

TEST #1

STA 4853

Name: _____

March 6, 2017

Please read the following directions.

DO NOT TURN THE PAGE UNTIL INSTRUCTED TO DO SO

Directions

- This exam is **closed book** and **closed notes**.
- There are 32 multiple choice questions.
- Circle the **single best** answer for each multiple choice question. Your choice should be made clearly.
- Always **circle the correct response**. (Sometimes the question has an empty blank or a box, but this is **NOT** where the answer goes.)
- There is no penalty for guessing.
- The exam has **13** pages.
- Each question is worth equal credit.

For each of the AR or MA parameters in a time series model, SAS PROC ARIMA reports an estimate, standard error, t -value, and p -value. The p -value is the approximate probability of getting a _____ by chance whose magnitude is as large as that observed when the true value of the _____ is zero.

Problem 1. Which of the following words correctly fills in the **first** blank above?

- a) estimate b) t -value c) variance estimate d) Ljung-Box test e) AIC
f) SBC g) standard error h) p -value i) autocorrelation j) parameter

Problem 2. Which of the following words correctly fills in the **second** blank above?

- a) p -value b) autocorrelation c) parameter d) variance estimate e) Ljung-Box test
f) AIC g) SBC h) estimate i) standard error j) t -value

Problem 3. For a random shock process, the theoretical ACF _____

- a) will have nearly all values within the two standard error band
b) will have an approximate cutoff after lag 1
c) will have an exact cutoff after lag 1
d) is non-stationary
e) is independent of future values
f) is exactly zero for all nonzero lags
g) decays rapidly to zero

Problem 4. Suppose $a_1, a_2, a_3, \dots$ are random shocks with mean zero and variance σ_a^2 , and we define

$$\begin{aligned}\tilde{z}_5 &= \psi_0 a_5 + \psi_1 a_4 + \psi_2 a_3 \\ \tilde{z}_3 &= \psi_0 a_3 + \psi_1 a_2 + \psi_2 a_1.\end{aligned}$$

What is the value of $E\tilde{z}_5\tilde{z}_3$?

- a) $\sigma_a^2(\psi_0\psi_1 + \psi_1\psi_2)$ b) $\sigma_a^2\psi_0\psi_2$ c) $\sigma_a^2(\psi_0^2 + \psi_1^2 + \psi_2^2)$
d) ψ_1^2 e) ψ_1^3 f) ψ_1^4 g) $\frac{\sigma_a^2}{1 - \phi_1^2}$ h) $\frac{C}{1 - \phi_1}$ i) ϕ_1^3

Problem 5. To generate a realization from an $AR(p)$ process, we need _____ starting values.

- a) 0 b) 1 c) 2 d) ϕ_1 e) ϕ_{11}
f) ϕ_p g) ϕ_{pp} h) p i) $p + 1$ j) $p - 1$

Problem 6. Suppose you fit a regression model and then construct a plot of the residuals versus fitted values (also known as the predicted values). In this plot the residuals are plotted on the y -axis and fitted values on the x -axis. If the regression assumptions are valid, we expect roughly that _____

- a) the plot will follow a straight line
- b) the fitted values will be normally distributed
- c) the residuals will decay to zero
- d) 95% of the fitted values will be between -2 and 2
- e) positive residuals will tend to be followed by positive residuals
- f) the residuals will form a band centered at zero with constant vertical width
- g) positive residuals will tend to be followed by negative residuals

Problem 7. Approximate 95% confidence intervals for a regression coefficient β_i use the value 1.96. If the regression assumptions are valid, an **exact** 95% confidence interval may be constructed by replacing 1.96 by _____

- a) a value obtained from the normal distribution
- b) a value obtained from a chi-square distribution
- c) a value obtained from a t -distribution
- d) 2
- e) 1.959964
- f) 1.644854
- g) 2.326348

Problem 8. In the SAS PROC REG output, the “Corrected Total Sums of Squares” is _____

- a) the total variability in the sample responses Y_i
- b) the total variability in the sample covariates X_i
- c) the total variability of the residuals ε_i
- d) the total variability of the fitted (predicted) values $\hat{Y}_i$
- e) equal to R-squared (R^2)
- f) equal to the Error Mean Square
- g) equal to the Root MSE
- h) equal to the Adjusted R-squared

Problem 9. Suppose z_t is an AR(2) process: $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t$. What is the correlation between z_{t-2} and a_t ?

