3.3.2.2 ACF

Note that the pattern, rather than the magnitude of the sequence $\gamma(j)$, is associated with the model form. Normalize the autocovariance sequence $\gamma(j)$ by computing autocorrelations

$$\rho(j) = \gamma(j) \; / \; \gamma(0)$$

Note that

$$\rho(0) = 1$$

for all series and that

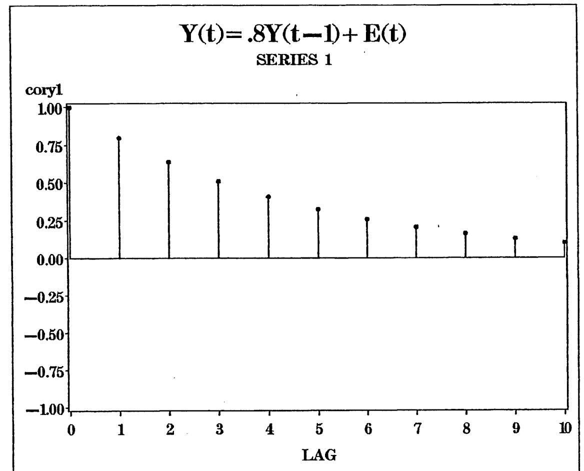
$$\rho(j) = \rho(-j)$$

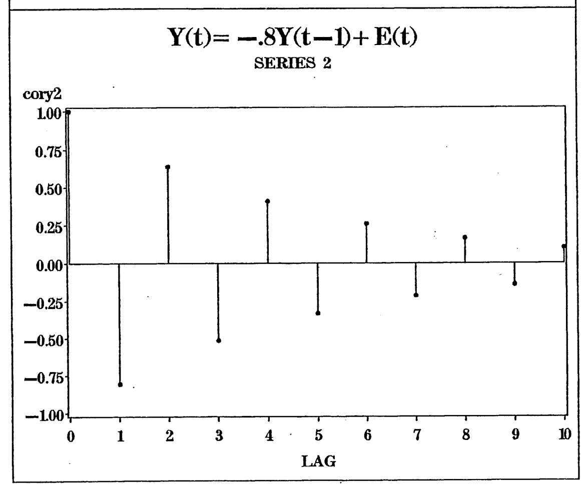
The ACFs for the eight series previously listed are listed below.

Series	Model, ACF
1	$Y_t = .8Y_{t-1} + e_t$, $\rho(j) = .8^{\lvert j \rvert}$
2	$Y_t = -.8Y_{t-1} + e_t$, $\rho(j) = (-.8)^{\lvert j \rvert}$
3	$Y_t = .3Y_{t-1} + .4Y_{t-2} + e_t$, $\rho(1) = .5000$, $\rho(j) = .3\rho(j-1) + .4\rho(j-2)$ for $j > 1$
4	$Y_t = .7Y_{t-1} - .49Y_{t-2} + e_t$, $\rho(1) = .4698$, $\rho(j) = .7\rho(j-1) - .49\rho(j-2)$ for $j > 1$
5	$Y_t = e_t + .8e_{t-1}$, $\rho(1) = .4878$, $\rho(j) = 0$ for $j > 1$
6	$Y_t = e_t - .3e_{t-1} - .4e_{t-2}$, $\rho(1) = -.144$, $\rho(2) = -.32$, $\rho(j) = 0$ for $j > 2$
7	$Y_t = e_t$, $\rho(0) = 1$, $\rho(j) = 0$ for $j > 0$
8	$Y_t - .6Y_{t-1} = e_t + .4e_{t-1}$, $\rho(0) = 1$, $\rho(1) = .7561$, $\rho(j) = .6\rho(j-1)$ for $j > 1$

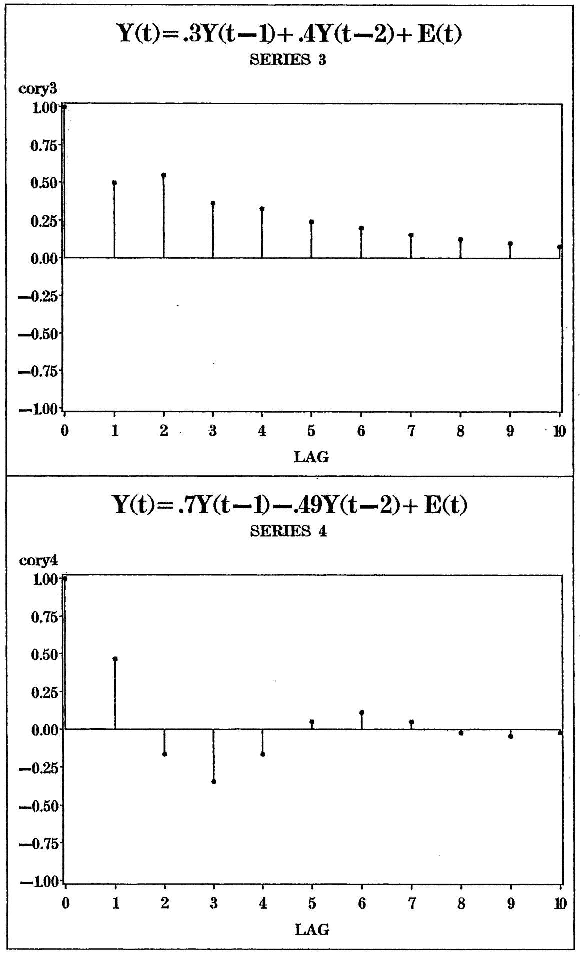
The ACFs are plotted in **Output 3.1**.

