Autoregressive Error Terms in Regression

Model $Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_k X_{kt} + Z_t$

Autoregressive error $Z_t = \alpha Z_{t-1} + e_t$
e_t white noise
Could be higher order AR(p)

Case 1: α known

$Z_t = Y_t - (\beta_0 + \beta_1 X_{1t} + \cdots + \beta_k X_{kt})$
$\alpha Z_{t-1} = \alpha Y_{t-1} - (\alpha\beta_0 + \alpha\beta_1 X_{1,t-1} + \cdots + \alpha\beta_k X_{k,t-1})$

subtract

$Z_t - \alpha Z_{t-1} = Y_t - \alpha Y_{t-1} - (\beta_0 - \alpha\beta_0) - \beta_1(X_{1t} - \alpha X_{1,t-1}) + \cdots + \beta_k(X_{kt} - \alpha X_{k,t-1})$
$\underline{Y_t - \alpha Y_{t-1}} = \beta_0 \underline{(1-\alpha)} + \beta_1(\underline{X_{1t} - \alpha X_{1,t-1}}) + \cdots + \beta_k(\underline{X_{kt} - \alpha X_{k,t-1}}) + e_t$

"New variables" as underlined.
New regression satisfies OLS assumptions (error is e_t)!
New regression has same coefficients as old!

What about first observation?
Z_1 independent of $e_2, e_3, ..., e_n$
$\sqrt{1-\alpha^2}\, Z_1$ independent of $e_2, e_3, ..., e_n$ and $\mathrm{Var}\{\sqrt{1-\alpha^2}\, Z_1\} = \sigma^2$
$\sqrt{1-\alpha^2}\, Y_1 = \beta_0\sqrt{1-\alpha^2} + \beta_1 X_{11}\sqrt{1-\alpha^2} + \cdots + \beta_k X_{k1}\sqrt{1-\alpha^2} + \sqrt{1-\alpha^2}\, Z_1$

"dependent"	<-------	-"independent"-	--	------------>
$\sqrt{1-\alpha^2}\, Y_1$	$\sqrt{1-\alpha^2}$	$X_{1,1}\sqrt{1-\alpha^2}$	$\cdots$	$X_{k,1}\sqrt{1-\alpha^2}$
$Y_2-\alpha Y_1$	$1-\alpha$	$X_{1,2}-\alpha X_{1,1}$		$X_{k,2}-\alpha X_{k,1}$
$Y_3-\alpha Y_2$	$1-\alpha$	$X_{1,3}-\alpha X_{1,2}$		$X_{k,3}-\alpha X_{k,2}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$Y_n-\alpha Y_{n-1}$	$1-\alpha$	$X_{1,n}-\alpha X_{1,n-1}$		$X_{k,n}-\alpha X_{k,n-1}$

Case 1A: α known, order 2

$Z_t = \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + e_t$

$V_1 = c\, Z_1$, where $c = \sigma/\sqrt{\gamma(0)}$, has variance σ^2.

$V_2 = a\, Z_2 + b\, Z_1$ has variance $(a^2 + b^2)\gamma(0) + 2ab\,\gamma(1)$. Set this = σ^2.
$E\{ Z_1 V_2 \} = a\,\gamma(1) + b\gamma(0)$. Set this = 0.
2 equations in 2 unknowns - solve for a,b.

$V_3 = e_3 = Z_3 - \alpha_1 Z_2 - \alpha_2 Z_1$ independent of Z_1, Z_2; variance is σ^2.
etc.

$$\text{New } \mathbf{Y} \text{ column} = \begin{pmatrix} c & 0 & 0 & & 0 \\ b & a & 0 & \cdots & 0 \\ -\alpha_2 & -\alpha_1 & 1 & & 0 \\ & \vdots & & & \\ 0 & 0 & 0 & & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{pmatrix}$$

Same for intercept, Xs. Regress new **Y** on new **X** matrix.

