be your tentative identification of the appropriate Box-Jenkins model for the data? Fully explain your answer. P = 0 . Q = 1 .

Explanation: The sample acf has one significant spike in it (don't count the definitional 1.0 at lag zero) and then it cuts off while the sample pacf is damping out (tailing off). This is emblematic of an ARMA(0,1) model.

(4)



10. Using **Computer Output # 1**, fill in the following P-Q box. Assume the data being analyzed is **monthly** data. 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.

Q

P	0	1	2
0	351.6999 354.3051 44.39 (0.0069)	285.6983 290.9087 18.90 (0.7072)	287.571 295.3865 19.81 (0.5957)
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)		

(12)

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

(2) Reasoning: The model of choice, given the P-Q box, is the ARMA(0,1) model because it has the smallest AIC and SBC goodness-of-fit measures and the residuals of the model are white noise.

11. Use **Computer Output #1** 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.

(4)
Overfitting Model 1 is ARMA(0 , 2).
The overfitting coefficient is −0.04405.
The T-statistic of the overfitting coefficient is −0.44.
Therefore the overfitting coefficient from this model is statistically (significant/**insignificant**). Circle one alternative.

(4)
Overfitting Model 2 is ARMA(1 , 1).
The overfitting coefficient is 0.03726.
The T-statistic of the overfitting coefficient is 0.35.
Therefore the overfitting coefficient from this model is statistically (significant/**insignificant**). Circle one alternative.

(2) My conclusion is the ARMA(0,1) model is the best model because both overfitting coefficients are statistically insignificant.

12. In the below space write out the final model that you have chosen for the Y time series in **Computer Output # 1** with accompanying t-statistics, standard errors, goodness-of-fit measures, and a test statistic for white noise residuals with accompanying p-value. (You can report your estimated model either in the intercept-form or the deviation-from-mean form.)

(4)

Intercept form:

$$y_t = 49.98908 - (-0.95652)\hat{a}_{t-1} + \hat{a}_t$$
(0.03345)
[t = −28.60]

Deviation-from-Mean Form:

$$y_t - 49.98908 = \hat{a}_t - (-0.95652)\hat{a}_{t-1}$$
(0.03345)
[t = −28.60]

