SAS/ETS® 14.3 User's Guide

The PDLREG Procedure

This document is an individual chapter from *SAS/ETS® 14.3 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2017. *SAS/ETS® 14.3 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/ETS® 14.3 User's Guide

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

September 2017

Chapter 27
The PDLREG Procedure

Contents

Overview: PDLREG Procedure

The PDLREG procedure estimates regression models for time series data in which the effects of some of the regressor variables are distributed across time. The distributed lag model assumes that the effect of an input variable X on an output Y is distributed over time. If you change the value of X at time t, Y will experience some immediate effect at time t, and it will also experience a delayed effect at times $t + 1$, $t + 2$, and so on up to time $t + p$ for some limit p.

The regression model supported by PROC PDLREG can include any number of regressors with distribution lags and any number of covariates. (Simple regressors without lag distributions are called covariates.) For example, the two-regressor model with a distributed lag effect for one regressor is written

$$y_t = \alpha + \sum_{i=0}^{p} \beta_i x_{t-i} + \gamma z_t + u_t$$

Here, x_t is the regressor with a distributed lag effect, z_t is a simple covariate, and u_t is an error term.

The distribution of the lagged effects is modeled by Almon lag polynomials. The coefficients b_i of the lagged values of the regressor are assumed to lie on a polynomial curve. That is,

$$b_i = \alpha_0^* + \sum_{j=1}^{d} \alpha_j^* i^j$$

where $d (\leq p)$ is the degree of the polynomial. For the numerically efficient estimation, the PDLREG procedure uses *orthogonal polynomials*. The preceding equation can be transformed into orthogonal polynomials,

$$b_i = \alpha_0 + \sum_{j=1}^{d} \alpha_j f_j(i)$$

where $f_j(i)$ is a polynomial of degree j in the lag length i, and α_j is a coefficient estimated from the data.

The PDLREG procedure supports endpoint restrictions for the polynomial. That is, you can constrain the estimated polynomial lag distribution curve so that $b_{-1} = 0$ or $b_{p+1} = 0$, or both. You can also impose linear restrictions on the parameter estimates for the covariates.

You can specify a minimum degree and a maximum degree for the lag distribution polynomial, and the procedure fits polynomials for all degrees in the specified range. (However, if distributed lags are specified for more that one regressor, you can specify a range of degrees for only one of them.)

The PDLREG procedure can also test for autocorrelated residuals and perform autocorrelated error correction by using the autoregressive error model. You can specify any order autoregressive error model and can specify several different estimation methods for the autoregressive model, including exact maximum likelihood.

The PDLREG procedure computes generalized Durbin-Watson statistics to test for autocorrelated residuals. For models with lagged dependent variables, the procedure can produce Durbin h and Durbin t statistics. You can request significance level p-values for the Durbin-Watson, Durbin h, and Durbin t statistics. For more information about these statistics, see Chapter 8, "The AUTOREG Procedure."

The PDLREG procedure assumes that the input observations form a time series. Thus, the PDLREG procedure should be used only for ordered and equally spaced time series data.

Getting Started: PDLREG Procedure

Use the MODEL statement to specify the regression model. The PDLREG procedure's MODEL statement is written like MODEL statements in other SAS regression procedures, except that a regressor can be followed by a lag distribution specification enclosed in parentheses.

For example, the following MODEL statement regresses Y on X and Z and specifies a distributed lag for X:

```
model y = x(4,2) z;
```

The notation X(4,2) specifies that the model includes X and 4 lags of X, with the coefficients of X and its lags constrained to follow a second-degree (quadratic) polynomial. Thus, the regression model specified by this MODEL statement is

$$y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + b_3 x_{t-3} + b_4 x_{t-4} + c z_t + u_t$$

$$b_i = \alpha_0 + \alpha_1 f_1(i) + \alpha_2 f_2(i)$$

where $f_1(i)$ is a polynomial of degree 1 in i and $f_2(i)$ is a polynomial of degree 2 in i.

Lag distribution specifications are enclosed in parentheses and follow the name of the regressor variable. The general form of the lag distribution specification is

regressor-name (length, degree, minimum-degree, end-constraint)

where

length	is the length of the lag distribution—that is, the number of lags of the regressor to use.
degree	is the degree of the distribution polynomial.
minimum-degree	is an optional minimum degree for the distribution polynomial.
end-constraint	is an optional endpoint restriction specification, which can have the value FIRST, LAST, or BOTH.

If the *minimum-degree* option is specified, the PDLREG procedure estimates models for all degrees between *minimum-degree* and *degree*.