Note: This is just an algorithm for doing GLS. The matrix

$$\boldsymbol{C}'\boldsymbol{C} = \begin{pmatrix} c & 0 & 0 & & 0 \\ b & a & 0 & \cdots & 0 \\ -\alpha_2 & -\alpha_1 & 1 & & 0 \\ & \vdots & & & \\ 0 & 0 & 0 & & 1 \end{pmatrix}' \begin{pmatrix} c & 0 & 0 & & 0 \\ b & a & 0 & \cdots & 0 \\ -\alpha_2 & -\alpha_1 & 1 & & 0 \\ & \vdots & & & \\ 0 & 0 & 0 & & 1 \end{pmatrix}$$

is inverse of $\boldsymbol{\Gamma}/\sigma^2$ where $\boldsymbol{\Gamma}$ is variance-covariance matrix of Z_t's. Thus $(\boldsymbol{X}'\boldsymbol{G}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{G}^{-1}\boldsymbol{Y}) = (\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{C}\,\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{C}\,\boldsymbol{Y})$ which is why regressing $\boldsymbol{C}\,\boldsymbol{Y}$ on $\boldsymbol{C}\,\boldsymbol{X}$ delivers the desired matrix.

Case 2: AR( ) parameters unknown.
Plan:

1. OLS regression **Y** on **X** => residuals
2. Fit AR(p) to residuals
3. Apply case 1 algorithm with **E**stimated AR parameters (EGLS)
4. Use the EGLS estimates to compute new residuals and iterate.

Fuller (Sec. 9.1 and 9.3) gives rather mild conditions under which the above scheme is asymptotically justified (i.e. the transformed regression statistics are asymptotically multivariate normal etc.) when the error series is stationary.

Note 1: If $\alpha = 1$ (residuals are random walk) then the whole model can be specified in differences. Note that we lose the first observation and the ability to estimate the intercept. Nevertheless, computing $Y_t - Y_{t-1}$ as a "new Y" and similarly for the X's is seen to give an equation in transformed variables that has the same β's as the original equation and has white noise errors.

Note 2: In assessing the contribution of our X's in predicting Y (especially if it is expensive to collect the X information) we note that $\hat{Y}_{t+1} = \hat{\beta}_0 + \hat{\beta}_1 X_{1,t+1} + \cdots + \hat{\beta}_k X_{k,t+1} + \hat{Z}_{t+1}$ where $\hat{Z}_{t+1}$ would be computed from recent residuals and the estimated AR(p) model. Thus if we divide the error sum of squares (errors from these predictions) by the total sum of squares then subtract from 1, we get a "Total R^2" much of which might come from the momentum in the Z's, especially if they form a nearly nonstationary process.

SAS PROC AUTOREG will also produce a "Regression R^2" which is the R^2 from the regression on the transformed variables and is a better measure of how much you are getting from just the X's.

Note 3: Durbin and Watson suggested the statistic $d = \frac{\sum_{2}^{n}(\hat{Z}_t - \hat{Z}_{t-1})^2}{\sum_{1}^{n}\hat{Z}_t^2}$ as a measure of autocorrelation. If the Z's are uncorrelated, and we divide numerator and denominator by n, we see that dw converges to 2, so a dw near 2 implies lack of autocorrelation at lag 1. 0<dw<2 implies positive lag 1 autocorrelation and 2<dw<4 implies negative. The dw theory does not allow lagged Ys in the model.

Note 4: The regression in the table above is a nonlinear function of α and the k+1 β's. As such it can be estimated by a general nonlinear likelihood maximization scheme. That is, we can iterate on all the parameters simultaneously rather than iterating on the βs and α separately.

(1) Run OLS on regression
  Estimators: Unbiased (because E{Z}=0}
  Consistent (converge to true βs)

(2) Regress residuals on lagged residuals to ESTIMATE AR coefficients

(3) Proceed as thought estimated AR = true AR.

