(portions excerpted from SAS System for Forecasting Time Series)

Likelihood based statistics

- $\ln(\mathcal{L})$ = log of likelihood: big is good.

- AIC = $-2\ln(\mathcal{L}) + 2(p+q)$ small is good
  Akaike's Information Criterion

- SBC = $-2\ln(\mathcal{L}) + (p+q)\ln(n)$
  Schwartz's Bayesian Criterion

- Compare different models for the same data
  Use same estimation method
  Do not compare to each other
  SBC more conservative.

Example: Baseball winning % for NL West pennant winner

```
Data Pennant;
 retain year 1900;   year=year+1; input winpct @@;
 title h=3 "Baseball";   split = (year>1968);  if year > 1920;
cards;
647 741 650 693 686 763 704 643 724 675 647 682 664 614 592 610
636 651 686 604 614 604 621 608 621 578 610 617 645 597 656 584
599 621 649 597 625 586 630 654 649 688 682 682 636 628 610 595
630 591 624 627 682 630 641 604 617 597 564 617 604 624 611 574
599 586 627 599 574 630 556 617 611 630 667 630 605 586 559 571
632 549 562 568 586 593 556 584 568 562 580 605 642
 ;
title2 "West";
proc arima; i var=winpct;
    e p=1 ml itprint grid; e p=2 ml; e q=2 ml;
    e q=1 p=2 ml;  run;
```

Box-Ljung on original data (identify statement)

```
                              Baseball
                                West
                        The ARIMA Procedure
                 Autocorrelation Check for White Noise

 To      Chi-        Pr >
Lag    Square   DF  ChiSq -------------Autocorrelations------------
  6     37.03    6 <.0001  0.435  0.444  0.207  0.142  0.151  0.090
 12     46.99   12 <.0001  0.122  0.141  0.131  0.130  0.151  0.153
 18     51.31   18 <.0001  0.097  0.049 -0.079 -0.087 -0.103  0.094
```

Estimate statement with itprint, grid options

```
                        Preliminary Estimation

                       Initial Autoregressive
                             Estimates

                                   Estimate
                         1          0.43521

                 Constant Term Estimate      344.7338
                 White Noise Variance Est    831.0392

                 Conditional Least Squares Estimation

Iteration       SSE         MU     AR1,1  Constant    Lambda    R Crit

        0     60476   610.3699   0.43521  344.7338   0.00001         1
        1     60470   610.7646   0.44120   341.297      1E-6  0.010208
        2     60470   610.7702   0.44123   341.281      1E-7  0.000114


                     Maximum Likelihood Estimation

  Iter     Loglike         MU     AR1,1  Constant    Lambda    R Crit

     0  -348.94925  610.7702   0.44123   341.281   0.00001         1
     1  -348.94762  610.7347   0.43512  344.9901      1E-6  0.006768
     2  -348.94762  610.7344   0.43524  344.9156      1E-7  0.000135

                ARIMA Estimation Optimization Summary

Estimation Method                                    Maximum Likelihood
Parameters Estimated                                                  2
Termination Criteria                            Maximum Relative Change
                                                           in Estimates
Iteration Stopping Value                                          0.001
Criteria Value                                                 0.000279
Alternate Criteria                                   Relative Change in
                                                     Objective Function
Alternate Criteria Value                                       3.086E-9
Maximum Absolute Value of Gradient                             9.014848
R-Square Change from Last Iteration                            0.000135
Objective Function                              Log Gaussian Likelihood
Objective Function Value                                       -348.948
Marquardt's Lambda Coefficient                                     1E-7
Numerical Derivative Perturbation Delta                           0.001
Iterations                                                            2
```

```
                    Maximum Likelihood Estimation

                              Standard                   Approx
Parameter       Estimate         Error    t Value    Pr > |t|      Lag
MU             610.73440       5.97709     102.18      <.0001        0
AR1,1            0.43524       0.10725       4.06      <.0001        1

                  Constant Estimate      344.9156
                  Variance Estimate      851.6998
                  Std Error Estimate      29.1839
                  AIC                    701.8952
                  SBC                    706.4762
                  Number of Residuals          73

                     Correlations of Parameter
                             Estimates

                   Parameter         MU     AR1,1
                   MU             1.000     0.024
                   AR1,1          0.024     1.000

                 Autocorrelation Check of Residuals

  To     Chi-        Pr >
 Lag   Square   DF  ChiSq -------------Autocorrelations------------

   6     9.32    5 0.0969 -0.129  0.309 -0.012  0.009  0.090 -0.015
  12    11.03   11 0.4405  0.044  0.069  0.045  0.033  0.066  0.074
  18    17.65   17 0.4111  0.043  0.057 -0.106 -0.037 -0.159  0.159
  24    20.25   23 0.6266  0.017  0.007  0.114 -0.071  0.074  0.023

                   Log-likelihood Surface on Grid
                   Near Estimates: AR1,1 (winpct)

              MU
              (winpct)     0.43024   0.43524   0.44024

               610.72940   -348.95   -348.95   -348.95
               610.73440   -348.95   -348.95   -348.95
               610.73940   -348.95   -348.95   -348.95

                     Model for variable winpct
                     Estimated Mean    610.7344

                       Autoregressive Factors
                    Factor 1:  1 - 0.43524 B**(1)
```

B= "backshift" $B(Y_t)=Y_{t-1}$.