Output 3.1 Plotting Actual Autocorrelations for Series 1–8

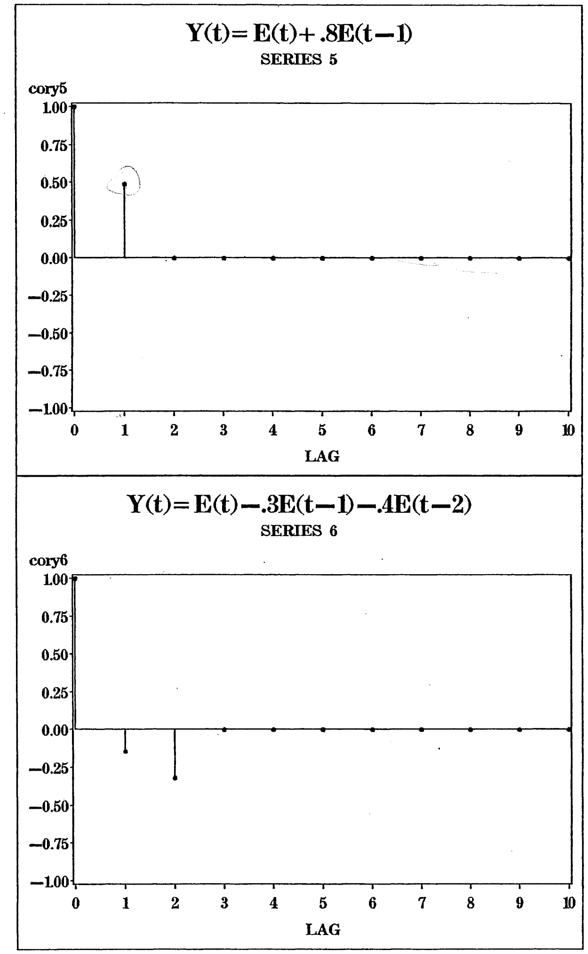




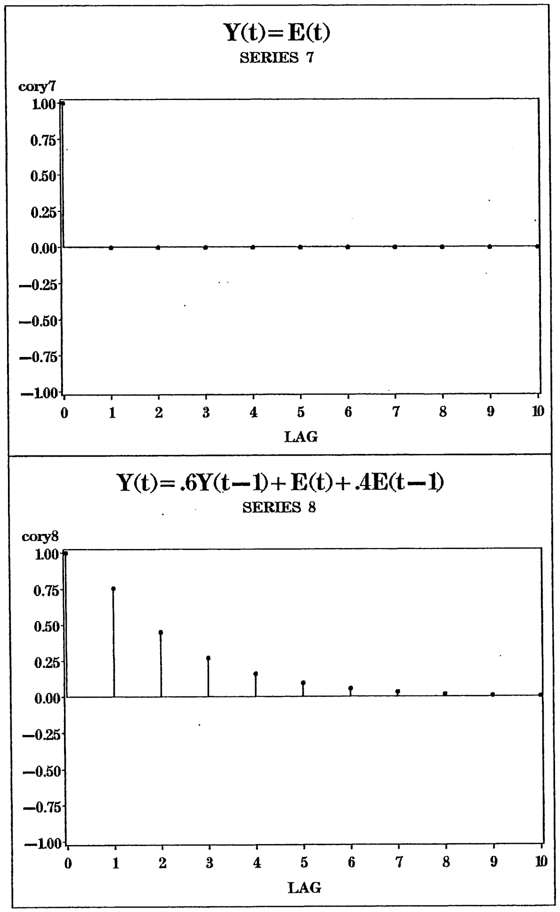
Output 3.1
Plotting Actual Autocorrelations for Series 1–8 (continued)



Output 3.1
Plotting Actual Autocorrelations for Series 1–8 (continued)



Output 3.1
Plotting Actual Autocorrelations for Series 1–8 (continued)



3.3.2.3 PACF

The PACF is motivated by the regression approach to the silver example in Chapter 2, "Simple Models: Autoregression." First, regress Y_t on Y_{t-1} and call the coefficient on Y_{t-1} $\hat{\pi}_1$. Next, regress Y_t on Y_{t-1}, Y_{t-2} and call the coefficient on Y_{t-2} $\hat{\pi}_2$. Continue in this manner, regressing Y_t on Y_{t-1}, $Y_{t-2}, \ldots, Y_{t-j}$ and calling the last coefficient $\hat{\pi}_j$. The $\hat{\pi}_j$ values are the estimated partial autocorrelations.

In an autoregression of order p, the coefficients $\hat{\pi}_j$ estimate 0s for all j>p. The theoretical partial autocorrelations π_j estimated by the $\hat{\pi}_j$ are obtained by solving equations similar to the regression normal equations.

$$\begin{bmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(j-1) \\ \gamma(1) & \gamma(0) & \ldots & \gamma(j-2) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \gamma(j-1) & \gamma(j-2) & \ldots & \gamma(0) \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_j \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \cdot \\ \cdot \\ \gamma(j) \end{bmatrix}$$

For each j, let $\pi_j = b_j$. (A new set of equations is needed for each j.) As with autocorrelations, the π_j sequence is useful for identifying the form of a time series model. The PACF is most useful for identifying AR processes because, for an AR(p), the PACF is 0 beyond lag p. For MA or mixed (ARMA) processes, the theoretical PACF does not become 0 after a fixed number of lags.