Example 1: NC Retail Sales (quarterly)

```
  Plot of qsales*date.  Symbol is value of qtr.

qsales |
       |                              4
       |                              3
 20000 +                            42
       |                           231
       |                 23    4 24
       |              423  42423
 15000 +           3421  1 1  1 1
       |       4232
       |    4 2   1 1
       |   231 1
 10000 +   1
       |
       -+------------+------------+------------+-
      81Q4         87Q2         92Q4         98Q2
```

Seasonal (qtr 1 low, qtr 4 high) Use dummy variables

Trend: Use t t^2 t^3

Autocorrelation: Try 5 lags (more than a year) and let SAS omit insignificant ones.

```
Data NCSALES;
input qsales t t2 s1 s2 s3 s4 date :yyq6.;
if t2 = . then t2=t*t;
t3=t8t2;
qtr=qtr(date);
format date YYQ4.;
cards;
         9485.68     2       4     1     0     0     0     1983Q1
        11164.09     3       9     0     1     0     0     1983Q2
        11797.33     4      16     0     0     1     0     1983Q3
        12459.68     5      25     0     0     0     1     1983Q4
        11176.09     6      36     1     0     0     0     1984Q1
        12854.22     7      49     0     1     0     0     1984Q2
                        (more data here)
        21130.73    48    2304     0     0     1     0     1994Q3
        22657.68    49    2401     0     0     0     1     1994Q4
             .      50    2500     1     0     0     0     1995Q1
             .      51    2601     0     1     0     0     1995Q2
             .      52    2704     0     0     1     0     1995Q3
             .      53    2809     0     0     0     1     1995Q4
   ;
proc reg; model qsales = t S1 S2 S3/dwprob;
output out=out1 p=pred lcl=l ucl=u;  run;
```

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	9556.93236	379.69796	25.17	<.0001
t	1	612.17585	57.41516	10.66	<.0001
t2	1	-21.51050	2.57455	-8.36	<.0001
t3	1	0.28845	0.03324	8.68	<.0001
s1	1	-1644.06586	196.35286	-8.37	<.0001
s2	1	-23.23254	195.28585	-0.12	0.9059
s3	1	-124.05869	194.64734	-0.64	0.5274

```
                    The REG Procedure
                     Model: MODEL1
               Dependent Variable: qsales

         Durbin-Watson D                 0.584
         Pr < DW                        <.0001
         Pr > DW                        1.0000
         Number of Observations             48
         1st Order Autocorrelation       0.662

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and
      Pr>DW is the p-value for testing negative autocorrelation.
```

```
proc autoreg data=NCSales;
 model qsales = t t2 t3 S1 S2 S3 / nlag=5 backstep dwprob;
run;
```

```
                   The AUTOREG Procedure

               Dependent Variable    qsales


             Ordinary Least Squares Estimates

      SSE                  9299521.8    DFE                      41
      MSE                     226818    Root MSE          476.25372
      SBC                 747.681588    AIC               734.58318
      Regress R-Square        0.9737    Total R-Square       0.9737
      Durbin-Watson           0.5836    Pr < DW              <.0001
      Pr > DW                 1.0000
NOTE: Pr<DW is the p-value for testing positive autocorrelation, and
      Pr>DW is the p-value for testing negative autocorrelation.
```

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	9557	379.6980	25.17	<.0001
t	1	612.1759	57.4152	10.66	<.0001
t2	1	-21.5105	2.5745	-8.36	<.0001
t3	1	0.2884	0.0332	8.68	<.0001
s1	1	-1644	196.3529	-8.37	<.0001
s2	1	-23.2325	195.2859	-0.12	0.9059
s3	1	-124.0587	194.6473	-0.64	0.5274

Estimates of Autocorrelations

```
Lag Covariance  Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

 0     193740     1.000000 |                    |********************|
 1     128334     0.662404 |                    |*************       |
 2     109532     0.565354 |                    |***********         |
 3    88836.7     0.458535 |                    |*********           |
 4    48426.2     0.249955 |                    |*****               |
 5    27365.1     0.141246 |                    |***                 |
```

Note: First part same as OLS. Shows strong autocorrelation.