$(1-0.4352B)(Y_t-\mu)=e_t \Rightarrow (Y_t-\mu)-0.4352(Y_{t-1}-\mu)=e_t \Rightarrow (Y_t-\mu)=0.4352(Y_{t-1}-\mu)+e_t$

```
                   Maximum Likelihood Estimation

                             Standard                 Approx
Parameter      Estimate         Error    t Value    Pr > |t|     Lag
MU            610.94760       8.00314      76.34      <.0001       0
AR1,1           0.29680       0.11428       2.60      0.0094       1
AR1,2           0.30822       0.11431       2.70      0.0070       2


                  Constant Estimate      241.3102
                  Variance Estimate      779.1167
                  Std Error Estimate     27.91266
                  AIC                    696.5505
                  SBC                    703.4218
                  Number of Residuals          73

                Correlations of Parameter Estimates

              Parameter         MU     AR1,1     AR1,2
              MU             1.000     0.017     0.016
              AR1,1          0.017     1.000    -0.443
              AR1,2          0.016    -0.443     1.000

                 Autocorrelation Check of Residuals

 To     Chi-       Pr >
Lag   Square  DF  ChiSq -------------Autocorrelations------------

  6     1.99   4 0.7379  0.033  0.067 -0.074 -0.101  0.034 -0.054
 12     3.35  10 0.9720  0.007  0.057  0.021  0.006  0.063  0.087
 18    13.71  16 0.6200  0.067  0.052 -0.125 -0.163 -0.171  0.168
 24    16.38  22 0.7965  0.061  0.022  0.100 -0.057  0.088  0.008

                     Model for variable winpct

                     Estimated Mean    610.9476
                       Autoregressive Factors
            Factor 1:  1 - 0.2968 B**(1) - 0.30822 B**(2)
```

$m^2 - 0.2968m - 0.30822 = (m - 0.7231)(m + .4263)$. <u>Fitted</u> model is stationary !!

$$Y_t = \frac{1}{(1-0.7231B)(1+.4263B)} e_t = \left[\frac{A}{(1-0.7231B)} + \frac{1-A}{(1+.4263B)}\right] e_t$$

A = .7231/(.7231+.4263)

```
                      Maximum Likelihood Estimation

                                 Standard                 Approx
Parameter       Estimate            Error    t Value    Pr > |t|     Lag

MU             610.79315          5.63555     108.38      <.0001       0
MA1,1           -0.30846          0.10851      -2.84      0.0045       1
MA1,2           -0.43694          0.10949      -3.99      <.0001       2


                     Constant Estimate      610.7931
                     Variance Estimate      779.7421
                     Std Error Estimate     27.92386
                     AIC                    696.6768
                     SBC                    703.5482
                     Number of Residuals          73

                    Autocorrelation Check of Residuals

 To     Chi-       Pr >
 Lag    Square  DF  ChiSq -------------Autocorrelations------------

   6      2.46   4 0.6517  0.046  0.059  0.139  0.044  0.067 -0.004
  12      4.58  10 0.9172  0.041  0.098  0.046  0.024  0.083  0.062
  18     12.38  16 0.7177  0.038  0.077 -0.055 -0.139 -0.125  0.186
  24     14.23  22 0.8932  0.028 -0.013  0.102 -0.045  0.063  0.007

                        Model for variable winpct

                        Estimated Mean    610.7931
                          Moving Average Factors
                Factor 1:  1 + 0.30846 B**(1) + 0.43694 B**(2)
```

$$m^2+0.30846m + 0.43694 = (m- A)(m - A^*)$$

A, A^* complex pair with $|A| = \sqrt{0.43694} = 0.6610<1$. Fitted model is invertible!!

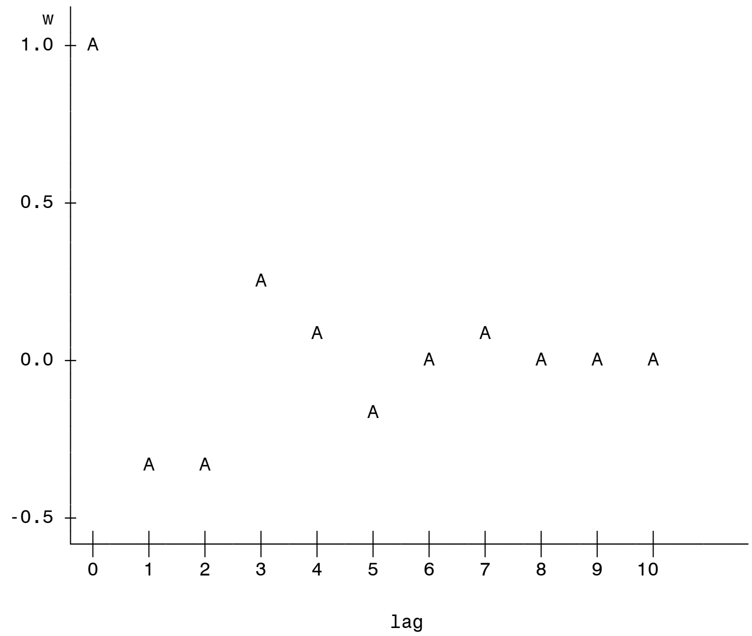
$$\begin{aligned}
e_t &= w_0(Y_t-\mu) + w_1(Y_{t-1}-\mu) + w_2(Y_{t-2}-\mu) + w_3(Y_{t-3}-\mu) + w_4(Y_{t-4}-\mu) + \cdots \\
.30846\, e_{t-1} &= .3046[w_0(Y_{t-1}-\mu) + w_1(Y_{t-2}-\mu) + w_2(Y_{t-3}-\mu) + w_3(Y_{t-4}-\mu) + w_4(Y_{t-5}-\mu) + \cdots \\
.43694\, e_{t-2} &= \qquad .43694[w_0(Y_{t-2}-\mu) + w_1(Y_{t-3}-\mu) + w_2(Y_{t-4}-\mu) + w_3(Y_{t-5}-\mu) + \cdots \\
\text{(sum)} &= 1(Y_t-\mu)
\end{aligned}$$

so we must have (with $w_{-1}=0$) $w_0=1$ and $w_j+.30846w_{j-1}+.43694w_{j-2} = 0$

Generate the ws: observe damped cyclical behavior - typical of complex roots!!

```
data weights;
w1 = 1; w2=0;w=1;lag=0; output;
do lag = 1 to 10;   w = -.30846*w1 - .43694*w2;
output; retain; w2=w1;w1=w; end;
proc plot; plot w*lag/hpos=60 vpos=20;
title "weights"; run;
```

Plot of w*lag. Legend: A = 1 obs, B = 2 obs, etc.



Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	610.94413	7.77268	78.60	<.0001	0
MA1,1	-0.17953	0.37373	-0.48	0.6310	1
AR1,1	0.13527	0.34434	0.39	0.6944	1
AR1,2	0.38109	0.16772	2.27	0.0231	2

Constant Estimate	295.4738
Variance Estimate	785.8795
Std Error Estimate	28.03354
AIC	698.1473
SBC	707.3092
Number of Residuals	73

```
                Autocorrelation Check of Residuals

 To      Chi-        Pr >
Lag    Square   DF  ChiSq -------------Autocorrelations------------
  6      1.64    3 0.6494  0.012  0.052 -0.030 -0.104  0.059 -0.052
 12      2.92    9 0.9673  0.008  0.064  0.022  0.003  0.063  0.077
 18     12.95   15 0.6058  0.065  0.053 -0.112 -0.156 -0.168  0.177
 24     15.88   21 0.7763  0.054  0.012  0.105 -0.069  0.093  0.004

                   Model for variable winpct

                    Estimated Mean    610.9441

                      Autoregressive Factors
          Factor 1:  1 - 0.13527 B**(1) - 0.38109 B**(2)

                      Moving Average Factors
                   Factor 1:  1 + 0.17953 B**(1)
```

	AIC	SBC
AR(1)	701.8952	706.4762
AR(2)	696.5505 ***	703.4218 ***
MA(2)	696.6768	703.5482
ARMA(2,1)	698.1473	707.3092
ARMA(1,1) not shown	699.4642	706.3356

Several new identification tools now available. This section discusses them.

<u>Additional Identification Tools:</u>

In addition to the ACF, IACF, and PACF, three methods called ESACF, SCAN, and MINIC are available for simultaneously identifying both the autoregressive and moving average orders. These consist of tables with rows labeled AR 0, AR 1 etc. and columns MA 0, MA 1 etc. You look at the table entries to find the row and column whose labels give the correct p and q. Tsay and Tiao (1984, 1985) develop the ESACF and SCAN methods and show they even work when the autoregressive operator has roots on the unit circle, in which case p+d rather than p is found. For $(Y_t - Y_{t-2}) - 0.7(Y_{t-1} - Y_{t-3}) = e_t$ ESACF and SCAN should give 3 as the autoregressive order. The key to showing their results is that standard estimation techniques give consistent estimators of the autoregressive operator coefficients even in the presence of unit roots.

These methods can be understood through an ARMA(1,1) example. Suppose you have the ARMA(1,1) process $Z_t - \alpha Z_{t-1} = e_t - \beta e_{t-1}$ where Z_t is the deviation from the mean at time t. The autocorrelations ρ(j) are $\rho(0)=1$, $\rho(1) = [(\alpha-\beta)(1-\alpha\beta)] / [1-\alpha^2 + (\beta-\alpha)^2]$, and $\rho(j) = \alpha\rho(j-1)$ for j>1.

The partial autocorrelations are motivated by the problem of finding the best linear predictor of Z_t based on $Z_{t-1}, \ldots, Z_{t-k}$. That is you want to find coefficients ϕ_{kj} for which $E\{ (Z_t - \phi_{k1}Z_{t-1} - \phi_{k2}Z_{t-2} - \ldots - \phi_{kk}Z_{t-k})^2\}$ is minimized. This is sometimes referred to as "performing a theoretical regression" of Z_t on $Z_{t-1}, Z_{t-2}, \ldots, Z_{t-k}$ or "projecting" Z_t onto the space spanned by $Z_{t-1}, Z_{t-2}, \ldots, Z_{t-k}$. It is accomplished by solving the matrix system of equations

$$\begin{pmatrix} 1 & \rho(1) & \cdots & \rho(k-1) \\ \rho(1) & 1 & \cdots & \rho(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \rho(k-2) & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{pmatrix}$$

Letting $\pi_k = \phi_{kk}$ for k=1,2, $\cdots$ produces the sequence π_k of partial autocorrelations.

At k=1 in the ARMA(1,1) example, you note that $\phi_{11} = \pi_1 = \rho(1) = [(\alpha-\beta)(1-\alpha\beta)] / [1-\alpha^2 + (\beta-\alpha)^2]$ which does not in general equal α. Therefore $Z_t - \phi_{11}Z_{t-1}$ is not $Z_t - \alpha Z_{t-1}$ and thus does not equal $e_t - \beta e_{t-1}$. The autocorrelations of $Z_t - \phi_{11}Z_{t-1}$ would not drop to 0 beyond the moving average order. Increasing k beyond 1 will not solve the problem. Still, it is clear that there is *some* linear combination of Z_t and Z_{t-1}, namely $Z_t - \alpha Z_{t-1}$, whose autocorrelations theoretically identify the order of the moving average part of your model. In general neither the π_k sequence nor any ϕ_{kj} sequence contain the autoregressive coefficients unless the process is a pure autoregression. You are looking for a linear combination $Z_t - C_1 Z_{t-1} - C_2 Z_{t-2} - \ldots - C_p Z_{t-p}$ whose autocorrelation is 0 for j exceeding the moving average order q (1 in our example). The trick is to discover p and the $C_j s$ from the data.