- a) ϕ_2
- b) ϕ_{22}
- c) ρ_2
- d) ϕ_1^2
- e) 0
- f) 1
- g) -1
- h) ϕ_2^2

Problem 10. In the regression output, if the estimate $\hat{\beta}_i$ has a t -value t_i which is large in magnitude, then the associated p -value will be _____

- a) greater than 2 b) greater than the standard error c) less than the standard error
d) greater than 1 e) less than 0 f) close to 0 g) close to 1

Problem 11. Let $X_1, X_2, \dots, X_n$ be a random sample. The sample standard deviation s_x is given by _____

- a) $\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})}$ b) $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})$
c) $\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ d) $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
e) $\frac{1}{n} \sum_{i=1}^n X_i$ f) $\sqrt{\frac{1}{n} \sum_{i=1}^n X_i}$
g) $\frac{1}{n} \sum_{i=1}^n X_i^2$ h) $\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}$

Problem 12. For a stationary AR(1) process

$$z_t = C + \phi_1 z_{t-1} + a_t,$$

it is true that $\text{Var}(z_t) =$ _____

- a) $\text{Var}(C) + \phi_1 \text{Var}(z_{t-1}) + \text{Var}(a_t)$ b) $C + \phi_1 \text{Var}(z_{t-1}) + \text{Var}(a_t)$
c) $\phi_1^2 \text{Var}(z_{t-1}) + \text{Var}(a_t)$ d) $\phi_1 \text{Var}(z_{t-1}) + \text{Var}(a_t)$
e) $C^2 + \phi_1^2 \text{Var}(z_{t-1}) + a_t$ f) $\phi_1 \text{Var}(z_{t-1}) + a_t$
g) $\text{Var}(C) + \phi_1 \text{Var}(z_{t-1}) + a_t$ h) $C + \phi_1 \text{Var}(z_{t-1}) + a_t$

Problem 13. For a stationary AR(1) process, which of the following is true for all $k \geq 1$?

- a) $\rho_k = C^2 \rho_{k-1}$ b) $\rho_k = \sigma_a^2 \rho_{k-1}$
c) $\phi_{kk} = \phi_1 \rho_{k-1}$ d) $\phi_{kk} = C \rho_{k-1}$
e) $\phi_{kk} = \sigma_a \rho_{k-1}$ f) $\phi_{kk} = \phi_1^2 \rho_{k-1}$
g) $\phi_{kk} = C^2 \rho_{k-1}$ h) $\phi_{kk} = \sigma_a^2 \rho_{k-1}$
i) $\rho_k = \phi_1 \rho_{k-1}$ j) $\rho_k = C \rho_{k-1}$
k) $\rho_k = \sigma_a \rho_{k-1}$ l) $\rho_k = \phi_1^2 \rho_{k-1}$

Problem 14. Suppose you perform a regression using SAS PROC REG, and you specify the DWPROB option in the MODEL statement. Then, in addition to the Durbin-Watson statistic, SAS will display the first order autocorrelation of the residuals $e_1, e_2, \dots, e_n$. What is the formula for the first order autocorrelation of the residuals?

- a) $\frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2}$ b) $\frac{\sum_{t=1}^n e_t^2}{\sum_{t=2}^n e_t e_{t-1}}$ c) $\sum_{t=2}^n e_t e_{t-1}$
- d) $\sum_{t=2}^n (e_t - \bar{e})(e_{t-1} - \bar{e})$ e) $\sum_{t=1}^{n-1} (e_t - \bar{e})(e_{t+1} - \bar{e})$ f) $\sum_{t=1}^n (e_t - \bar{e})^2$
- g) $\sum_{t=1}^{n-1} e_t e_{t+1}$ h) $\sum_{t=1}^n e_t^2$ i) $\frac{\sum_{t=1}^n e_t^2}{\sum_{t=1}^{n-1} e_t e_{t+1}}$
-

MINIC is a _____ a time series. It finds the values of p and q of the ARMA(p, q) model which minimizes an estimate of the _____.

Problem 15. Which of the following phrases correctly fills in the **first** blank above?

- a) method for estimating the parameters of an ARMA(p, q) model for
b) method to help identify the orders p and q of an appropriate ARMA model for
c) test of the normality of the residuals obtained by fitting an ARMA(p, q) model to
d) test of the significance of an ARMA(p, q) model for
e) test of the residual ACF when fitting an ARMA(p, q) model to
f) test of the stationarity of an ARMA(p, q) model for

Problem 16. Which of the following choices correctly fills in the **second** blank above?