The AUTOREG Procedure

Backward Elimination of Autoregressive Terms

Lag	Estimate	t Value	Pr > |t|
5	0.074212	0.45	0.6579
3	-0.152076	-0.85	0.4014
4	0.168893	1.19	0.2425
2	-0.225536	-1.45	0.1562

Preliminary MSE 108731

```
                 Estimates of Autoregressive Parameters

                                         Standard
             Lag      Coefficient           Error    t Value
               1        -0.662404        0.118451      -5.59

                           Yule-Walker Estimates
      SSE                  4736903.38    DFE                     40
      MSE                      118423    Root MSE         344.12583
      SBC                  719.750619    AIC             704.781011
      Regress R-Square         0.9558    Total R-Square      0.9866
      Durbin-Watson            1.9768    Pr < DW             0.3409
      Pr > DW                  0.6591
NOTE: Pr<DW is the p-value for testing positive autocorrelation, and
      Pr>DW is the p-value for testing negative autocorrelation.


                                      Standard                 Approx
   Variable        DF     Estimate       Error    t Value    Pr > |t|

   Intercept        1         9668    558.1976      17.32      <.0001
   t                1     613.2735     92.5406       6.63      <.0001
   t2               1     -22.3063      4.2045      -5.31      <.0001
   t3               1       0.3061      0.0542       5.65      <.0001
   s1               1        -1620    105.4686     -15.36      <.0001
   s2               1     -11.0318    118.3319      -0.09      0.9262
   s3               1    -119.7767    102.4078      -1.17      0.2491
```

Tests are now valid (asymptotically).