The lagged residual from the theoretical regression of Z_t on Z_{t-1} is $R_{1,t-1} = Z_{t-1} - \phi_{11}Z_{t-2}$ which is a linear combination of Z_{t-1} and Z_{t-2} so regressing Z_t on Z_{t-1} and $R_{1,t-1}$ produces regression coefficients, say C_{21} and C_{22}, which give the same fit, or projection, as regressing Z_t on Z_{t-1} and Z_{t-2}. That is $C_{21}Z_{t-1} + C_{22}R_{1,t-1} = C_{21}Z_{t-1} + C_{22}(Z_{t-1} - \phi_{11}Z_{t-2}) = \phi_{21}Z_{t-1} + \phi_{22}Z_{t-2}$. Thus it must be that $\phi_{21} = C_{21} + C_{22}$ and $\phi_{22} = -\phi_{11}C_{22}$. In matrix form

$$\begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & -\phi_{11} \end{pmatrix} \begin{pmatrix} C_{21} \\ C_{22} \end{pmatrix}$$

Noting that $\rho(2) = \alpha\,\rho(1)$, the ϕ_{2j} coefficients satisfy

$$\begin{pmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} = \rho(1) \begin{pmatrix} 1 \\ \alpha \end{pmatrix}$$

Relating this to the Cs and noting that $\rho(1) = \phi_{11}$, you have

$$\begin{pmatrix} 1 & \phi_{11} \\ \phi_{11} & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -\phi_{11} \end{pmatrix} \begin{pmatrix} C_{21} \\ C_{22} \end{pmatrix} = \phi_{11} \begin{pmatrix} 1 \\ \alpha \end{pmatrix}$$

or

$$\begin{pmatrix} C_{21} \\ C_{22} \end{pmatrix} = \frac{\phi_{11}}{\phi_{11}(\phi_{11}^2\text{-}1)} \begin{pmatrix} 0 & \phi_{11}^2 - 1 \\ -\phi_{11} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ \alpha \end{pmatrix} = \begin{pmatrix} \alpha \\ \frac{\alpha - \phi_{11}}{\phi_{11}^2\text{-}1} \end{pmatrix}$$

You now "filter" Z using only $C_{21} = \alpha$, that is, you compute Z_t - C_{21} Z_{t-1} which is just Z_t - α Z_{t-1} and this in turn is a moving average of order 1. Its lag 1 autocorrelation (it is nonzero) will appear in the AR 1 row and MA 0 column of the ESACF table. Let the residual from this regression be denoted $R_{2,t}$. The next step is to regress Z_t on Z_{t-1}, $R_{1,t-2}$, and $R_{2,t-1}$. In this regression, the theoretical coefficient of Z_{t-1} will again be α but its estimate may differ somewhat from the one obtained previously. Notice the use of the lagged value of $R_{2,t}$ and the second lag of the first round residual $R_{1,t-2} = Z_{t-2} - \phi_{11} Z_{t-3}$. The lag 2 autocorrelation of Z_t - α Z_{t-1}, which is 0, will be written in the MA 1 column of the AR 1 row. For the ESACF of a general ARMA(p,q) in the AR p row, once your regression has at least q lagged residuals the first p theoretical C_{kj} will be the p autoregressive coefficients and the filtered series will be a MA(q) so its autocorrelations will be 0 beyond lag q.

The entries in the AR k row of the ESACF table are computed as follows:

(1) Regress Z_t on Z_{t-1}, Z_{t-2}, ..., Z_{t-k} with residual $R_{1,t}$
Coefficients: C_{11},C_{12},...,C_{1k}
(2) Regress Z_t on Z_{t-1}, Z_{t-2}, ..., Z_{t-k}, $R_{1,t-1}$ with residual $R_{2,t}$
Second round coefficients: C_{21},...,C_{2k}, (and $C_{2,k+1}$)
Record in MA 0 column, the lag 1 autocorrelation of
$Z_t - C_{21}Z_{t-1} - C_{22}Z_{t-2} - \ldots - C_{2k}Z_{t-k}$
(3) Regress Z_t on Z_{t-1}, Z_{t-2}, ..., Z_{t-k}, $R_{1,t-2}$, $R_{2,t-1}$ with residual $R_{3,t}$
Third round coefficients: C_{31},...,C_{3k}, (and $C_{3,k+1}$,$C_{3,k+2}$)
Record in MA 1 column, the lag 2 autocorrelation of
$Z_t - C_{31}Z_{t-1} - C_{32}Z_{t-2} - \ldots - C_{3k}Z_{t-k}$
etc.

Notice that at each step, you lag all residuals that were previously included as regressors and add the lag of the most recent residual to your regression. The estimated C coefficients and resulting filtered series differ at each step. Looking down the ESACF table of an AR(p,q), theoretically row p should be the first row in which a string of 0's appears and it should start at the MA q column. Finding that row and the first 0 entry in it puts you in row p column q of the ESACF. The model is now identified.

Here is a theoretical ESACF table for an ARMA(1,1) with "X" for nonzero numbers

	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	X	X	X	X	X	X
AR 1	X	0^*	0	0	0	0
AR 2	X	X	0	0	0	0
AR 3	X	X	X	0	0	0
AR 4	X	X	X	X	0	0

The string of 0s slides to the right as the AR row number moves beyond p so there appears a triangular array of 0s whose "point" 0^* is at the correct (p,q) combination.

In practice, the theoretical regressions are replaced by least squares regressions so the ESACF table will only have numbers near 0 where the theoretical ESACF table has 0s. A recursive algorithm is used to quickly compute the needed coefficients without having to compute so many actual regressions. PROC ARIMA will also use asymptotically valid standard errors based on Bartlett's formula to deliver a table of approximate p-values for the ESACF entries and will suggest values of p and q as a tentative identification. See Tsay and Tiao (1984) for further details.