You can solve the previous set of equations for the catalog of series. When you observe an estimated PACF $\hat{\pi}_j$, compare its behavior to the behavior shown next to choose a model. The following is a list of actual partial autocorrelations for Series 1–8:

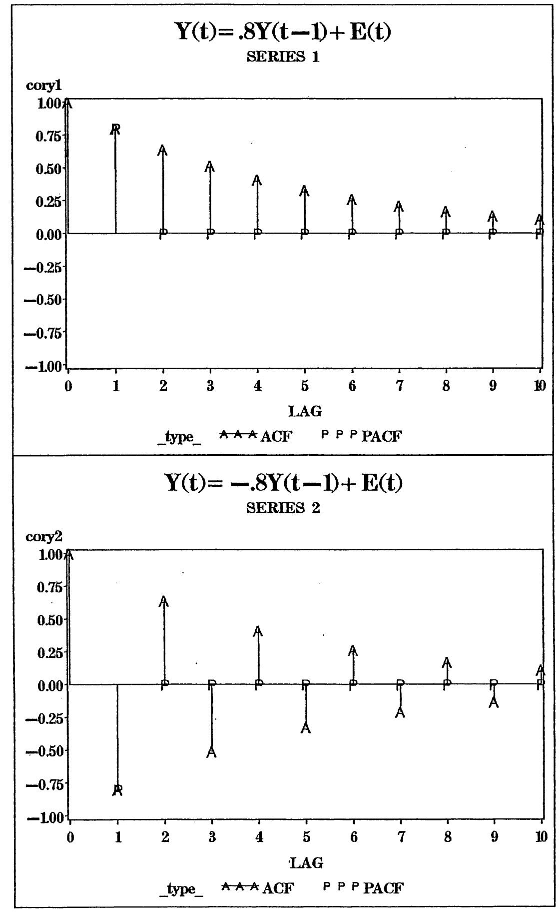
		Lag				
Series	Model	1	2	3	4	5
1	$Y_t = .8Y_{t-1} + e_t$	0.8	0	0	0	0
2	$Y_t = -.8\ Y_t + e_t$	−0.8	0	0	0	0
3	$Y_t = .3Y_{t-1} + .4Y_{t-2} + e_t$	0.5	0.4	0	0	0
4	$Y_t = .7Y_{t-1} - .49Y_{t-2} + e_t$	0.4698	−0.4900	0	0	0
5	$Y_t = e_t + .8\ e_{t-1}$	0.4878	−0.3123	0.2215	−0.1652	0.1267
6	$Y_t = e_t - .3\ e_{t-1} - .4e_{t-2}$	−0.144	−0.3480	−0.1304	−0.1634	−0.0944
7	$Y_t = e_t$	0	0	0	0	0
8	$Y_t = .6Y_{t-1} + e_t + .4e_{t-1}$	0.7561	−0.2756	0.1087	−0.0434	0.0173

Plots of these values against lag number, with A used as a plot symbol for the ACF and P for the PACF, are given in **Output 3.2**. A list of actual autocorrelations for Series 1–8 follows:

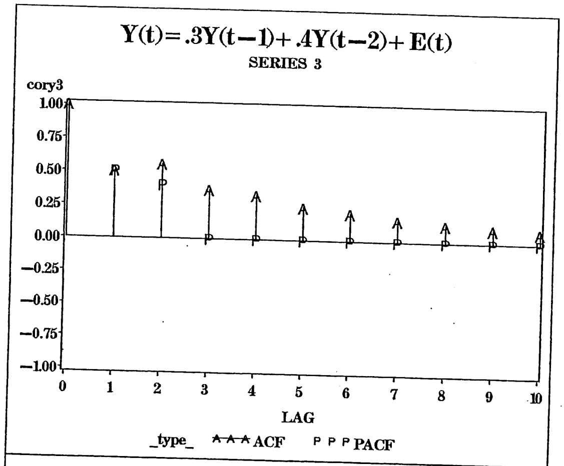
		Lag				
Series	Model	1	2	3	4	5
1	$Y_t = .8Y_{t-1} + e_t$	0.8	0.64	0.512	0.410	0.328
2	$Y_t = -.8\ Y_t + e_t$	−0.8	0.64	−0.512	0.410	−0.328
3	$Y_t = .3Y_{t-1} + .4Y_{t-2} + e_t$	0.500	0.550	0.365	0.330	0.245
4	$Y_t = .\ 7Y_{t-1} - 49Y_{t-2} + e_t$	0.470	−0.161	−0.343	−0.161	0.055
5	$Y_t = e_t + .8\ e_{t-1}$	0.488	0	0	0	0
6	$Y_t = e_t - .3\ e_{t-1} - 4e_{t-2}$	−0.144	−0.32	0	0	0
7	$Y_t = e_t$	0	0	0	0	0
8	$Y_t = .6Y_{t-1} + e_t + .4e_{t-1}$	0.756	0.454	0.272	0.163	0.098

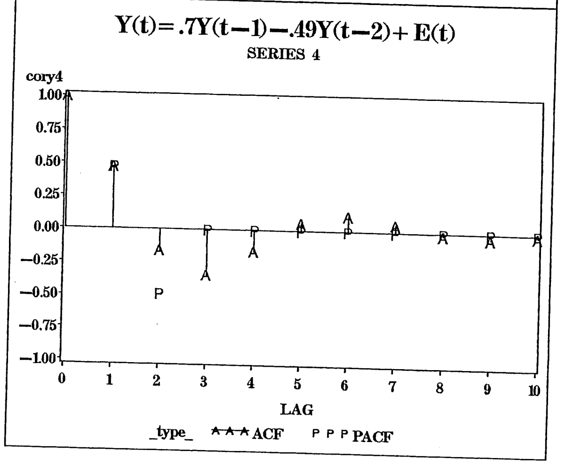
Output 3.2 shows the plots.