- a) Likelihood b) Variance Estimate c) Standard Error Estimate d) SBC
e) PACF f) Constant Estimate g) Chi-Square h) ACF

Problem 17. Suppose you have a time series $z_1, z_2, \dots, z_n$ with a very large value of n (say, in the millions) which has been generated by some stationary ARMA process. If you use OLS (ordinary least squares) to fit the regression model

$$z_t = \beta_0 + \beta_1 z_{t-1} + \beta_2 z_{t-2} + \beta_3 z_{t-3} + \varepsilon_t,$$

the parameter estimate $\hat{\beta}_3$ will be very close to _____

- a) ρ_2 b) ρ_3 c) ρ_4 d) σ_a^2 e) ϕ_{22} f) ϕ_{33}
g) ϕ_{44} h) μ_z i) σ_z^2 j) θ_3 k) θ_2 l) θ_4

Problem 18. Suppose we know the values of a time series $\{z_t\}$ at times $1, 2, \dots, n$, that is, we know $z_1, z_2, z_3, \dots, z_n$. The series $\{z_{t-5}\}$ (z_t lagged by 5) will have _____

- a) 4 missing values at the end of the series
- b) 5 missing values at the end of the series
- c) 4 missing values at the beginning of the series
- d) 5 missing values at the beginning of the series
- e) no missing values

Problem 19. For a stationary process $\{z_t\}$, the quantity $\text{Cov}(z_t, z_{t-k})$ is represented by the symbol _____

- a) ψ_k
- b) γ_k
- c) ϕ_{kk}
- d) θ_k
- e) ϕ_k
- f) ρ_k

Problem 20. If the values of X and Y are obtained for a random sample of n individuals: $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, the sample **covariance** between X and Y is defined to be _____

- a) $\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2 \cdot \sum_{i=1}^n (Y_i - \bar{Y})^2}$
- b) $\frac{1}{n} \sum_{i=1}^n X_i Y_i$
- c) $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 (Y_i - \bar{Y})^2$
- d) $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$
- e) $\frac{c(X, Y)}{s_x s_y}$
- f) $\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \cdot \sum_{i=1}^n (Y_i - \bar{Y})^2}}$

Problem 21. An ARMA(2,1) process is defined by _____

- a) $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t - \theta_1 a_{t-1}$
- b) $z_t = C + \phi_1 z_{t-1} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$
- c) $z_t = C + \phi_1 z_1 + \phi_2 z_2 + a_t - \theta_1 a_1$
- d) $z_t = C + \phi_1 z_1 + a_t - \theta_1 a_1 - \theta_2 a_2$
- e) $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} - \theta_1 a_{t-1}$
- f) $z_t = C + \phi_1 z_{t-1} - \theta_1 a_{t-1} - \theta_2 a_{t-2}$
- g) $z_t = C + \phi_1 z_1 + \phi_2 z_2 - \theta_1 a_1$
- h) $z_t = C + \phi_1 z_1 - \theta_1 a_1 - \theta_2 a_2$

Problem 22. An ARMA(p, q) process is stationary if and only if all the solutions of _____ the unit circle in the complex plane.

- a) $\phi(B) = 0$ lie strictly inside
- b) $\theta(B) = 0$ lie strictly inside
- c) $\theta(B)/\phi(B) = 0$ lie strictly inside
- d) $\phi(B) = 0$ lie strictly outside
- e) $\theta(B) = 0$ lie strictly outside
- f) $\theta(B)/\phi(B) = 0$ lie strictly outside
- g) $\phi(B) = 0$ lie on the boundary of
- h) $\theta(B) = 0$ lie on the boundary of
- i) $\theta(B)/\phi(B) = 0$ lie on the boundary of