Tsay and Tiao (1985) suggest a second table called SCAN. It is computed using canonical correlations. For the ARMA(1,1) model, recall that the autocovariances are $\gamma(0)$, $\gamma(1)$,$\gamma(2)= \alpha\gamma(1)$, $\gamma(3)= \alpha^2\gamma(1)$, $\gamma(4)= \alpha^3\gamma(1)$, etc. so the covariance matrix of Y_t, Y_{t-1}, ... , Y_{t-5} is

$$\Gamma = \begin{pmatrix} \gamma(0) & \gamma(1) & \alpha\gamma(1) & \alpha^2\gamma(1) & \alpha^3\gamma(1) & \alpha^4\gamma(1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \alpha\gamma(1) & \alpha^2\gamma(1) & \alpha^3\gamma(1) \\ [\ \alpha\gamma(1)\ ] & [\ \gamma(1)\ ] & \gamma(0) & \gamma(1) & \alpha\gamma(1) & \alpha^2\gamma(1) \\ [\ \underline{\alpha^2\gamma(1)}\ ] & [\ \underline{\alpha\gamma(1)}\ ] & \underline{\gamma(1)} & \gamma(0) & \gamma(1) & \alpha\,\gamma(1) \\ \underline{\alpha^3\gamma(1)} & \underline{\alpha^2\gamma(1)} & \underline{\alpha\gamma(1)} & \gamma(1) & \gamma(0) & \gamma(1) \\ \underline{\alpha^4\gamma(1)} & \underline{\alpha^3\gamma(1)} & \underline{\alpha^2\gamma(1)} & \alpha\gamma(1) & \gamma(1) & \gamma(0) \end{pmatrix}$$

The entries in square brackets form the 2x2 submatrix of covariances between the vectors (Y_t, Y_{t-1}) and (Y_{t-2}, Y_{t-3}). That submatrix **A**, the variance matrix $\mathbf{C}_{11}$ of (Y_t, Y_{t-1}), and the variance matrix $\mathbf{C}_{22}$ of (Y_{t-2}, Y_{t-3}) are

$$\mathbf{A} = \gamma(1)\begin{pmatrix} \alpha & 1 \\ \alpha^2 & \alpha \end{pmatrix} \qquad \mathbf{C}_{11} = \mathbf{C}_{22} = \begin{pmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{pmatrix}$$

The best linear predictor of (Y_t, $Y_{t-1})'$ based on (Y_{t-2}, $Y_{t-3})'$ is $\mathbf{A}'\mathbf{C}_{22}^{-1}$(Y_{t-2}, $Y_{t-3})'$ with prediction error variance matrix $\mathbf{C}_{11} - \mathbf{A}'\mathbf{C}_{22}^{-1}\mathbf{A}$. Because matrix $\mathbf{C}_{11}$ represents the variance of (Y_t, Y_{t-1}), the matrix $\mathbf{C}_{11}^{-1}\ \mathbf{A}'\mathbf{C}_{22}^{-1}\mathbf{A}$ is analogous to a regression R^2

statistic. Its eigenvalues are called squared canonical correlations between (Y_t, $Y_{t-1})'$ and (Y_{t-2}, $Y_{t-3})'$.

Recall that, for a square matrix **M**, if a column vector **H** exists such that **MH** = b**H** then **H** is called an *eigenvector* and the scalar b is the corresponding *eigenvalue* of matrix **M**. Using **H** = $(1, -\alpha)'$ you see that **AH** = $(0,0)'$ so that $\mathbf{C}_{11}^{-1}\mathbf{A}'\mathbf{C}_{22}^{-1}\mathbf{AH} = 0\mathbf{H}$, that is, $\mathbf{C}_{11}^{-1}\mathbf{A}'\mathbf{C}_{22}^{-1}\mathbf{A}$ has an eigenvalue 0. The number of 0 eigenvalues of **A** is the same as the number of 0 eigenvalues of $\mathbf{C}_{11}^{-1}\mathbf{A}'\mathbf{C}_{22}^{-1}\mathbf{A}$. This is true for general time series covariance matrices.

The matrix **A** has first column that is α times the second which implies these equivalent statements:

(1) The 2x2 matrix **A** is not of full rank (its rank is 1)
(2) The 2x2 matrix **A** has at least one eigenvalue 0
(3) The 2x2 matrix $\mathbf{C}_{11}^{-1}\mathbf{A}'\mathbf{C}_{22}^{-1}\mathbf{A}$ has at least one eigenvalue 0
(4) The vectors (Y_t, Y_{t-1}) and (Y_{t-2}, Y_{t-3}) have at least one squared canonical correlation that is 0.

The fourth of these statements is easily seen. The linear combinations $Y_t - \alpha Y_{t-1}$ and its second lag $Y_{t-2} - \alpha Y_{t-3}$ have correlation 0 because each is a MA(1). The smallest canonical correlation is obtained by taking linear combinations of (Y_t, Y_{t-1}) and (Y_{t-2}, Y_{t-3}) and finding the pair with correlation closest to 0. Since there exist linear combinations in the two sets that are *uncorrelated*, the smallest canonical correlation must be 0. Again you have a method of finding a linear combination whose autocorrelation sequence is 0 beyond the moving average lag q.