Introductory Example

The following statements generate simulated data for variables Y and X. Y depends on the first three lags of X, with coefficients .25, .5, and .25. Thus, the effect of changes of X on Y takes effect 25% after one period, 75% after two periods, and 100% after three periods.

```
data test;
   xl1 = 0; xl2 = 0; xl3 = 0;
   do t = -3 to 100;
      x = ranuni(1234);
      y = 10 + .25 * xl1 + .5 * xl2 + .25 * xl3
             + .1 * rannor(1234);
      if t > 0 then output;
      xl3 = xl2; xl2 = xl1; xl1 = x;
   end;
run;
```

The following statements use the PDLREG procedure to regress Y on a distributed lag of X. The length of the lag distribution is 4, and the degree of the distribution polynomial is specified as 3.

```
proc pdlreg data=test;
   model y = x( 4, 3 );
run;
```

The PDLREG procedure first prints a table of statistics for the residuals of the model, as shown in Figure 27.1. For an explanation of these statistics, see Chapter 8, “The AUTOREG Procedure.”

Figure 27.1 Residual Statistics

The PDLREG Procedure

Dependent Variable	y

Ordinary Least Squares Estimates			
SSE	0.86604442	**DFE**	91
MSE	0.00952	**Root MSE**	0.09755
SBC	-156.72612	**AIC**	-169.54786
MAE	0.07761107	**AICC**	-168.88119
MAPE	0.73971576	**HQC**	-164.3651
Durbin-Watson	1.9920	**Total R-Square**	0.7711

The PDLREG procedure next prints a table of parameter estimates, standard errors, and t tests, as shown in Figure 27.2.

Figure 27.2 Parameter Estimates

Parameter Estimates					
Variable	**DF**	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**
Intercept	1	10.0030	0.0431	231.87	<.0001
x0**	1	0.4406	0.0378	11.66	<.0001
x1**	1	0.0113	0.0336	0.34	0.7377
x2**	1	-0.4108	0.0322	-12.75	<.0001
x3**	1	0.0331	0.0392	0.84	0.4007

The table in Figure 27.2 shows the model intercept and the estimated parameters of the lag distribution polynomial. The parameter labeled X**0 is the constant term, α_0, of the distribution polynomial. X**1 is the linear coefficient, α_1; X**2 is the quadratic coefficient, α_2; and X**3 is the cubic coefficient, α_3.

The parameter estimates for the distribution polynomial are not of interest in themselves. Since the PDLREG procedure does not print the orthogonal polynomial basis that it constructs to represent the distribution polynomial, these coefficient values cannot be interpreted.

However, because these estimates are for an orthogonal basis, you can use these results to test the degree of the polynomial. For example, this table shows that the X**3 estimate is not significant; the p-value for its t ratio is 0.4007, while the X**2 estimate is highly significant ($p < .0001$). This indicates that a second-degree polynomial might be more appropriate for this data set.

The PDLREG procedure next prints the lag distribution coefficients and a graphical display of these coefficients, as shown in Figure 27.3.

Figure 27.3 Coefficients and Graph of Estimated Lag Distribution

<table>
<tr><th colspan="6">Estimate of Lag Distribution</th></tr>
<tr><th>Variable</th><th>Estimate</th><th>Standard Error</th><th>t Value</th><th>Approx Pr > |t|</th><th>-0.04 0.4167</th></tr>
<tr><td>x(0)</td><td>-0.040150</td><td>0.0360</td><td>-1.12</td><td>0.2677</td><td>|***| |</td></tr>
<tr><td>x(1)</td><td>0.324241</td><td>0.0307</td><td>10.55</td><td><.0001</td><td>| |***************************** |</td></tr>
<tr><td>x(2)</td><td>0.416661</td><td>0.0239</td><td>17.45</td><td><.0001</td><td>| |*************************************|</td></tr>
<tr><td>x(3)</td><td>0.289482</td><td>0.0315</td><td>9.20</td><td><.0001</td><td>| |************************** |</td></tr>
<tr><td>x(4)</td><td>-0.004926</td><td>0.0365</td><td>-0.13</td><td>0.8929</td><td>| | |</td></tr>
</table>

The lag distribution coefficients are the coefficients of the lagged values of X in the regression model. These coefficients lie on the polynomial curve defined by the parameters shown in Figure 27.2. Note that the estimated values for X(1), X(2), and X(3) are highly significant, while X(0) and X(4) are not significantly different from 0. These estimates are reasonably close to the true values used to generate the simulated data.