Problem 23. The process

$$z_t = \phi_1 z_{t-1} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3}$$

can be written in backshift form as _____

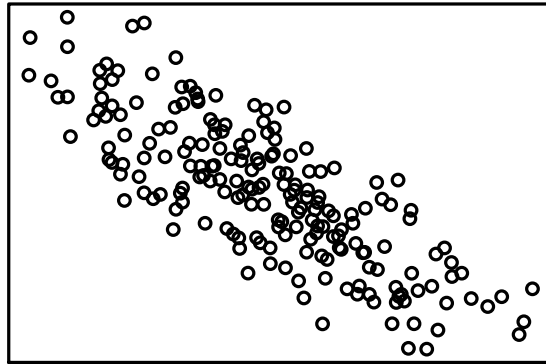
- a) $(1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)z_t = (1 - \phi_1 B)a_t$
- b) $(1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)z_t = (1 + \phi_1 B)a_t$
- c) $(1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3)z_t = (1 - \phi_1 B)a_t$
- d) $(1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3)z_t = (1 + \phi_1 B)a_t$
- e) $(1 + \phi_1 B)z_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)a_t$
- f) $(1 + \phi_1 B)z_t = (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3)a_t$
- g) $(1 - \phi_1 B)z_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3)a_t$
- h) $(1 - \phi_1 B)z_t = (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3)a_t$

Problem 24. In PROC ARIMA, the ML (maximum likelihood) method of estimation should be used _____

- a) whenever computational speed is very important
- b) as a precaution whenever the data are **not** normally distributed
- c) whenever the CLS method is too slow
- d) when the shocks a_t are independent and approximately normally distributed with constant variance
- e) only when the series is non-stationary
- f) only when all the parameters are statistically significant

Problem 25. For a stationary time series $z_1, \dots, z_n$, suppose the scatterplot of z_t versus z_{t-4} looks like the plot given below. Then you know _____

- a) the variance of the series z_t tends to increase with time t
- b) the variance of the series z_t tends to decrease with time t
- c) the residuals are **not** normally distributed
- d) the residuals **are** normally distributed
- e) the partial autocorrelation at lag 4 is positive
- f) the partial autocorrelation at lag 4 is negative
- g) the autocorrelation at lag 4 is positive
- h) the autocorrelation at lag 4 is negative
- i) the mean of the series z_t tends to increase with time t
- j) the mean of the series z_t tends to decrease with time t



Problem 26. In regression, large values of H (the leverage) identify _____

- a) covariates which can be dropped
- b) covariates with serial correlation
- c) covariates which should be retained
- d) serial correlation in the residuals
- e) cases with unusual response values
- f) influential cases in the data
- g) cases with unusual covariate values
- h) influential covariates in the model

Problem 27. Suppose you are analyzing a time series $z_1, z_2, \dots, z_n$ using SAS PROC ARIMA. In the output produced by the IDENTIFY statement, the values of the Ljung-Box test statistics $Q(6)$, $Q(12)$, $Q(18)$, $Q(24)$ are given along with their corresponding P -values. If $\{z_t\}$ is actually just a random shock sequence, then we expect that _____

- a) $H_0 : \rho_1 = \rho_2 = \dots = \rho_6 = 0$ will be rejected
- b) $H_0 : Q(6) = Q(12) = Q(18) = Q(24) = 0$ will be rejected
- c) all four P -values will be large
- d) all four P -values will be small
- e) $Q(6)$, $Q(12)$, $Q(18)$, $Q(24)$ will all be large
- f) most of the ACF values will lie outside the band
- g) most of the PACF values will lie outside the band

Suppose that, based on a random sample of **400** observations, you fit a regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ for a response variable Y on two covariates X_1 and X_2 . Use the information in the table below to answer the next two questions.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t -value	$\text{Pr} > t $
Intercept	1	4.0	0.5		$< .0001$
X1	1	40.0	10.0		$< .0001$
X2	1	24.0	16.0		.1336

Problem 28. The t -value entries in the table have been left blank. What is the t -value for the X1 row of the table?

- a) 1.5 b) 4.0 c) 16.0 d) .25 e) .0625 f) 2.0 g) 0.5

Problem 29. Assuming all the regression assumptions are valid, compute an approximate 95% confidence interval for β_1 .

- a) (20.4, 59.6) b) (38.04, 41.96) c) (39.951, 40.049) d) (9.804, 10.196)
e) (8.04, 11.96) f) (-68.4, 88.4) g) (39.02, 40.98) h) (6.08, 13.92)

Problem 30. A process is generated by

$$z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t - \theta_1 a_{t-1}.$$

What are the requirements for this process to be stationary?