Output 3.2 Plotting Actual Autocorrelations and Actual Partial Autocorrelations for Series 1–8

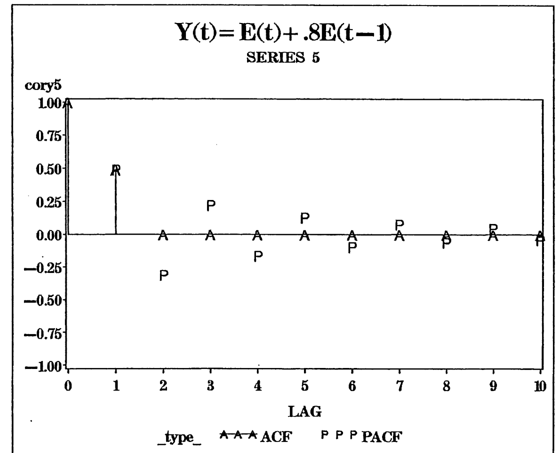


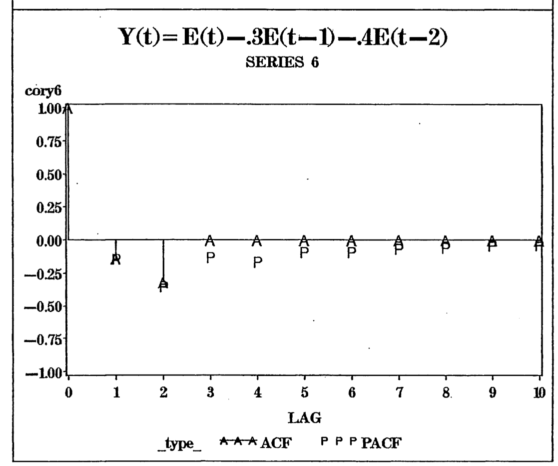
Output 3.2 *Plotting Actual Autocorrelations and Actual Partial Autocorrelations for Series 1–8 (continued)*



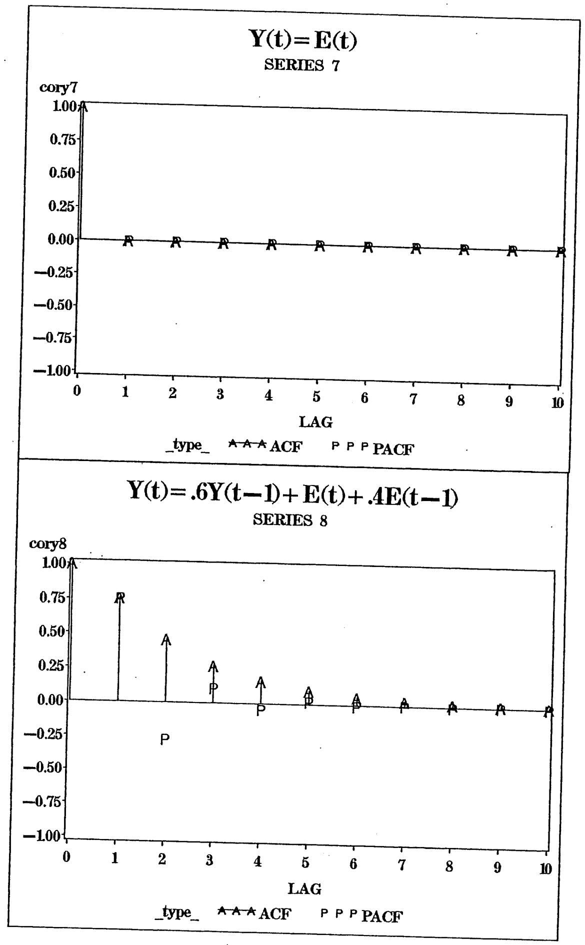


Output 3.2 *Plotting Actual Autocorrelations and Actual Partial Autocorrelations for Series 1–8 (continued)*





Output 3.2 Plotting Actual Autocorrelations and Actual Partial Autocorrelations for Series 1–8 (continued)



3.3.2.4 Estimated ACF

Begin the PROC ARIMA analysis by estimating the three functions defined above. Use these estimates to identify the form of the model. Define the estimated autocovariance C(*j*) as

$$C(j) = \Sigma(Y_t - \overline{Y})(Y_{t+j} - \overline{Y}) / n$$

where the summation is from 1 to $n-j$ and $\overline{Y}$ is the mean of the entire series. Define the estimated autocorrelation by