$Q_{24} = 18.90$ ($p = 0.7072$)
AIC = 285.6983 SBC = 290.9087

Computer Output # 1

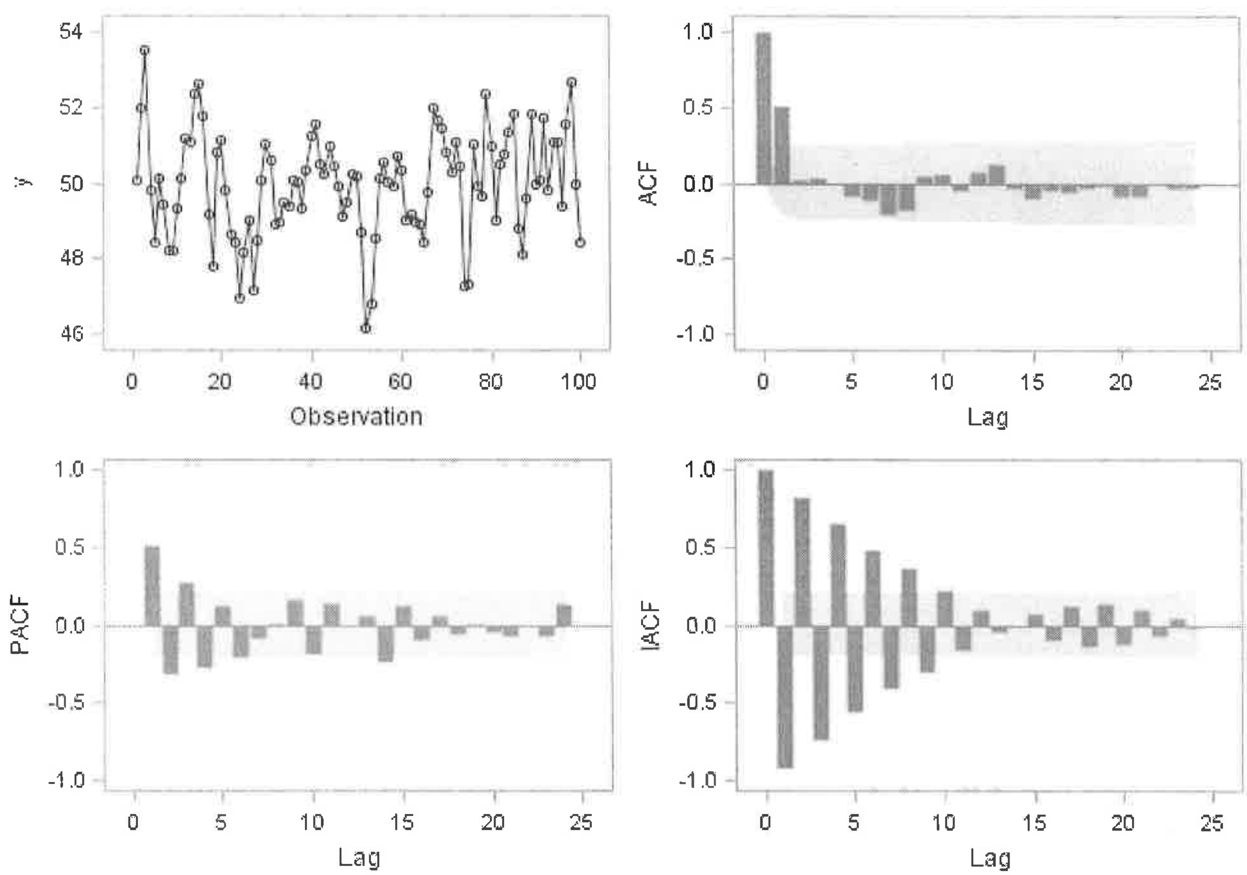
The SAS System

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					
				Approx	

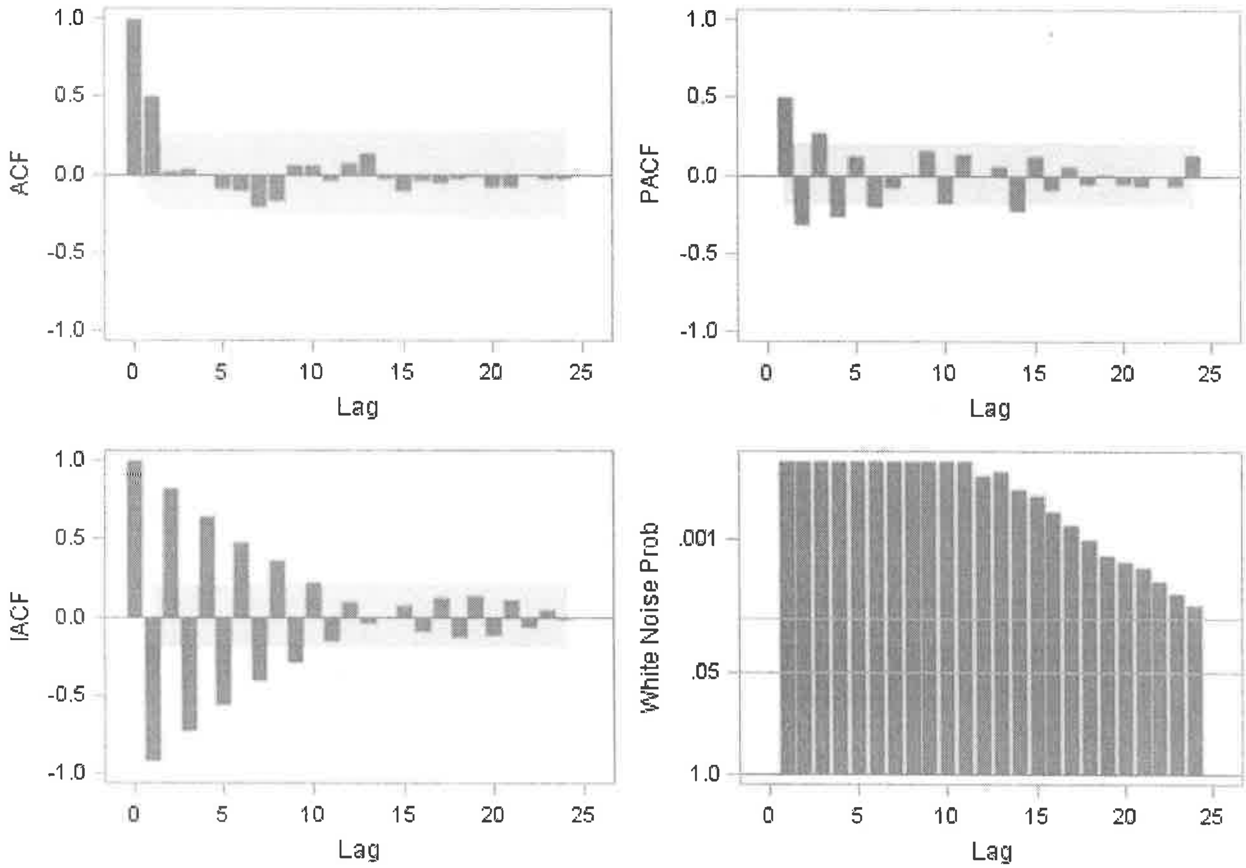
Parameter	Estimate	Standard Error	t Value	Pr > |t|	Lag
MU	49.96876	0.13974	357.59	<.0001	0

Constant Estimate	49.96876
Variance Estimate	1.952621
Std Error Estimate	1.397362
AIC	351.6999
SBC	354.3051
Number of Residuals	100

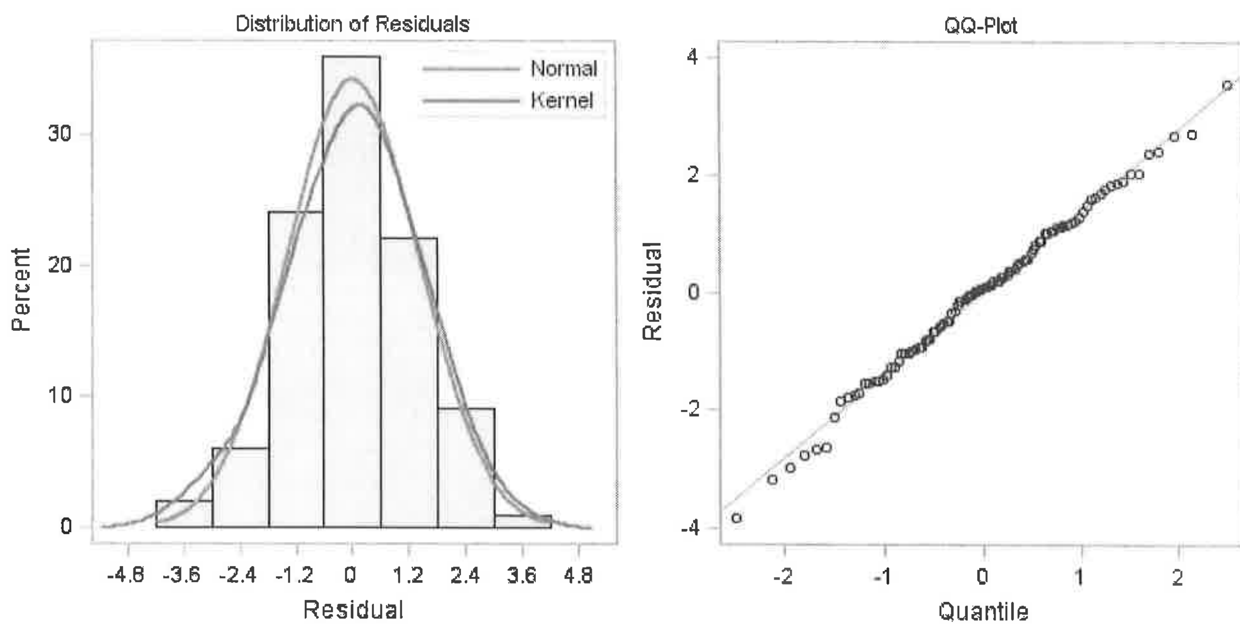
*** AIC and SBC do not include log determinant.**