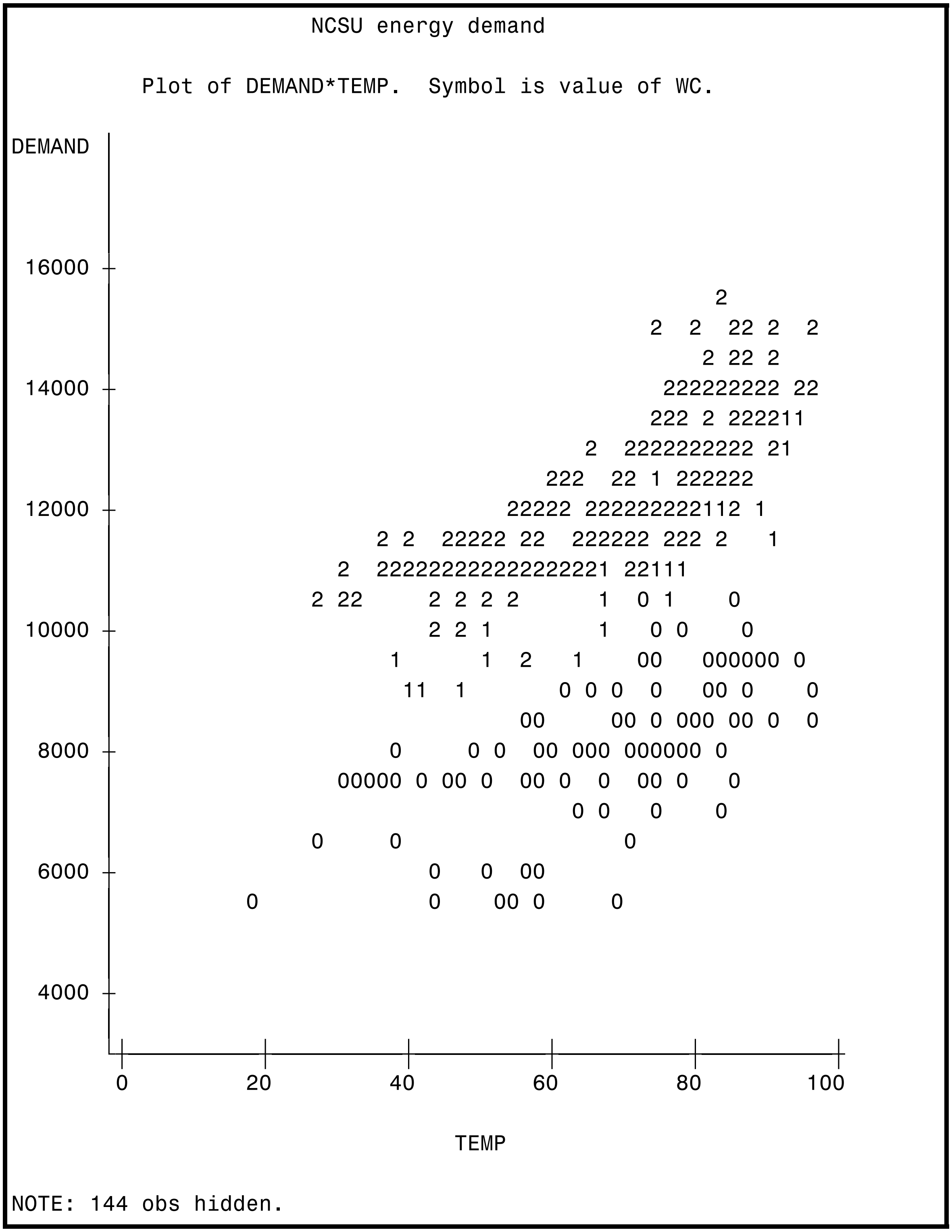
Example 2: NCSU energy usage

Variables:

Demand: Energy used on campus
Class: Classes held that day (dummy variable)
Work: All work days (includes classes)
Temperature
Weekday dummies (base=Wednesday). These add little to R^2

```
  /* ---------------------------------------------------------------
  |  NCSU campuswide energy consumption 1980-81 school year.        |
  |  Indicators for workdays and class days.                        |
   ----------------------------------------------------------------*/

proc format; value dow 1="Sun" 2="Mon" 3="Tues" 4="Wed"
      5="Thurs" 6="Fri" 7="Sat";
options pagesize=55 ls=64;
data energy;
input  DAY TEMP DEMAND T  WORK  CLASS  DATE :date7.  Y 67 ;
   put @20 date;
TEMPSQ=(TEMP-65)**2;
*** compare all days to Wednesday ***;
SAT   = (T=7);   SUN  = (T=1);
MON   = (T=2);   TUES = (T=3);
THURS = (T=5);   FRI  = (T=6);
 format date date7.;   WC = work+class;
cval="green";
 if WC=0 then cval="red";
 if WC=1 then cval = "cyan";
 cards;
          1     83       8217    1      0       0      01JUL79    2
          2     87      12545    2      1       1      02JUL79    1
          3     85      12649    3      1       1      03JUL79    1

                           (more data)

         30     88      13789    2      1       1      30JUN80    1
;

PROC GPLOT data=energy; PLOT DEMAND*TEMP=WC;
  title 'NCSU Energy Demand';
  title2 c=red "Off  " c=cyan "  Work, no Class " c=green " Class ";
  Symbol1 v=dot i=none c=red;
  Symbol2 v=dot i=none c=cyan;;
  symbol3 v=dot i=none c=green;

PROC GPLOT data=energy; PLOT DEMAND*T=WC  ;
        format t dow.;

PROC AUTOREG data=energy; MODEL DEMAND = TEMP TEMPSQ CLASS WORK
   SAT--FRI   /NLAG=15 BACKSTEP;
   ** day of week adds about .002 to R**2  ***;
run;
```

A simple plot of energy usage versus temperature using as a symbol the sum of the class and workday seasonal dummies (0-> not a work day, 1-> work day without classes, 2-> a class day).