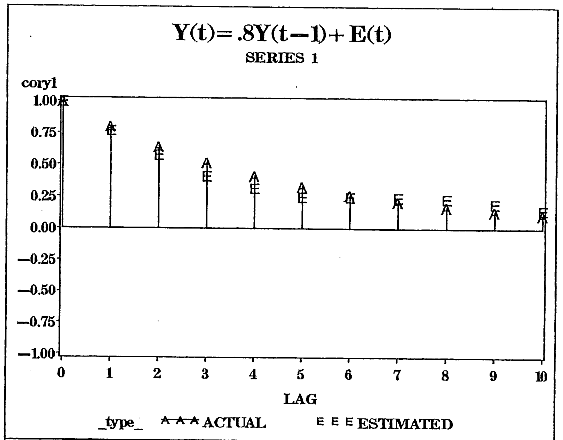
$$r(j) = C(j) / C(0)$$

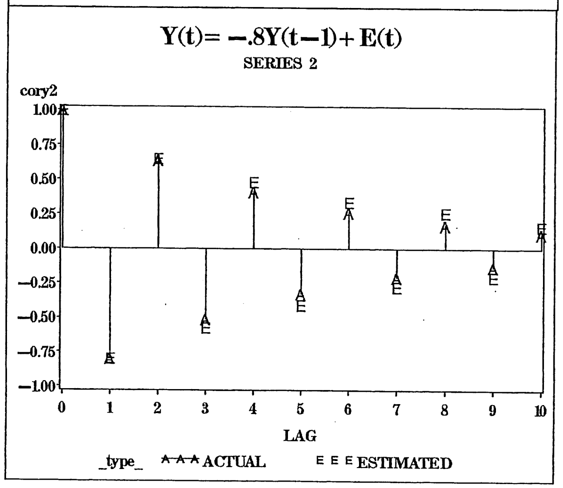
Compute standard errors for autocorrelations in PROC ARIMA as follows:

- For autocorrelation $r(j)$, assign a variance $\left(\Sigma r^2(i)\right)/n$ where the summation runs from $-j+1$ to $j-1$.
- The standard error is the square root of this variance.
- This is the appropriate variance under the hypothesis that $\gamma(i) = 0$ for $i \geq j$ while $\gamma(i) \neq 0$ for $i < j$.

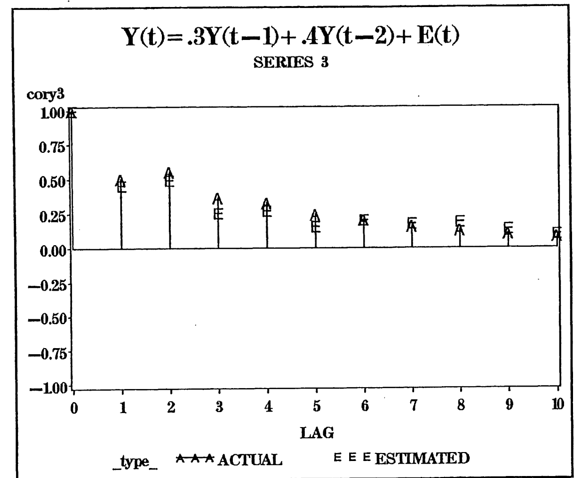
The group of plots in **Output 3.3** illustrates the actual (A) and estimated (E) ACFs for the series. Each data series contains 150 observations. The purpose of the plots is to indicate the amount of sampling error in the estimates.

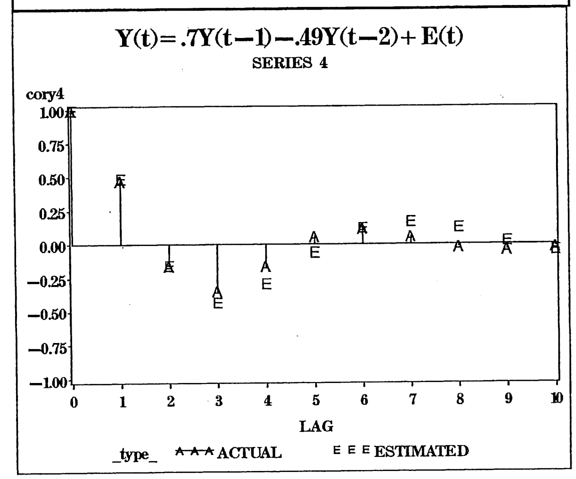
Output 3.3
Plotting Actual and Estimated Autocorrelations for Series 1–8



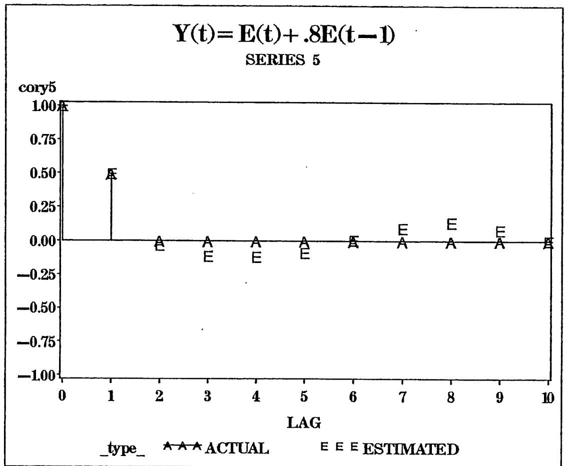


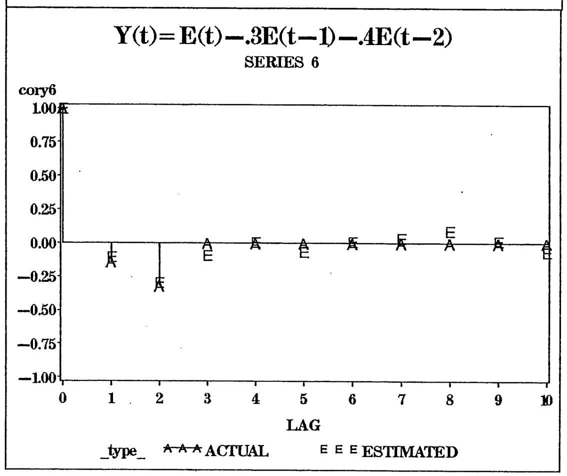
Output 3.3
Plotting Actual and Estimated Autocorrelations for Series 1–8 (continued)