- a) $|\phi_1| < 1, |\theta_1| < 1$
b) $|\phi_2| < 1, |\theta_1| < 1$
c) $|\phi_1| < 1$
d) $|\theta_1| < 1$
e) $|\phi_2| < 1, \phi_2 + \phi_1 < 1, \phi_2 - \phi_1 < 1, |\theta_1| < 1$
f) $|\phi_1| < 1, \phi_1 + \phi_2 < 1, \phi_1 - \phi_2 < 1, |\theta_1| < 1$
g) $|\phi_2| < 1, \phi_2 + \phi_1 < 1, \phi_2 - \phi_1 < 1$
h) $|\phi_1| < 1, \phi_1 + \phi_2 < 1, \phi_1 - \phi_2 < 1$

The last two problems use the three pages of SAS output attached to the end of this exam.

Problem 31. The first page of output has the title “IDENTIFY A GOOD MODEL”. This page gives the usual output produced by the IDENTIFY statement in PROC ARIMA for a series z_1 of length 200. Based on this output, a reasonable model for this series is _____

- | | | | |
|----------|--------------|--------------|--------------|
| a) MA(2) | b) MA(3) | c) MA(4) | d) AR(1) |
| e) AR(2) | f) ARMA(1,1) | g) ARMA(2,2) | h) ARMA(1,3) |

Problem 32. The last two pages of output have the title “TRY AR(1) and MA(3) MODELS”. SAS PROC ARIMA was used to fit an AR(1) and an MA(3) model to a time series. (This series is **not** the series from the previous problem.) These pages give parts of the output produced by the ESTIMATE statement for these two models. The first page gives some of the output from fitting the AR(1) model; the second page gives output from the MA(3) model.

Suppose you relied on the AIC or SBC to choose between these two models. Which of the following statements is true?

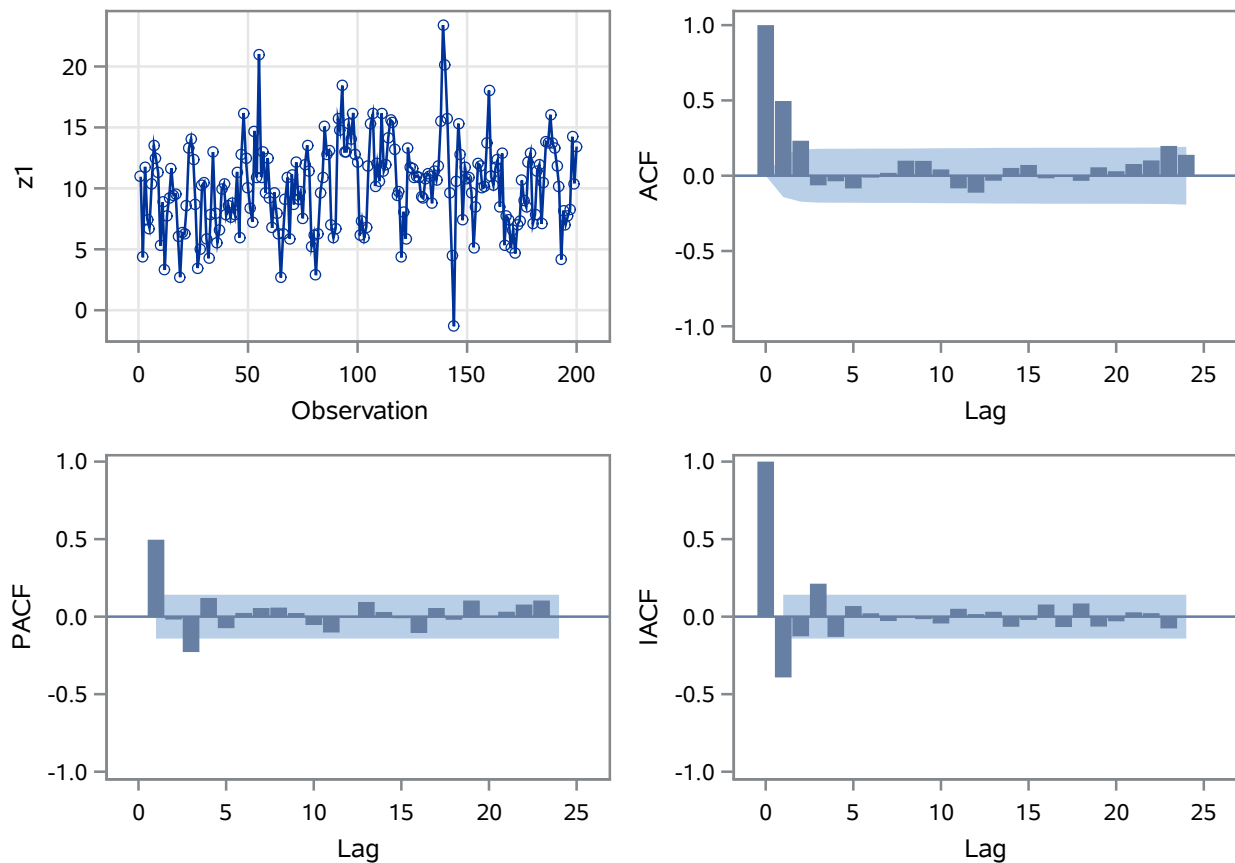
- a) AIC selects AR(1); SBC selects MA(3)
- b) AIC selects MA(3); SBC selects AR(1)
- c) Both AIC and SBC select MA(3)
- d) Both AIC and SBC select AR(1)
- e) Both AIC and SBC reject the null hypothesis
- f) Neither AIC or SBC reject the null hypothesis