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

Residual Correlation Diagnostics for y



Residual Normality Diagnostics for y



Model for variable y

Estimated Mean	49.96876

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

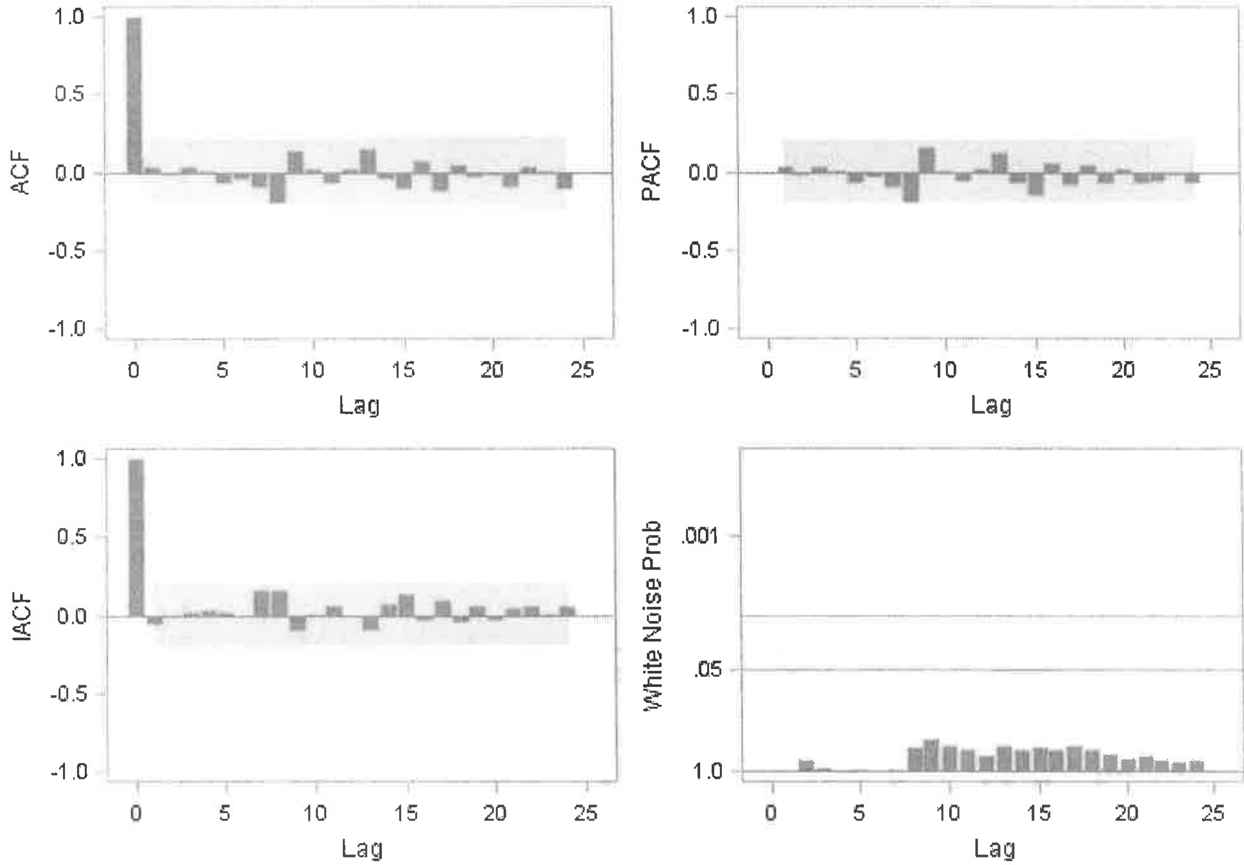
Constant Estimate	49.98908
Variance Estimate	0.999309
Std Error Estimate	0.999654
AIC	285.6983
SBC	290.9087
Number of Residuals	100