In general, construct an arbitrarily large covariance matrix of Y_t, Y_{t-1},Y_{t-2},... and let $\mathbf{A}_{j,m}$ be the mxm matrix whose upper left element is in row $j+1$, column 1 of the original matrix. In this notation, the **A** with square bracketed elements is denoted $\mathbf{A}_{2,2}$. and that with the underlined elements is $\mathbf{A}_{3,3}$. Again there is a full rank 3x2 matrix **H** for which $\mathbf{A}_{3,3}\mathbf{H}$ has all 0 elements, namely

$$\mathbf{A}_{3,3}\mathbf{H} = \begin{pmatrix} \alpha^2\gamma(1) & \alpha\gamma(1) & \gamma(1) \\ \alpha^3\gamma(1) & \alpha^2\gamma(1) & \alpha\gamma(1) \\ \alpha^4\gamma(1) & \alpha^3\gamma(1) & \alpha^2\gamma(1) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\alpha & 1 \\ 0 & -\alpha \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

showing that matrix $\mathbf{A}_{3,3}$ has (at least) 2 eigenvalues that are 0 with the columns of **H** being the corresponding eigenvectors. Similarly, using $\mathbf{A}_{3,2}$ and **H** = $(1, -\alpha)$

$$\mathbf{A}_{3,2}\mathbf{H} = \begin{pmatrix} \alpha^2\gamma(1) & \alpha\gamma(1) \\ \alpha^3\gamma(1) & \alpha^2\gamma(1) \end{pmatrix} \begin{pmatrix} 1 \\ -\alpha \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

so $\mathbf{A}_{3,2}$ has (at least) one 0 eigenvalue as does $\mathbf{A}_{j,2}$ for all j>1. In fact all $\mathbf{A}_{j,m}$ with j>1 and m>1 have at least one 0 eigenvalue for this example. For general ARIMA(p,q) models, all $\mathbf{A}_{j,m}$ with j>q and m>p have at least one 0 eigenvalue. This provides the key to the SCAN table. If you make a table whose m^{th} row, j^{th} column entry is the smallest canonical correlation derived from $\mathbf{A}_{j,m}$ you have this table for the current example:

	m=1	m=2	m=3	m=4
j=1	X	X	X	X
j=2	X	0	0	0
j=3	X	0	0	0
j=4	X	0	0	0

	p=0	p=1	p=2	p=3
q=0	X	X	X	X
q=1	X	0	0	0
q=2	X	0	0	0
q=3	X	0	0	0

where the Xs represent nonzero numbers. Relabeling the rows and columns with q=j-1 and p=m-1 gives the SCAN (smallest canonical correlation) table. It has a rectangular array of 0s whose upper left corner is at the p and q corresponding to the correct model, ARMA(1,1) for the current example. The first column of the SCAN table consists of the autocorrelations and the first row the partial autocorrelations.

In PROC ARIMA, entries of the 6x6 variance-covariance matrix $\mathbf{\Gamma}$ above would be replaced by estimated autocovariances. To see why the 0s appear for an ARMA(p,q) whose autoregressive coefficients are α_i, you notice from the Yule Walker equations that $\gamma(j) - \alpha_1\gamma(j-1) - \alpha_2\gamma(j-2) - \ldots - \alpha_p\gamma(j-p)$ is zero for j>q. Therefore, in the variance covariance matrix for such a process, any mxm submatrix with m>p whose upper left element is at row j column 1 of the original matrix will have at least one 0 eigenvalue with eigenvector $(1, -\alpha_1, -\alpha_2, \ldots, -\alpha_p, 0, 0, \ldots, 0)'$ if j>q . Hence 0 will appear in the theoretical table whenever m>p and j>q. Approximate standard errors are obtained by applying Bartlett's formula to the series filtered by the autoregressive coefficients, which in turn can be extracted from the **H** matrix (eigenvectors). An asymptotically valid test, again making use of Bartlett's formula, is available and PROC ARIMA displays a table of the resulting p-values.

The MINIC method simply attempts to fit models over a grid of p and q choices, and records the SBC information criterion for each fit in a table. The Schwartz Bayesian Information Criterion is SBC = n ln(s^2) + (p+q) ln(n) where p and q are the autoregressive and moving average orders of the candidate model and s^2 an estimate of the innovations variance. Although some references refer to Schwartz' criterion, perhaps normalized by n, as BIC here the symbol SBC is used so that Schwartz' criterion will not be confused the BIC criterion of Sawa (1978). Sawa's BIC, used as a model selection tool in PROC REG, is n ln(s^2) + 2[$(k+2)\frac{n}{n-k} - \left(\frac{n}{n-k}\right)^2$] for a full regression model with n observations and k parameters. The MINIC technique chooses p and q giving the smallest SBC. It is possible, of course, that the fitting will fail due to singularities in which case the SBC is set to missing.

The fitting of models in computing MINIC follows a clever algorithm suggested by Hannan and Rissanen (1982) using ideas dating back to Durbin (1960). First, using the Yule-Walker equations, a long autoregressive model is fit to the data. For the ARMA(1,1) example of this section it is seen that

$$Y_t = (\alpha - \beta)[Y_{t-1} + \beta Y_{t-2} + \beta^2 Y_{t-3} + ...] + e_t$$

and as long as $|\beta| < 1$ the coefficients on lagged Y will die off quite quickly indicating that a truncated version of this infinite autoregression will approximate the e_t process well. To the extent that this is true, the Yule-Walker equations for a length k (k large) autoregression can be solved to give estimates, say $\hat{b}_j$, of the coefficients of the Y_{t-j} terms and a residual series $\hat{e}_t = Y_t - \hat{b}_1 Y_{t-1} - \hat{b}_2 Y_{t-2} - \ldots - \hat{b}_k Y_{t-k}$ that is close to the actual e_t series. Next, for a candidate model of order p,q, regress Y_t on Y_{t-1}, ... ,Y_{t-p} , $\hat{e}_{t-1}$, $\hat{e}_{t-2}$,..., $\hat{e}_{t-q}$. Letting $\hat{\sigma}^2_{pq}$ be 1/n times the error sum of squares for this regression, pick p and q to minimize the SBC criterion SBC = n ln($\hat{\sigma}^2_{pq}$) + (p+q) ln(n). The length of the autoregressive model for the $\hat{e}_t$ series can be selected by minimizing the AIC criterion.