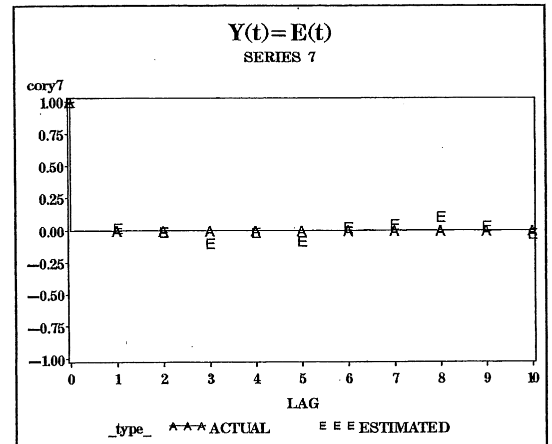


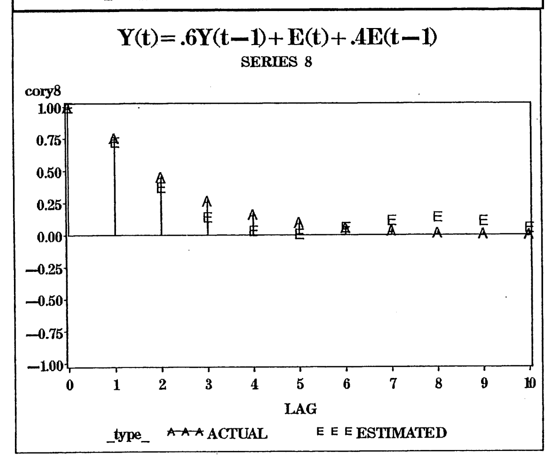
Output 3.3
Plotting Actual and Estimated Autocorrelations for Series 1–8 (continued)





Output 3.3
Plotting Actual and Estimated Autocorrelations for Series 1–8 (continued)



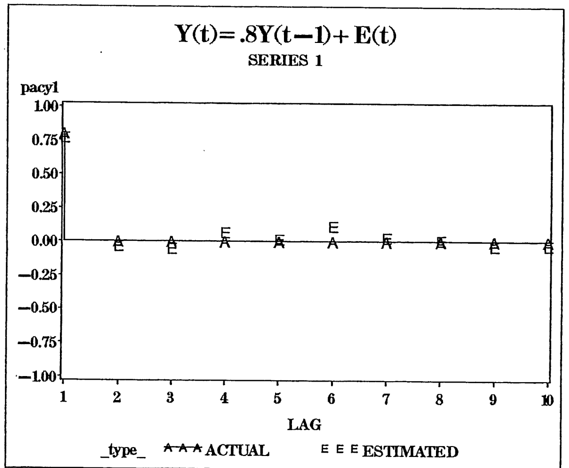


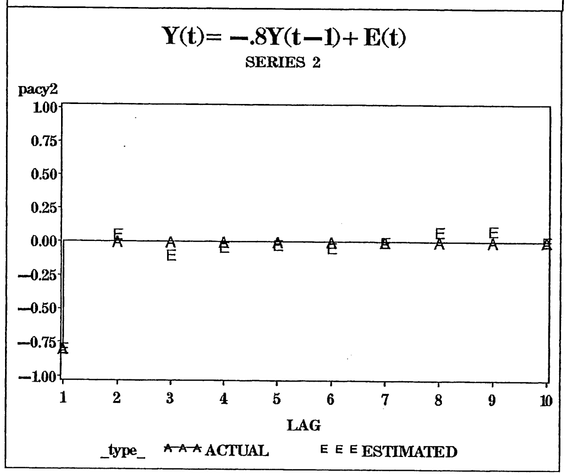
3.3.2.5 Estimated PACF

The partial autocorrelations are defined in **Section 3.3.2.3** as solutions to equations involving the covariances $\gamma(j)$. To estimate these partial autocorrelations, substitute estimated covariances C(*j*) for the actual covariances and solve. For *j* large enough that the actual partial autocorrelation π_j is 0 or nearly 0, an approximate standard error for the estimated partial autocorrelation is $n^{-1/2}$.

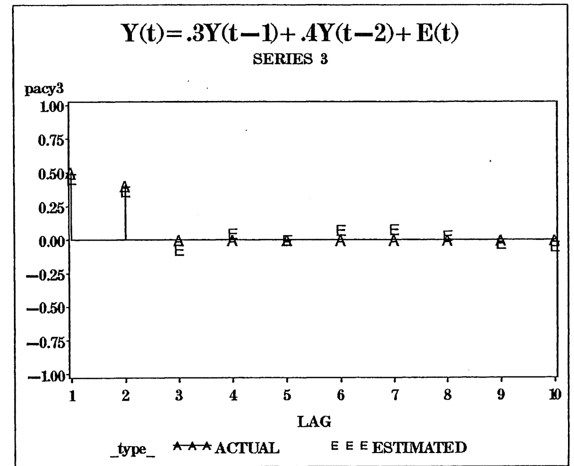
The next group of plots, in **Output 3.4,** illustrate the actual (A) and estimated (E) PACFs for the series.