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

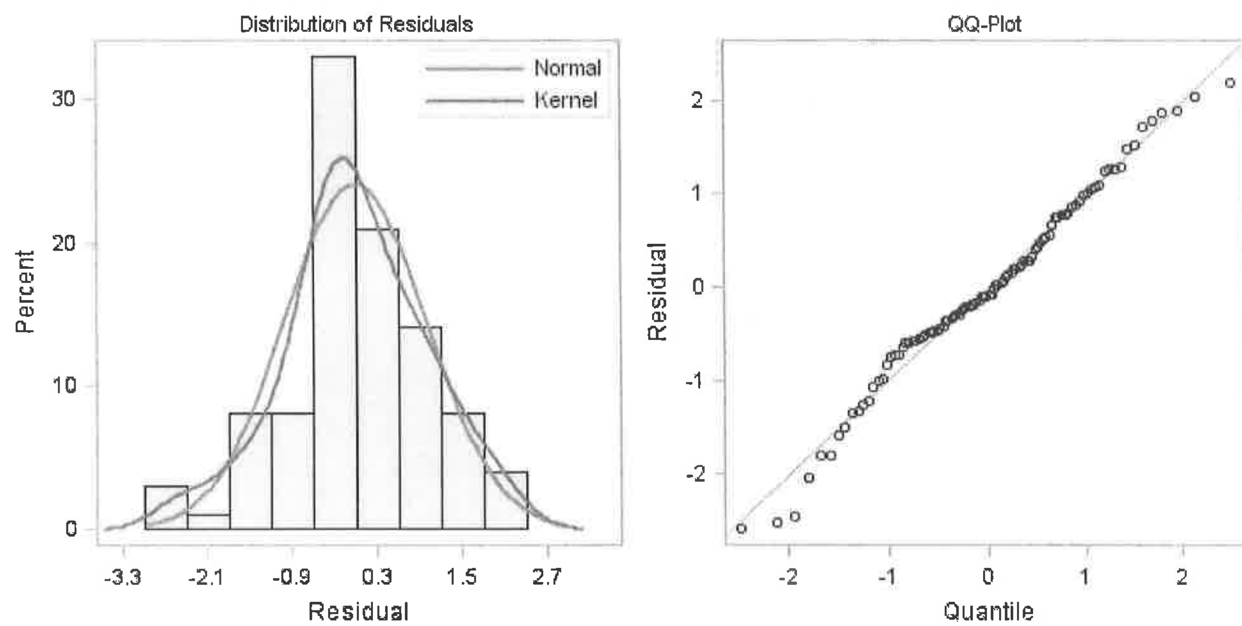
Correlations of Parameter Estimates		
Parameter	MU	MA1,1
MU	1.000	-0.046
MA1,1	-0.046	1.000

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

Residual Correlation Diagnostics for y



Residual Normality Diagnostics for y



Model for variable y

Estimated Mean	49.98908

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

Warning: Estimates did not improve after a ridge was encountered in the objective function. The iteration process has been terminated.

Warning: Estimates may not have converged.

ARIMA Estimation Optimization Summary	
Estimation Method	Conditional Least Squares
Parameters Estimated	3
Termination Criteria	Maximum Relative Change in Estimates
Iteration Stopping Value	0.001
Criteria Value	1.76E-15
Maximum Absolute Value of Gradient	0.075249
R-Square Change from Last Iteration	0.000517
Objective Function	Sum of Squared Residuals
Objective Function Value	97.80768
Marquardt's Lambda Coefficient	1E12
Numerical Derivative Perturbation Delta	0.001
Iterations	15
Warning Message	Estimates may not have converged.