Autoreg Procedure

Dependent Variable = DEMAND

Ordinary Least Squares Estimates

SSE	2.5387E8	DFE	361
MSE	703238.1	Root MSE	838.5929
SBC	5990.765	AIC	5971.252
Reg Rsq	0.8680	Total Rsq	0.8680
Durbin-Watson	0.6324		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	3856.671264	206.8	18.650	0.0001
TEMP	1	57.833448	2.6720	21.644	0.0001
TEMPSQ	1	0.824102	0.1523	5.412	0.0001
CLASS	1	1545.669155	157.1	9.841	0.0001
WORK	1	2638.304419	165.6	15.934	0.0001

Estimates of Autocorrelations

Lag	Covariance	Correlation
0	693631.1	1.000000
1	473562.1	0.682729
2	367778.3	0.530222
3	308403.7	0.444622
4	307188	0.442869
5	333956.1	0.481461
6	342626.4	0.493961
7	365283.9	0.526626
8	298846.6	0.430844
9	299775.9	0.432183
10	252863.5	0.364550
11	221802.3	0.319770
12	221137.3	0.318811
13	220904.3	0.318475
14	276792.2	0.399048
15	231104.3	0.333180

Backward Elimination of Autoregressive Terms

Lag	Estimate	t-Ratio	Prob
2	0.002017	0.0335	0.9733
12	-0.005618	-0.0938	0.9253
3	0.004789	0.0890	0.9291
10	0.054058	0.9333	0.3513
4	-0.060362	-1.1469	0.2522
6	-0.079523	-1.3325	0.1836
13	0.057813	1.0993	0.2724
11	0.076009	1.6618	0.0974
15	0.098888	1.8942	0.0590

Preliminary MSE = 314640.5

Estimates of the Autoregressive Parameters

Lag	Coefficient	Std Error	t Ratio
1	-0.54064918	0.044220	-12.226
5	-0.11926166	0.044812	-2.661
7	-0.20036592	0.054040	-3.708
8	0.16453292	0.058817	2.797
9	-0.10386205	0.050991	-2.037
14	-0.10107126	0.044566	-2.268

Expected Autocorrelations

Lag	Autocorr
0	1.0000
1	0.6883
2	0.5369
3	0.4527
4	0.4365
5	0.4771
6	0.4799
7	0.5188
8	0.4278
9	0.4298
10	0.4072
11	0.3919
12	0.3988
13	0.3992
14	0.4427

```
                    Autoreg Procedure
                  Yule-Walker Estimates

         SSE            96233172    DFE              355
         MSE            271079.4    Root MSE    520.6528
         SBC            5672.801    AIC         5629.872
         Reg Rsq          0.8881    Total Rsq     0.9500
         Durbin-Watson    1.7564

Variable     DF        B Value   Std Error   t Ratio Approx Prob
Intercept     1    5857.458298       374.1    15.659      0.0001
TEMP          1      31.686338      3.6911     8.585      0.0001
TEMPSQ        1       0.687270      0.1232     5.579      0.0001
CLASS         1    1185.053609       119.9     9.882      0.0001
WORK          1    2763.301101       123.7    22.345      0.0001

                  Expected Autocorrelations
                         Lag Autocorr
                           0   1.0000
                           1   0.6883
                           2   0.5369
                           3   0.4527
                           4   0.4365
                           5   0.4771
                           6   0.4799
                           7   0.5188
                           8   0.4278
                           9   0.4298
                          10   0.4072
                          11   0.3919
                          12   0.3988
                          13   0.3992
                          14   0.4427
```

Note: AUTOREG assumes model is $Z_t + \alpha_1 Z_{t-1} + \cdots + \alpha_p Z_{t-p} = e_t$. Parameter estimates have opposite signs from Box-Jenkins ARIMA parameters.

Note: AUTOREG is the only SAS/ETS procedure that includes a variance for the regression parameters in its prediction interval computations. Adds $\mathbf{X}_{t+j}\hat{\beta}$ variance to $\hat{Z}_{t+j}$ prediction error variance, but future **X** matrix row, $\mathbf{X}_{t+j}$ treated as a *constant* but is usually really a forecast in practice. Demos: NCSales.sas and Energy.sas