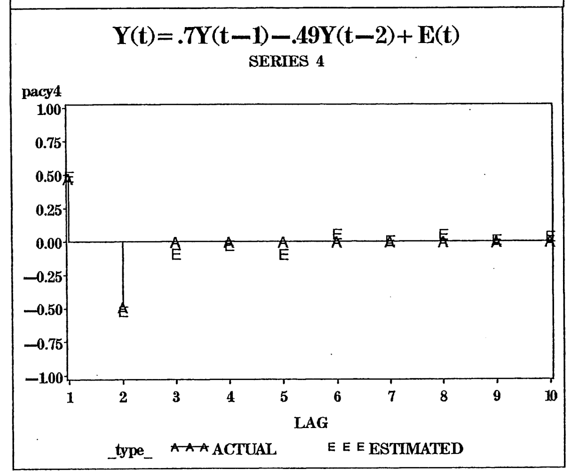
Output 3.4
Plotting Actual and Estimated Partial Autocorrelations for Series 1–8



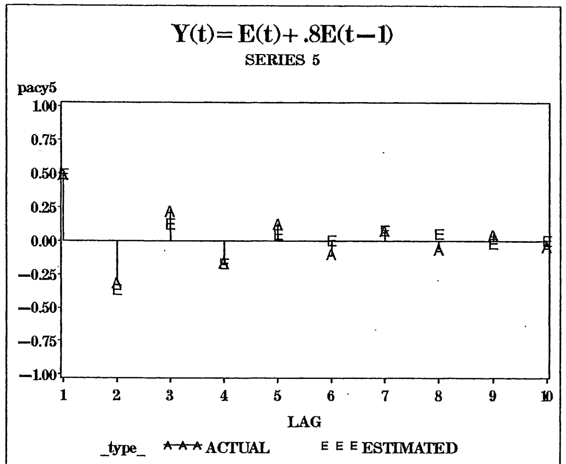


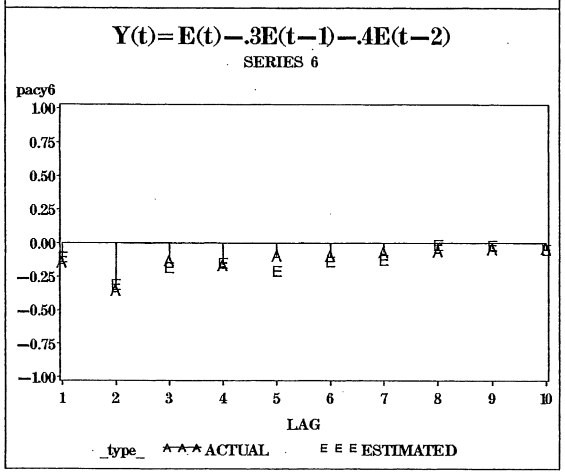
Output 3.4
Plotting Actual and Estimated Partial Autocorrelations for Series 1–8 (continued)



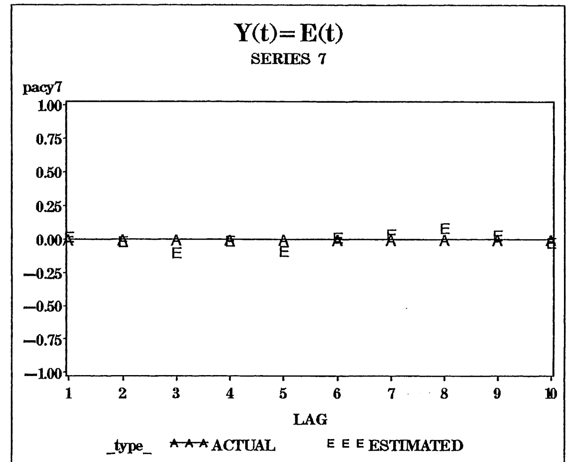


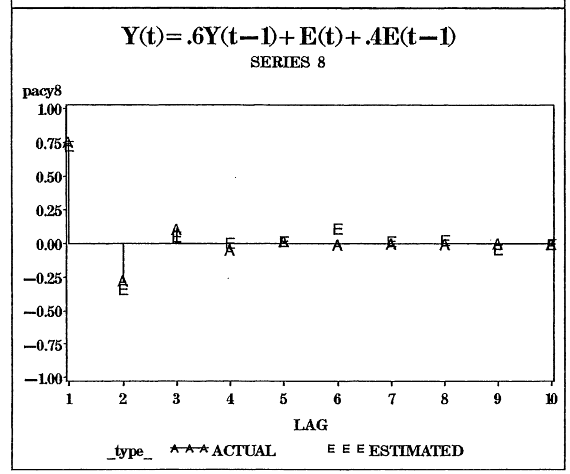
Output 3.4
Plotting Actual and Estimated Partial Autocorrelations for Series 1–8 (continued)





Output 3.4
Plotting Actual and Estimated Partial Autocorrelations for Series 1–8 (continued)





3.3.2.6 IACF

The IACF of an ARMA(p,q) model is defined as the ACF of the ARMA(q,p) model you obtain if you switch sides with the MA and AR operators. Thus, the inverse autocorrelation of

$$(1-.8B)(Y_t-\mu)=e_t$$

is defined as the ACF of

$$Y_t-\mu = e_t-.8e_{t-1}$$

In the catalog of Series 1–8, for example, the IACF of Series 3 is the same as the ACF of Series 6 and vice versa.

3.3.2.7 Estimated IACF

Suppose you know that a series comes from an AR(3) process. Fit an AR(3) model to obtain estimated coefficients—for example,

$$Y_t-\mu = .300(Y_{t-1}-\mu)+.340(Y_{t-2}-\mu)-.120(Y_{t-3}-\mu)+e_t$$

The inverse model is the moving average

$$Y_t-\mu = e_t-.300e_{t-1}-.340e_{t-2}+.120e_{t-3}$$

The inverse autocovariances are estimated by

$$(1+.300^2+.340^2+.120^2)\sigma^2$$

at lag 0,

$$(-.300+(.300)(.340)-(.340)(.120))\sigma^2$$

at lag 1,

$$(-.340-(.300)(.120))\sigma^2$$

at lag 2, and $.120\sigma^2$ at lag 3.