The graphical display of the lag distribution coefficients plots the estimated lag distribution polynomial reported in Figure 27.2. The roughly quadratic shape of this plot is another indication that a third-degree distribution curve is not needed for this data set.

Syntax: PDLREG Procedure

The following statements can be used with the PDLREG procedure:

PROC PDLREG *option* **;**
  BY *variables* **;**
  MODEL *dependent = effects / options* **;**
  OUTPUT OUT= *SAS-data-set keyword = variables* **;**
  RESTRICT *restrictions* **;**

Functional Summary

The statements and options used with the PDLREG procedure are summarized in Table 27.1.

Table 27.1 Functional Summary

Description	Statement	Option
Data Set Options		
Specify the input data set	PROC PDLREG	DATA=
Write predicted values to an output data set	OUTPUT	OUT=
BY-Group Processing		
Specify BY-group processing	BY	
Printing Control Options		
Request all print options	MODEL	ALL

Table 27.1 *continued*

Description	Statement	Option
Print transformed coefficients	MODEL	COEF
Print correlations of the estimates	MODEL	CORRB
Print covariances of the estimates	MODEL	COVB
Print DW statistics up to order *j*	MODEL	DW=*j*
Print the marginal probability of DW statistics	MODEL	DWPROB
Print inverse of Toeplitz matrix	MODEL	GINV
Print inverse of the crossproducts matrix	MODEL	I
Print details at each iteration step	MODEL	ITPRINT
Print Durbin *t* statistic	MODEL	LAGDEP
Print Durbin *h* statistic	MODEL	LAGDEP=
Suppress printed output	MODEL	NOPRINT
Print partial autocorrelations	MODEL	PARTIAL
Print standardized parameter estimates	MODEL	STB
Print crossproducts matrix	MODEL	XPX
Model Estimation Options		
Specify order of autoregressive process	MODEL	NLAG=
Suppress intercept parameter	MODEL	NOINT
Specify convergence criterion	MODEL	CONVERGE=
Specify maximum number of iterations	MODEL	MAXITER=
Specify estimation method	MODEL	METHOD=
Output Control Options		
Specify confidence limit size	OUTPUT	ALPHACLI=
Specify confidence limit size for structural predicted values	OUTPUT	ALPHACLM=
Output transformed intercept variable	OUTPUT	CONSTANT=
Output lower confidence limit for predicted values	OUTPUT	LCL=
Output lower confidence limit for structural predicted values	OUTPUT	LCLM=
Output predicted values	OUTPUT	P=
Output predicted values of the structural part	OUTPUT	PM=
Output residuals from the predicted values	OUTPUT	R=
Output residuals from the structural predicted values	OUTPUT	RM=
Output transformed variables	OUTPUT	TRANSFORM=
Output upper confidence limit for the predicted values	OUTPUT	UCL=
Output upper confidence limit for the structural predicted values	OUTPUT	UCLM=

PROC PDLREG Statement

PROC PDLREG *option* **;**

The PROC PDLREG statement has the following option:

DATA=*SAS-data-set*
specifies the name of the SAS data set containing the input data. If you do not specify the DATA= option, the most recently created SAS data set is used.

In addition, you can place any of the following MODEL statement options in the PROC PDLREG statement, which is equivalent to specifying the option for every MODEL statement: ALL, COEF, CONVERGE=, CORRB, COVB, DW=, DWPROB, GINV, ITPRINT, MAXITER=, METHOD=, NOINT, NOPRINT, and PARTIAL.

BY Statement

BY *variables* **;**

A BY statement can be used with PROC PDLREG to obtain separate analyses on observations in groups defined by the BY variables.

MODEL Statement

MODEL *dependent = effects / options* **;**

The MODEL statement specifies the regression model. The keyword MODEL is followed by the dependent variable name, an equal sign, and a list of independent *effects*. Only one MODEL statement is allowed.

Every variable in the model must be a numeric variable in the input data set. Specify an independent effect with a variable name optionally followed by a polynomial lag distribution specification.

Specifying Independent Effects

The general form of an *effect* is

variable **(***length, degree, minimum-degree, constraint* **)**

The term in parentheses following the variable name specifies a polynomial distributed lag (PDL) for the variable. The PDL specification is as follows:

length specifies the number of lags of the variable to include in the lag distribution.

degree specifies the maximum degree of the distribution polynomial. If not specified, the degree defaults to the lag length.

minimum-degree specifies the minimum degree of the polynomial. By default *minimum-degree* is the same as *degree*.

constraint specifies endpoint restrictions on the polynomial. The value of *constraint* can be FIRST, LAST, or BOTH. If a value is not specified, there are no endpoint restrictions.