The ARIMA Procedure

Name of Variable = z1	
Mean of Working Series	10.13957
Standard Deviation	3.629443
Number of Observations	200

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	63.62	6	<.0001	0.496	0.232	-0.064	-0.037	-0.084	-0.013
12	72.43	12	<.0001	0.019	0.100	0.098	0.042	-0.084	-0.113
18	74.69	18	<.0001	-0.033	0.051	0.071	-0.017	-0.005	-0.034
24	92.60	24	<.0001	0.056	0.030	0.078	0.101	0.198	0.139

Trend and Correlation Analysis for z1



The ARIMA Procedure

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	10.14649	0.20000	50.73	<.0001	0
AR1,1	0.51737	0.02709	19.10	<.0001	1

Constant Estimate	4.896957
Variance Estimate	9.339647
Std Error Estimate	3.056084
AIC	5074.455
SBC	5084.271
Number of Residuals	1000

Correlations of Parameter Estimates		
Parameter	MU	AR1,1
MU	1.000	0.002
AR1,1	0.002	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	11.65	5	0.0400	0.007	0.006	0.017	-0.096	-0.043	-0.010
12	17.23	11	0.1012	0.037	0.022	0.002	0.026	0.054	0.005
18	21.97	17	0.1859	-0.025	-0.029	-0.001	-0.018	0.053	-0.010
24	26.34	23	0.2851	-0.035	-0.002	0.011	0.045	0.012	-0.027
30	33.97	29	0.2404	0.055	-0.053	-0.017	-0.028	-0.020	-0.008
36	39.96	35	0.2591	-0.052	0.006	-0.012	0.007	-0.036	-0.040
42	46.21	41	0.2658	-0.000	-0.066	-0.011	0.038	0.003	0.002
48	59.58	47	0.1031	-0.006	-0.088	-0.026	-0.059	-0.026	-0.008

The ARIMA Procedure

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	10.14701	0.18992	53.43	<.0001	0
MA1,1	-0.52029	0.03121	-16.67	<.0001	1
MA1,2	-0.27966	0.03417	-8.18	<.0001	2
MA1,3	-0.17471	0.03130	-5.58	<.0001	3

Constant Estimate	10.14701
Variance Estimate	9.267278
Std Error Estimate	3.04422
AIC	5068.71
SBC	5088.341
Number of Residuals	1000

Correlations of Parameter Estimates				
Parameter	MU	MA1,1	MA1,2	MA1,3
MU	1.000	-0.000	-0.001	-0.002
MA1,1	-0.000	1.000	0.444	0.197
MA1,2	-0.001	0.444	1.000	0.446
MA1,3	-0.002	0.197	0.446	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.19	3	0.5346	0.002	-0.004	-0.007	-0.006	-0.045	-0.009
12	8.46	9	0.4889	0.042	0.021	-0.005	0.022	0.059	0.006
18	13.25	15	0.5829	-0.021	-0.031	0.001	-0.017	0.054	-0.009
24	17.77	21	0.6632	-0.036	-0.008	0.022	0.039	0.008	-0.031
30	25.26	27	0.5600	0.056	-0.050	-0.020	-0.030	-0.014	-0.011
36	31.70	33	0.5316	-0.056	0.001	-0.011	0.002	-0.040	-0.036
42	37.31	39	0.5473	-0.001	-0.066	-0.015	0.028	0.006	-0.007
48	50.14	45	0.2770	-0.007	-0.085	-0.025	-0.059	-0.024	-0.014