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

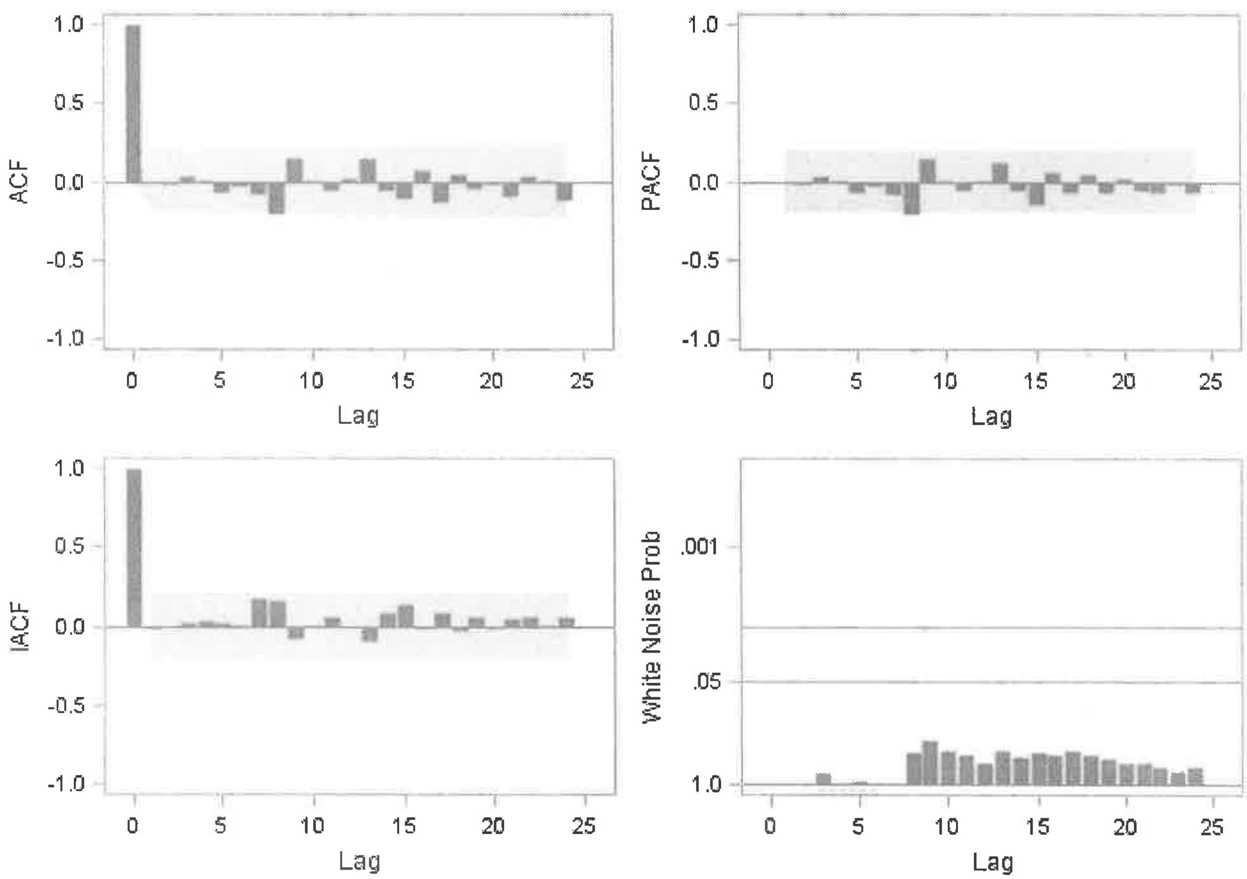
Number of Residuals	100

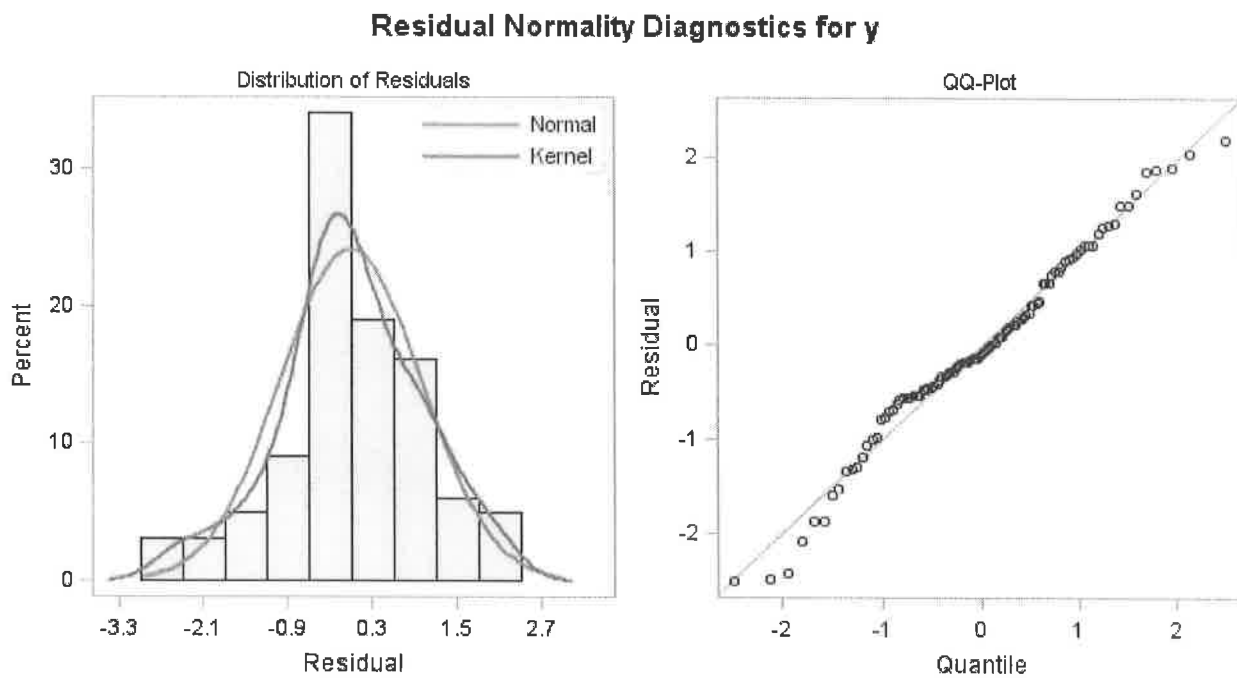
*** AIC and SBC do not include log determinant.**

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

Residual Correlation Diagnostics for y





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	49.95608	0.24369	205.00	<.0001	0
AR1,1	0.51126	0.08752	5.84	<.0001	1

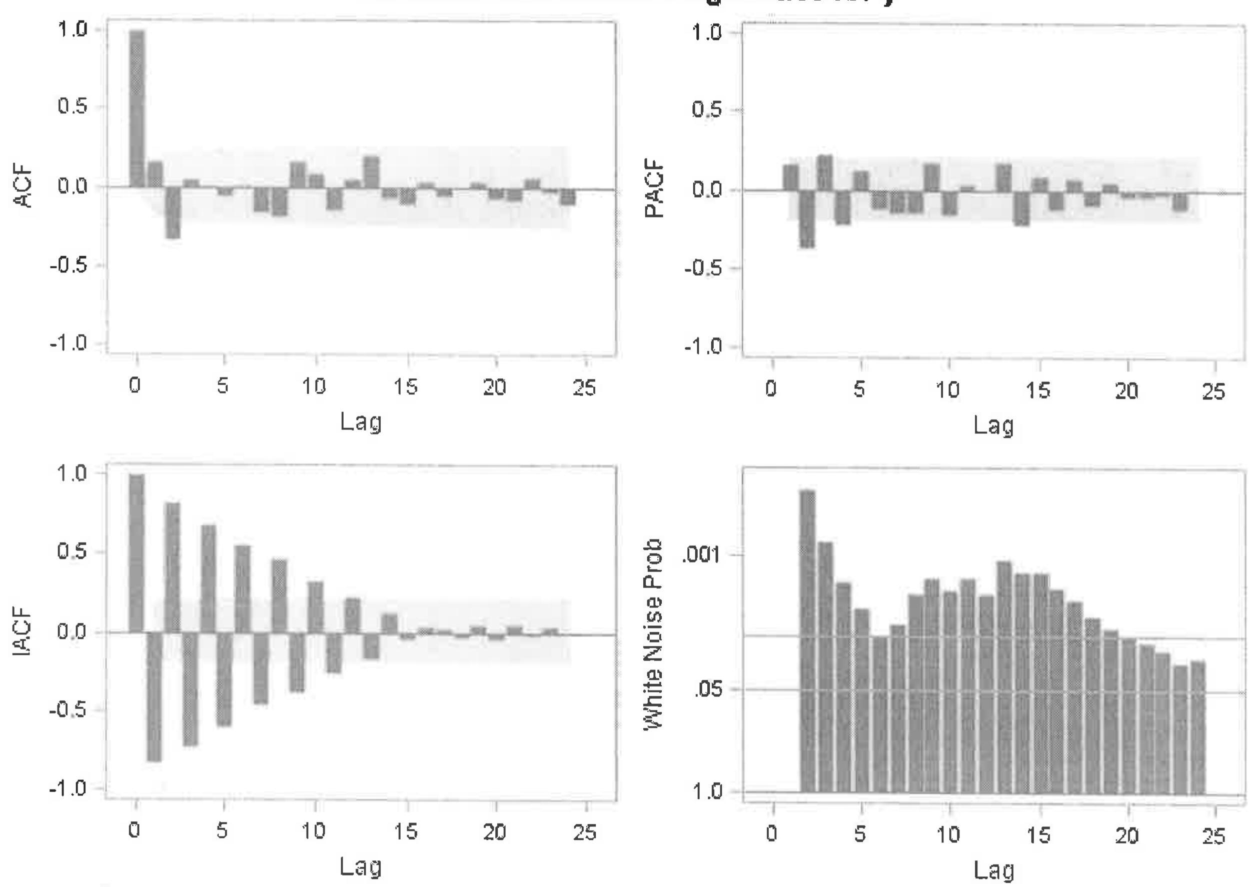
Constant Estimate	24.41574
Variance Estimate	1.463036
Std Error Estimate	1.20956
AIC	323.8188
SBC	329.0292
Number of Residuals	100