If you do not specify the *degree* or *minimum-degree* parameter, but you do specify endpoint restrictions, you must use commas to show which parameter, *degree* or *minimum-degree*, is left out.

MODEL Statement Options

The following options can appear in the MODEL statement after a slash (/).

ALL
prints all the matrices computed during the analysis of the model.

COEF
prints the transformation coefficients for the first p observations. These coefficients are formed from a scalar multiplied by the inverse of the Cholesky root of the Toeplitz matrix of autocovariances.

CORRB
prints the matrix of estimated correlations between the parameter estimates.

COVB
prints the matrix of estimated covariances between the parameter estimates.

DW=*j*
prints the generalized Durbin-Watson statistics up to the order of j. The default is DW=1. When you specify the LAGDEP or LAGDEP=*name* option, the Durbin-Watson statistic is not printed unless you specify the DW= option.

DWPROB
prints the marginal probability of the Durbin-Watson statistic.

CONVERGE=*value*
sets the convergence criterion. If the maximum absolute value of the change in the autoregressive parameter estimates between iterations is less than this amount, then convergence is assumed. The default is CONVERGE=0.001.

GINV
prints the inverse of the Toeplitz matrix of autocovariances for the Yule-Walker solution.

I
prints $(\mathbf{X}'\mathbf{X})^{-1}$, the inverse of the crossproducts matrix for the model; or, if restrictions are specified, it prints $(\mathbf{X}'\mathbf{X})^{-1}$ adjusted for the restrictions.

ITPRINT
prints information on each iteration.

LAGDEP
LAGDV
prints the t statistic for testing residual autocorrelation when regressors contain lagged dependent variables.

LAGDEP=*name*

LAGDV=*name*

prints the Durbin h statistic for testing the presence of first-order autocorrelation when regressors contain the lagged dependent variable whose name is specified as LAGDEP=*name*. When the h statistic cannot be computed, the asymptotically equivalent t statistic is given.

MAXITER=*number*

sets the maximum number of iterations allowed. The default is MAXITER=50.

METHOD=*value*

specifies the type of estimates for the autoregressive component. The values of the METHOD= option are as follows:

ML	specifies the maximum likelihood method.
ULS	specifies unconditional least squares.
YW	specifies the Yule-Walker method.
ITYW	specifies iterative Yule-Walker estimates.

The default is METHOD=ML if you specified the LAGDEP or LAGDEP= option; otherwise, METHOD=YW is the default.

NLAG=*m*

NLAG=(*number-list* **)**

specifies the order of the autoregressive process or the subset of autoregressive lags to be fit. If you do not specify the NLAG= option, PROC PDLREG does not fit an autoregressive model.

NOINT

suppresses the intercept parameter from the model.

NOPRINT

suppresses the printed output.

PARTIAL

prints partial autocorrelations if the NLAG= option is specified.

STB

prints standardized parameter estimates. Sometimes known as a standard partial regression coefficient, a standardized parameter estimate is a parameter estimate multiplied by the standard deviation of the associated regressor and divided by the standard deviation of the regressed variable.

XPX

prints the crossproducts matrix, $\mathbf{X}'\mathbf{X}$, used for the model. $\mathbf{X}$ refers to the transformed matrix of regressors for the regression.

OUTPUT Statement

OUTPUT OUT=*SAS-data-set keyword=option . . .* **;**

The OUTPUT statement creates an output SAS data set that contains variables as specified by the following keyword options. For a description of the associated computations for these options, see the section "Predicted Values" in Chapter 8, "The AUTOREG Procedure."

ALPHACLI=*number*
: sets the confidence limit size for the estimates of future values of the current realization of the response time series to *number*, where *number* is less than one and greater than zero. The resulting confidence interval has 1–*number* confidence. The default value for *number* is 0.05, corresponding to a 95% confidence interval.

ALPHACLM=*number*
: sets the confidence limit size for the estimates of the structural or regression part of the model to *number*, where *number* is less than one and greater than zero. The resulting confidence interval has 1–*number* confidence. The default value for *number* is 0.05, corresponding to a 95% confidence interval.

OUT=*SAS-data-set*
: names the output data.

The following specifications are of the form *keyword=names*, where *keyword* specifies the statistic to include in the output data set and *names* gives names to the variables that contain the statistics.