In general, you do not know the order p of the process, nor do you know the form (it may be MA or ARMA). Use the fact (see **Section 3.3.1**) that any invertible ARMA series can be represented as an infinite-order AR and therefore can be approximated by an AR(p) with p large.

Set p to the minimum of the NLAG value and one-half the number of observations after differencing. Then do the following:

- Fit AR(p) to the data.
- Using the estimated coefficients, compute covariances for corresponding MA series as illustrated above for p=3.
- Assign standard errors of $n^{-1/2}$ to the resulting estimates.

3.3.3 Chi-Square Check of Residuals

In the identification stage, PROC ARIMA uses the autocorrelations to form a statistic whose approximate distribution is chi-square under the null hypothesis that the series is white noise. The test is the Ljung modification of the Box-Pierce Q statistic. Both Q statistics are described in Box, Jenkins, and Riensel (1994) and the Ljung modification in Ljung and Box (1978, p. 297). The formula for this statistic is

$$n(n+2)\Sigma_{j=1}^{k} r^2(j)/(n-j)$$

where $r(j)$ is the estimated autocorrelation at lag j and k can be any positive integer. In PROC ARIMA several ks are used.

Later in the modeling stage, PROC ARIMA calculates the same statistic on the model residuals to test the hypothesis that they are white noise. The statistic is compared to critical values from a chi-square distribution. If your model is correct, the residuals should be white noise and the chi-square statistic should be small (the PROB value should be large). A significant chi-square statistic indicates that your model does not fit well.

3.3.4 Summary of Model Identification

At the identification stage, you compute the ACF, PACF, and IACF. Behavior of the estimated functions is the key to model identification. The behavior of functions for different processes is summarized in the following table:

***Table 3.1** Summary of Model Identification*

	MA(q)	AR(p)	ARMA(p, q)	White noise
ACF	D(q)	T	T	0
PACF	T	D(p)	T	0
IACF	T	D(p)	T	0

where

- D(q) means the function drops off to 0 after lag q
- T means the function tails off exponentially
- 0 means the function is 0 at all nonzero lags.

3.4 Examples and Instructions

The following pages contain results for 150 observations generated from each of the eight sample series discussed earlier. Thus, the ACFs correspond to the Es in **Output 3.3**. Even with 150 observations, considerable variation occurs.

To obtain all of the output shown for the first series Y1, use these SAS statements:

```
PROC ARIMA DATA=SERIES;
   IDENTIFY VAR=Y1 NLAG=10;
RUN;
```

The VAR= option is required. The NLAG= option gives the number of autocorrelations to be computed and defaults to 24. When you fit an ARIMA(p,d,q), NLAG+1 must be greater than p+d+q to obtain initial parameter estimates. For the ARMA(p,q) models discussed so far, d is 0.

The following options can also be used:

NOPRINT
: suppresses printout. This is useful because you must use an IDENTIFY statement prior to an ESTIMATE statement. If you have seen the output on a previous run, you may want to suppress it with this option.

CENTER
: subtracts the series mean from each observation prior to the analysis.

DATA=*SASdataset*
: specifies the SAS data set to be analyzed (the default is the most recently created SAS data set).

3.4.1 IDENTIFY Statement for Series 1–8

The following SAS statements, when used on the generated data, produce **Output 3.5**:

```
PROC ARIMA DATA=SERIES;
   IDENTIFY VAR=Y1 NLAG=10;
   IDENTIFY VAR=Y2 NLAG=10;
   more SAS statements
   IDENTIFY VAR=Y8 NLAG=10;
RUN;
```

Try to identify all eight of these series. These are presented in **Section 3.3.2.1**, so you can check your diagnosis against the actual model. For example, look at Y6. First, observe that the calculated Q statistic ❶ is 17.03, which would be compared to a chi-square distribution with six degrees of freedom. The 5% critical value is 12.59, so you have significant evidence against the null hypothesis that the considered model is adequate. Because no model is specified, this Q statistic simply tests the hypothesis that the original data are white noise. The number 0.0092 ❷ is the area under the chi-square distribution to the right of the calculated 17.03. Because 0.0092 is less than .05, without recourse to a chi-square table, you see that 17.03 is to the right of the 5% critical value. Either way, you decide that Y6 is not a white noise series. Contrast this with Y7, where the calculated statistic 2.85 ❸ has an area 0.8269 ❹ to its right; 2.85 is far to the left of the critical value and nowhere near significance. Therefore, you decide that Y7 is a white noise series.

A model is needed for Y6. The PACF and IACF are nonzero through several lags, which means that an AR diagnosis requires perhaps seven lags. A model with few parameters is preferable. The ACF is near 0 after two lags, indicating that you may choose an MA(2). Because an MA model has a persistently nonzero PACF and IACF, the MA(2) diagnosis seems appropriate. At this stage, you have identified the form of the model and can assign the remainder of the analysis to PROC ARIMA. You must identify the model because PROC ARIMA does not do it automatically.

The generated series has 150 observations; note the width of the standard error bands on the autocorrelations. Even with 150 observations, reading fine detail from the ACF is unlikely. Your goal is to use these functions to limit your search to a few plausible models rather than to pinpoint one model at the identification stage.