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

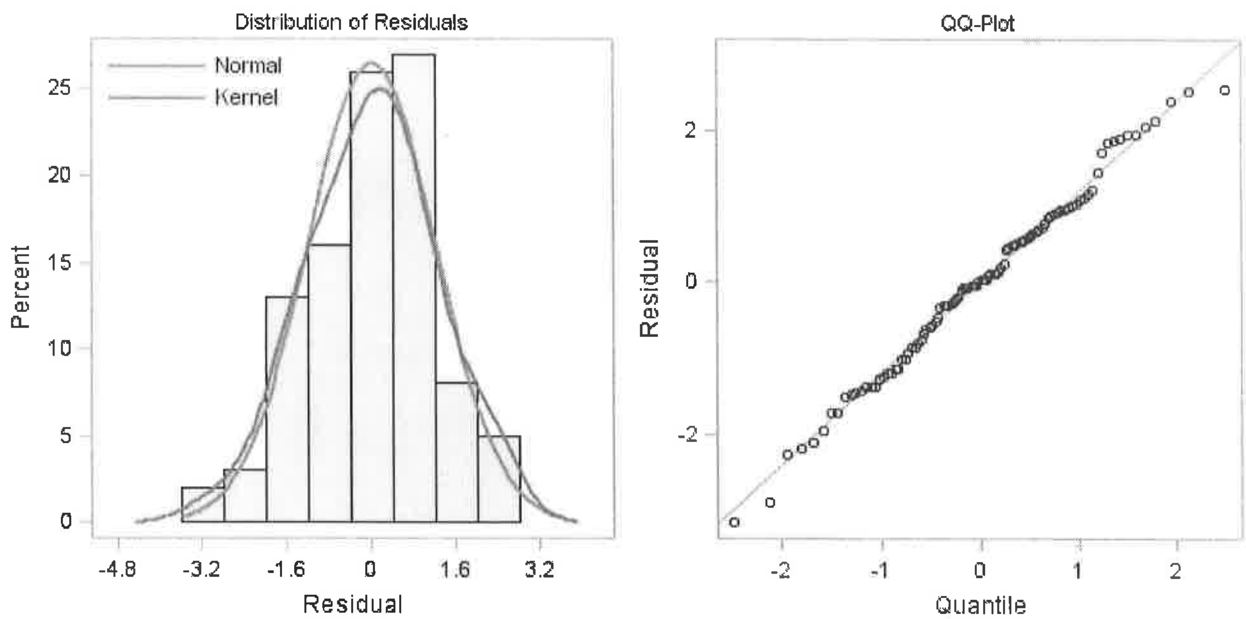
Correlations of Parameter Estimates		
Parameter	**MU**	**AR1,1**
MU	1.000	-0.020
AR1,1	-0.020	1.000