CONSTANT=*variable*
: writes the transformed intercept to the output data set.

LCL=*name*
: requests that the lower confidence limit for the predicted value (specified in the PREDICTED= option) be added to the output data set under *name*.

LCLM=*name*
: requests that the lower confidence limit for the structural predicted value (specified in the PREDICTEDM= option) be added to the output data set under *name*.

PREDICTED=*name*

P=*name*
: stores the predicted values in the output data set under *name*.

PREDICTEDM=*name*

PM=*name*
: stores the structural predicted values in the output data set under *name*. These values are formed from only the structural part of the model.

RESIDUAL=*name*

R=*name*
: stores the residuals from the predicted values based on both the structural and time series parts of the model in the output data set under *name*.

RESIDUALM=*name*

RM=*name*

requests that the residuals from the structural prediction be given.

TRANSFORM=*variables*

requests that the specified variables from the input data set be transformed by the autoregressive model and put in the output data set. If you need to reproduce the data suitable for reestimation, you must also transform an intercept variable. To do this, transform a variable that only takes the value 1 or use the CONSTANT= option.

UCL=*name*

stores the upper confidence limit for the predicted value (specified in the PREDICTED= option) in the output data set under *name*.

UCLM=*name*

stores the upper confidence limit for the structural predicted value (specified in the PREDICTEDM= option) in the output data set under *name*.

For example, the SAS statements

```
proc pdlreg data=a;
   model y=x1 x2;
   output out=b p=yhat r=resid;
run;
```

create an output data set named B. In addition to the input data set variables, the data set B contains the variable YHAT, whose values are predicted values of the dependent variable Y, and RESID, whose values are the residual values of Y.

RESTRICT Statement

RESTRICT *equation , . . . , equation* **;**

The RESTRICT statement places restrictions on the parameter estimates for covariates in the preceding MODEL statement. A parameter produced by a distributed lag cannot be restricted with the RESTRICT statement.

Each restriction is written as a linear equation. If you specify more than one restriction in a RESTRICT statement, the restrictions are separated by commas.

You can refer to parameters by the name of the corresponding regressor variable. Each name used in the equation must be a regressor in the preceding MODEL statement. Use the keyword INTERCEPT to refer to the intercept parameter in the model.

RESTRICT statements can be given labels. You can use labels to distinguish results for different restrictions in the printed output. Labels are specified as follows:

label : **RESTRICT** *. . .*

The following is an example of the use of the RESTRICT statement, in which the coefficients of the regressors X1 and X2 are required to sum to 1:

```
proc pdlreg data=a;
   model y = x1 x2;
   restrict x1 + x2 = 1;
run;
```

Parameter names can be multiplied by constants. When no equal sign appears, the linear combination is set equal to 0. Note that the parameters associated with the variables are restricted, not the variables themselves. Here are some examples of valid RESTRICT statements:

```
restrict x1 + x2 = 1;
restrict x1 + x2 - 1;
restrict 2 * x1 = x2 + x3 , intercept + x4 = 0;
restrict x1 = x2 = x3 = 1;
restrict 2 * x1 - x2;
```

Restricted parameter estimates are computed by introducing a Lagrangian parameter λ for each restriction (Pringle and Rayner 1971). The estimates of these Lagrangian parameters are printed in the parameter estimates table. If a restriction cannot be applied, its parameter value and degrees of freedom are listed as 0.

The Lagrangian parameter, λ, measures the sensitivity of the SSE to the restriction. If the restriction is changed by a small amount ϵ, the SSE is changed by $2\lambda\epsilon$.

The *t* ratio tests the significance of the restrictions. If λ is zero, the restricted estimates are the same as the unrestricted ones.

You can specify any number of restrictions in a RESTRICT statement, and you can use any number of RESTRICT statements. The estimates are computed subject to all restrictions specified. However, restrictions should be consistent and not redundant.

Details: PDLREG Procedure

Missing Values

The PDLREG procedure skips any observations at the beginning of the data set that have missing values. The procedure uses all observations with nonmissing values for all the independent and dependent variables such that the lag distribution has sufficient nonmissing lagged independent variables.