To illustrate, 1000 observations on an ARMA(1,1) with α= .8 and β = .4 are generated and analyzed

```
data a;
   a = .8; b=.4; y1=0; e1=0;
do t=1 to 1000;
  e = normal(1726555);
  Y = a*y1 + e -b*e1;
  output; e1=e; y1=y;
end;
i var=y nlag=1 minic p=(0:5) q=(0:5) ;
i var=y nlag=1 esacf p=(0:5) q=(0:5) ;
i var=y nlag=1 scan  p=(0:5) q=(0:5) ;
```

giving the results in the output below.

ESACF, SCAN and MINIC Displays

Minimum Information Criterion

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.28456	0.177502	0.117561	0.059353	0.028157	0.003877
AR 1	-0.0088	-0.04753	-0.04502	-0.0403	-0.03565	-0.03028
AR 2	-0.03958	-0.04404	-0.04121	-0.0352	-0.03027	-0.02428
AR 3	-0.04837	-0.04168	-0.03537	-0.02854	-0.02366	-0.01792
AR 4	-0.04386	-0.03696	-0.03047	-0.02372	-0.01711	-0.01153
AR 5	-0.03833	-0.03145	-0.02461	-0.0177	-0.01176	-0.00497

Error series model: AR(9)
Minimum Table Value: BIC(3,0) = -0.04837

Extended Sample Autocorrelation Function

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.5055	0.3944	0.3407	0.2575	0.2184	0.1567
AR 1	-0.3326	-0.0514	0.0564	-0.0360	0.0417	-0.0242
AR 2	-0.4574	-0.2993	0.0197	0.0184	0.0186	-0.0217
AR 3	-0.1207	-0.2357	0.1902	0.0020	0.0116	0.0006
AR 4	-0.4074	-0.1753	0.1942	-0.0132	0.0119	0.0015
AR 5	0.4836	0.1777	-0.0733	0.0336	0.0388	-0.0051

ESACF Probability Values

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.0001	0.0001	0.0001	0.0001	0.0001	0.0010
AR 1	0.0001	0.1489	0.1045	0.3129	0.2263	0.4951
AR 2	0.0001	0.0001	0.5640	0.6013	0.5793	0.6003
AR 3	0.0001	0.0001	0.0001	0.9598	0.7634	0.9874
AR 4	0.0001	0.0001	0.0001	0.7445	0.7580	0.9692
AR 5	0.0001	0.0001	0.0831	0.3789	0.2880	0.8851

ARMA(p+d,q) Tentative Order Selection Tests
(5% Significance Level)

ESACF	p+d	q
	1	1
	4	3
	5	2

Squared Canonical Correlation Estimates

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.2567	0.1563	0.1170	0.0670	0.0483	0.0249
AR 1	0.0347	0.0018	0.0021	0.0008	0.0011	0.0003
AR 2	0.0140	0.0023	0.0002	0.0002	0.0002	0.0010
AR 3	0.0002	0.0007	0.0002	0.0001	0.0002	0.0001
AR 4	0.0008	0.0010	0.0002	0.0002	0.0002	0.0002
AR 5	0.0005	0.0001	0.0002	0.0001	0.0002	0.0004

SCAN Chi-Square[1] Probability Values

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.0001	0.0001	0.0001	0.0001	0.0001	0.0010
AR 1	0.0001	0.2263	0.1945	0.4097	0.3513	0.5935
AR 2	0.0002	0.1849	0.7141	0.6767	0.7220	0.3455
AR 3	0.6467	0.4280	0.6670	0.9731	0.6766	0.9877
AR 4	0.3741	0.3922	0.6795	0.6631	0.7331	0.7080
AR 5	0.4933	0.8558	0.7413	0.9111	0.6878	0.6004

ARMA(p+d,q) Tentative Order Selection Tests
(5% Significance Level)

SCAN	p+d	q
	1	1
	3	0

The tentative order selections in ESACF and SCAN simply look at all triangles (rectangles) for which every element is insignificant at the specified level (0.05 by default). These are listed in descending order of size, size being the number of elements in the triangle or rectangle. In our example ESACF and SCAN list the correct (1,1) order at the top of the list. The MINIC criterion uses k=9, a preliminary AR(9) model, to create the estimated white noise series, then selects (p,q)= (3,0) as the order, this also being one choice given by the SCAN option. The second smallest SBC, -.04753, occurs at the correct (p,q) = (1,1).

A Monte Carlo Investigation:

As a check on the relative merits of these methods, 50 ARMA(1,1) series each of length 500 are generated for each of the 12 (α,β) pairs obtained by choosing α and β from {-.9, -.3, .3, .9} such that $\alpha \neq \beta$. This gives 600 series. For each, the ESACF, SCAN and MINIC methods are used, the output window with the results is saved, then read in a data step to extract the estimated p and q for each method. The whole experiment is repeated with series of length 50. A final set of 600 runs for $Y_t = .5Y_{t-4} + e_t + .3\, e_{t-1}$ using n=50 gives the last three rows. Asterisks indicate the correct model.