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

Residual Correlation Diagnostics for y



Residual Normality Diagnostics for y



Model for variable y	
Estimated Mean	49.95608

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

Conditional Least Squares Estimation

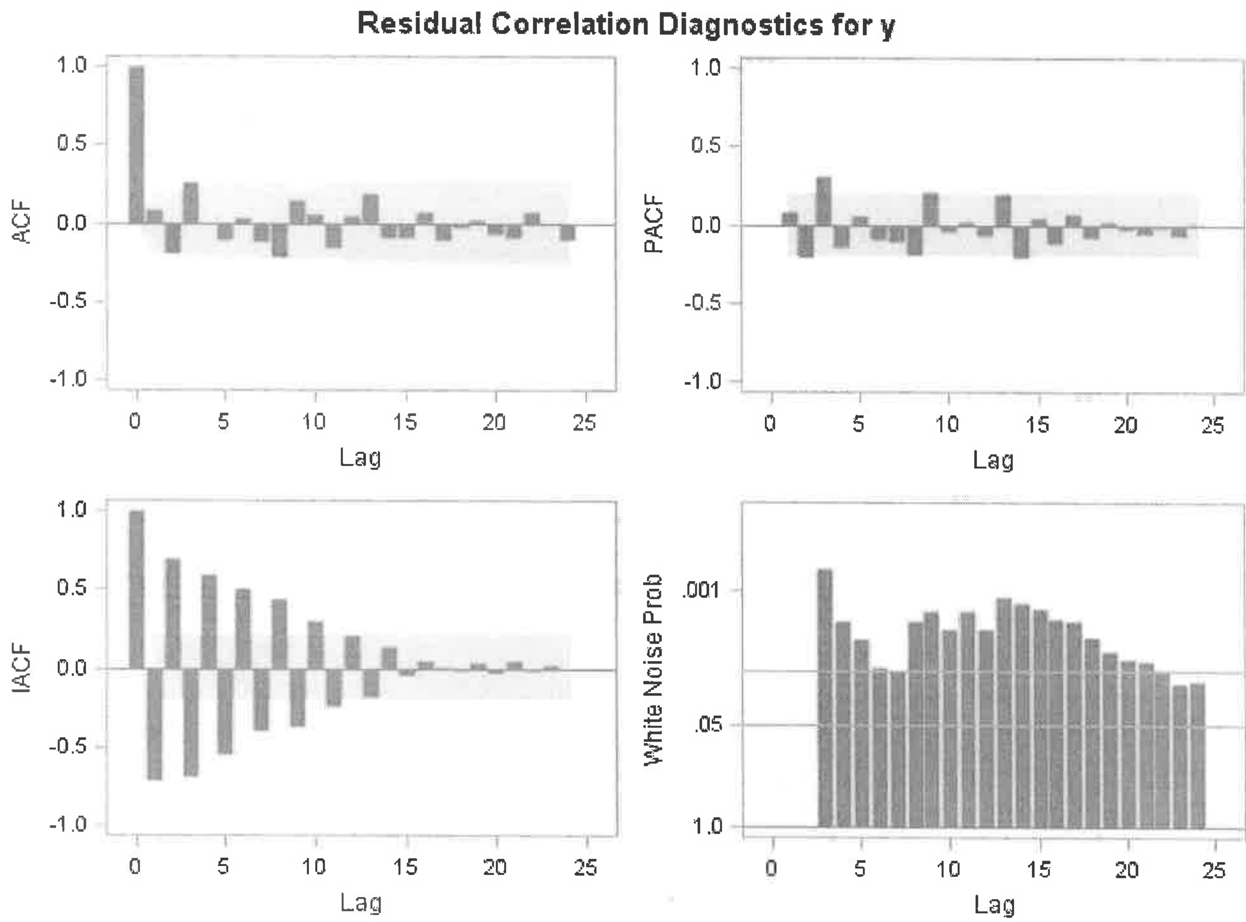
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	49.94664	0.17676	282.56	<.0001	0
AR1,1	0.67900	0.09679	7.01	<.0001	1
AR1,2	-0.32773	0.09680	-3.39	0.0010	2

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

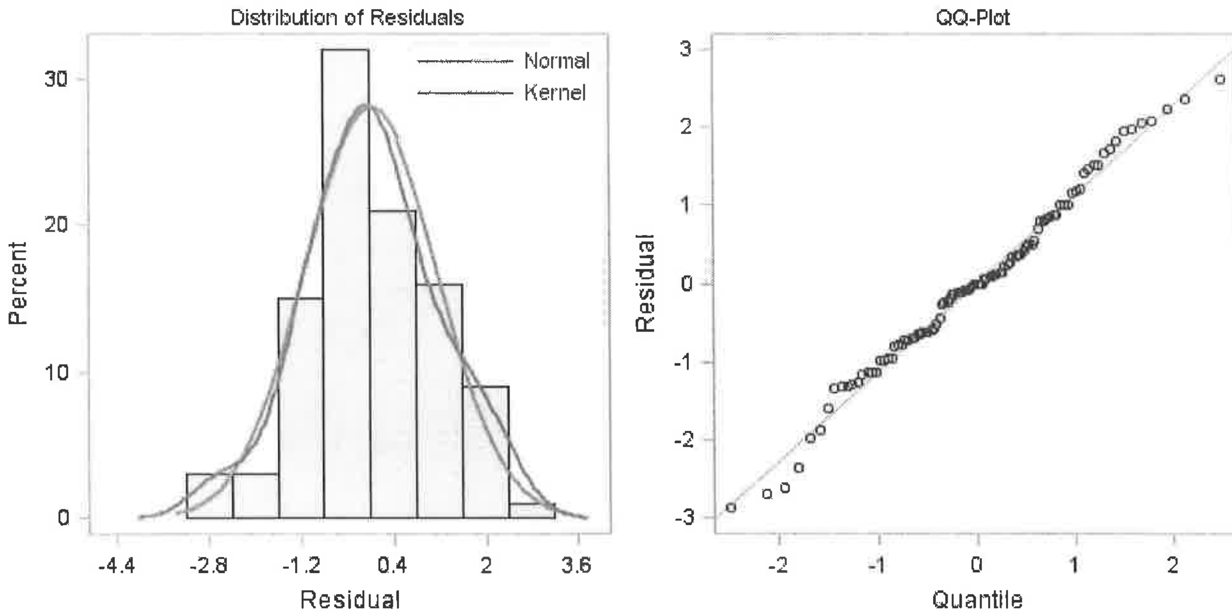
*** AIC and SBC do not include log determinant.**

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



Residual Normality Diagnostics for y



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

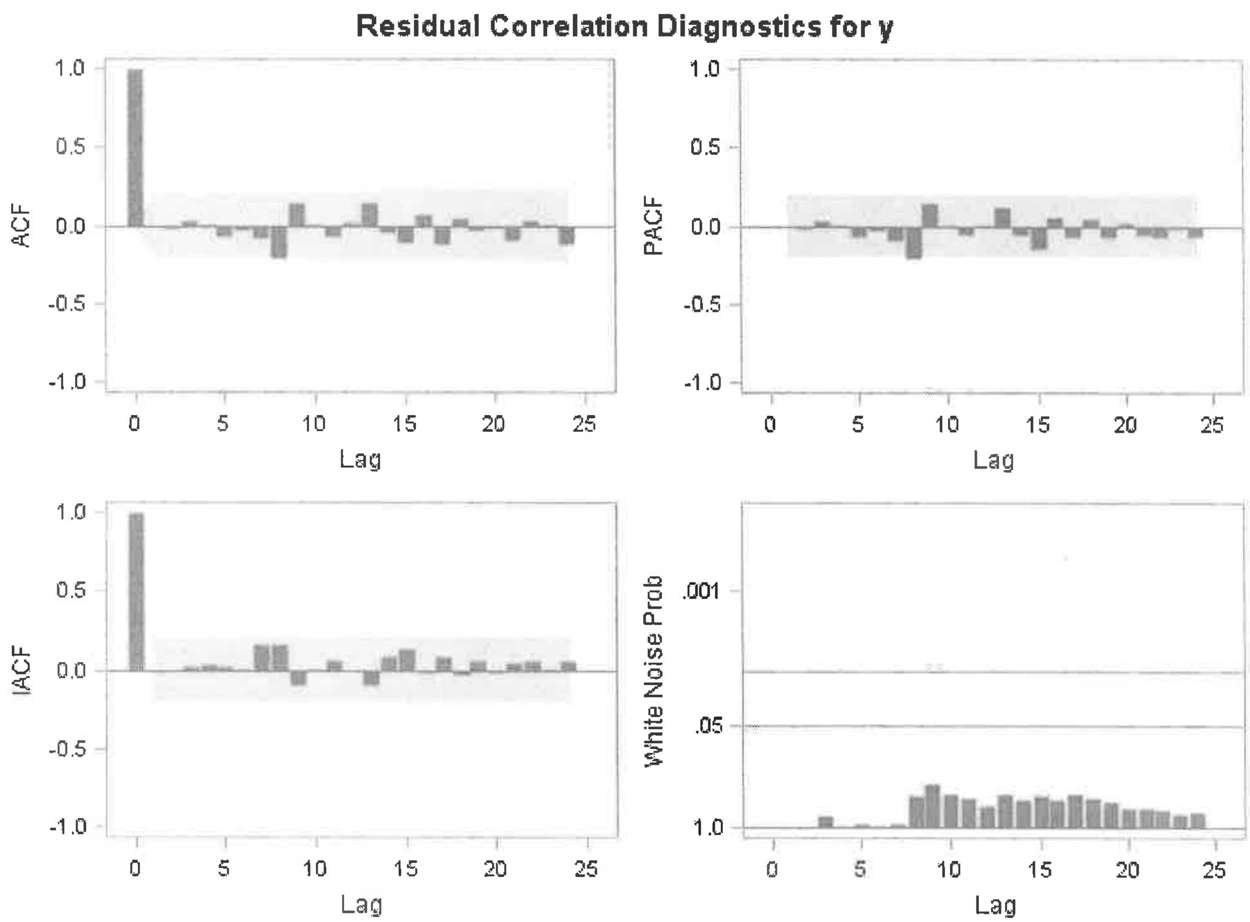
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	49.99101	0.19201	260.35	<.0001	0
MA1,1	-0.95435	0.03590	-26.59	<.0001	1
AR1,1	0.03726	0.10720	0.35	0.7289	1

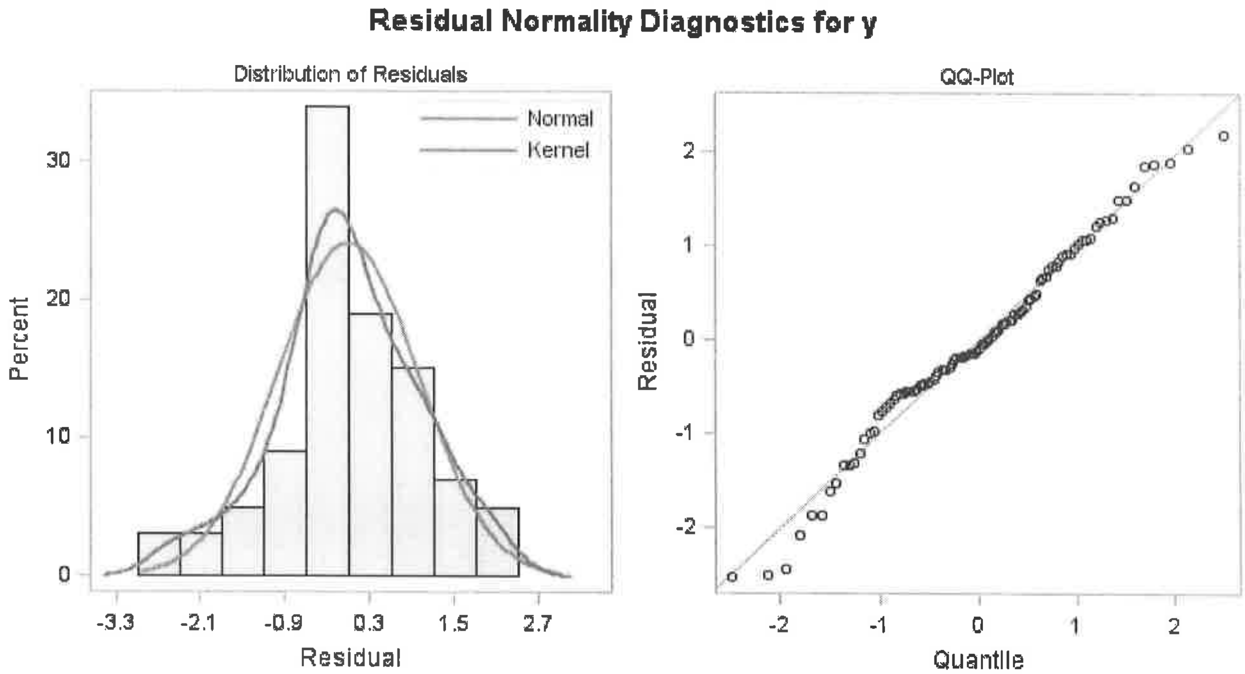
Constant Estimate	48.12834
Variance Estimate	1.008322
Std Error Estimate	1.004152
AIC	287.5705
SBC	295.386
Number of Residuals	100

*** 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)