	|<- ARMA(1,1) n=500 ->|			|<- ARMA(1,1) n=50 ->|			|<- ARMA(4,1) n=50 ->|		
pq	BIC	ESACF	SCAN	BIC	ESACF	SCAN	BIC	ESACF	SCAN
00	2	1	1	25	40	25	69	64	35
01	0	0	0	48	146	126	28	46	33
02	0	0	0	17	21	8	5	9	11
03	0	0	0	7	4	16	4	6	2
04	0	0	0	7	3	2	41	20	35
05	0	0	0	6	1	0	14	0	2
10	1	0	0	112	101	145	28	15	38
11	* 252 ***	441 ***	461 ***	53 **	165 **	203 *	5	47	78
12	13	23	8	16	7	9	1	10	30
13	13	18	8	12	0	2	0	1	3
14	17	5	3	2	0	1	3	0	0
15	53	6	0	5	0	0	4	0	0
20	95	6	12	91	41	18	26	16	19
21	9	6	25	9	22	14	2	42	25
22	24	46	32	4	8	7	3	62	121
23	1	0	1	1	2	1	1	2	8
24	4	0	2	3	0	1	2	1	0
25	10	0	1	6	0	0	0	0	0
30	35	2	9	50	6	8	30	9	21
31	5	3	11	1	10	3	3	23	27
32	3	6	1	3	4	0	3	21	7
33	3	15	13	0	2	0	1	16	2
34	5	2	0	1	0	0	0	0	0
35	4	0	0	2	0	0	0	0	0
40	5	0	0	61	6	6	170	66	98
41	3	0	5	3	4	0	* 10 ***	52 ***	0 *
42	2	4	2	1	2	0	4	24	0
43	5	3	0	0	0	0	0	22	0
44	1	4	1	1	0	0	0	0	0
45	6	0	0	5	0	0	1	0	0
50	5	0	0	32	3	2	116	6	5
51	0	1	2	10	1	0	18	13	0
52	3	0	1	2	0	0	2	6	0
53	9	2	0	2	0	0	5	0	0
54	6	1	0	2	1	0	1	0	0
55	6	0	0	0	0	0	0	0	0
totals	600	595	599	600	600	597	600	599	600

It is reassuring that the methods almost never underestimate p or q when n is 500. For the ARMA(1,1) with parameters in this range, it appears that SCAN does slightly better than ESACF with both being superior to MINIC. The SCAN and ESACF columns do not always add to 600 because for some cases, no rectangle or triangle can be found with all elements insignificant. Because SCAN compares the *smallest* normalized squared canonical correlation to a distribution (χ^2_1) that is appropriate for a randomly selected one, it is very conservative. By analogy, even if 5% of men exceed 6 feet in height, finding a random sample of 10 men whose *shortest* member exceeds 6 feet in height would be extremely rare. Thus the appearance of a significant bottom right corner element in the SCAN table, which would imply no rectangle of insignificant values, happens rarely - not the 30 times you would expect from 600(.05)=30.

The conservatism of the test also implies that for p and q moderately large there is a fairly good chance that a rectangle (triangle) of "insignificant" terms will appear by change having p or q too small. Indeed for 600 replicates of the model $Y_t = 0.5Y_{t-4} + e_t + 0.3\, e_{t-1}$ using n=50, we see that (p,q)=(4,1) is rarely chosen by any technique with SCAN giving no correct choices. There does not seem to be a universally preferable choice among the three.

As a real data example, we use monthly interbank loans in billions of dollars. The data were downloaded from the Federal Reserve website. The data seem stationary when differences of logarithms are used (see the discussion of Dow Jones Industrial Average). When data are differenced, the models are referred to as ARIMA(p,d,q) where the I stands for "integrated" and d is the amount of differencing required to make the data appear stationary. Thus d=1 indicates a first difference and d=2 indicates that the differences were themselves differenced, a "second difference." The data seemed to have d=1 (more on deciding how much differencing to use is coming up). To identify the log transformed variable, called LOANS in the data set, use this code to get the SCAN table.

```
 PROC ARIMA;
   IDENTIFY VAR=LOANS SCAN P=(0:5) Q=(0:5);
RUN;
```

The output below shows the SCAN results. They indicate several possible models.

SCAN Table for Interbank Loans

Squared Canonical Correlation Estimates

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.9976	0.9952	0.9931	0.9899	0.9868	0.9835
AR 1	<.0001	0.0037	0.0397	0.0007	0.0024	0.0308
AR 2	0.0037	0.0003	0.0317	0.0020	<.0001	0.0133
AR 3	0.0407	0.0309	0.0274	0.0126	0.0134	0.0125
AR 4	0.0004	0.0053	0.0076	0.0004	0.0022	0.0084
AR 5	0.0058	0.0003	0.0067	0.0022	<.0001	0.0078

SCAN Chi-Square[1] Probability Values

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
AR 1	0.9125	0.2653	0.0003	0.6474	0.3936	0.0019
AR 2	0.2618	0.7467	0.0033	0.4419	0.9227	0.0940
AR 3	0.0002	0.0043	0.0136	0.0856	0.0881	0.1302
AR 4	0.7231	0.1942	0.1562	0.7588	0.4753	0.1589
AR 5	0.1613	0.7678	0.1901	0.4709	0.9708	0.1836

The ARIMA Procedure

ARMA(p+d,q) Tentative Order Selection Tests

----SCAN---	
p+d	q
4	0
2	3

(5% Significance Level)

The SCAN table was computed on log transformed but not differenced data. Therefore, assuming d=1 as appears to be true for this data, the listed number p+d represents p+1 and SCAN suggests ARIMA(3,1,0) or ARIMA(1,1,3).

```
IDENTIFY VAR=LOANS(1);
  ESTIMATE P=3     ML;
  ESTIMATE P=1 Q=3 ML;
 RUN;
```

The Chi-Square checks for both of these models are insignificant at all lags, indicating both models fit well. Both models have some insignificant parameters and could be refined by omitting some lags if desired.