Polynomial Distributed Lag Estimation

The simple finite distributed lag model is expressed in the form

$$y_t = \alpha + \sum_{i=0}^{p} \beta_i x_{t-i} + \epsilon_t$$

When the lag length (p) is long, severe multicollinearity can occur. Use the Almon or *polynomial distributed lag* model to avoid this problem, since the relatively low-degree d ($\leq p$) polynomials can capture the true lag distribution. The lag coefficient can be written in the Almon polynomial lag

$$\beta_i = \alpha_0^* + \sum_{j=1}^{d} \alpha_j^* i^j$$

Emerson (1968) proposed an efficient method of constructing orthogonal polynomials from the preceding polynomial equation as

$$\beta_i = \alpha_0 + \sum_{j=1}^{d} \alpha_j f_j(i)$$

where $f_j(i)$ is a polynomial of degree j in the lag length i. The polynomials $f_j(i)$ are chosen so that they are orthogonal,

$$\sum_{i=1}^{n} w_i f_j(i) f_k(i) = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}$$

where w_i is the weighting factor, and $n = p + 1$. PROC PDLREG uses the equal weights ($w_i = 1$) for all i. To construct the orthogonal polynomials, the following recursive relation is used:

$$f_j(i) = (A_j i + B_j) f_{j-1}(i) - C_j f_{j-2}(i) j = 1, \ldots, d$$

The constants A_j, B_j, and C_j are determined as follows,

$$\begin{aligned} A_j &= \left\{ \sum_{i=1}^{n} w_i i^2 f_{j-1}^2(i) - \left(\sum_{i=1}^{n} w_i i f_{j-1}^2(i) \right)^2 \right. \\ &\quad \left. - \left(\sum_{i=1}^{n} w_i i f_{j-1}(i) f_{j-2}(i) \right)^2 \right\}^{-1/2} \\ B_j &= -A_j \sum_{i=1}^{n} w_i i f_{j-1}^2(i) \\ C_j &= A_j \sum_{i=1}^{n} w_i i f_{j-1}(i) f_{j-2}(i) \end{aligned}$$

where $f_{-1}(i) = 0$ and $f_0(i) = 1/\sqrt{\sum_{i=1}^{n} w_i}$.

PROC PDLREG estimates the orthogonal polynomial coefficients, $\alpha_0, \ldots, \alpha_d$, to compute the coefficient estimate of each independent variable (X) with distributed lags. For example, if an independent variable is specified as X(9,3), a third-degree polynomial is used to specify the distributed lag coefficients. The third-degree polynomial is fit as a constant term, a linear term, a quadratic term, and a cubic term. The four terms are constructed to be orthogonal. In the output produced by the PDLREG procedure for this case, parameter estimates with names X**0, X**1, X**2, and X**3 correspond to $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2$, and $\hat{\alpha}_3$, respectively. A test using the t statistic and the approximate p-value ("Approx Pr > $|t|$") associated with X**3 can determine whether a second-degree polynomial rather than a third-degree polynomial is appropriate. The estimates of the 10 lag coefficients associated with the specification X(9,3) are labeled X(0), X(1), X(2), X(3), X(4), X(5), X(6), X(7), X(8), and X(9).

Autoregressive Error Model Estimation

The PDLREG procedure uses the same autoregressive error model estimation methods as the AUTOREG procedure. These two procedures share the same computational resources for computing estimates. For more information about estimation methods for autoregressive error models, see Chapter 8, "The AUTOREG Procedure."

OUT= Data Set

The OUT= data set produced by the PDLREG procedure's OUTPUT statement is similar in form to the OUT= data set produced by the AUTOREG procedure. For more information about the OUT= data set, see Chapter 8, "The AUTOREG Procedure."

Printed Output

The PDLREG procedure prints the following items:

1. the name of the dependent variable
2. the ordinary least squares (OLS) estimates
3. the estimates of autocorrelations and of the autocovariance, and if line size permits, a graph of the autocorrelation at each lag. The autocorrelation for lag 0 is 1. These items are printed if you specify the NLAG= option.
4. the partial autocorrelations if the PARTIAL and NLAG= options are specified. The first partial autocorrelation is the autocorrelation for lag 1.
5. the preliminary mean square error, which results from solving the Yule-Walker equations if you specify the NLAG= option
6. the estimates of the autoregressive parameters, their standard errors, and the ratios of estimates to standard errors (t) if you specify the NLAG= option
7. the statistics of fit for the final model if you specify the NLAG= option. These include the error sum of squares (SSE), the degrees of freedom for error (DFE), the mean square error (MSE), the root mean square error (Root MSE), the mean absolute error (MAE), the mean absolute percentage error (MAPE), the Schwarz information criterion (SBC), Akaike's information criterion (AIC), Akaike's information criterion corrected (AICC), the regression R^2 (Regress R-Square), the total R^2 (Total R-Square), and the Durbin-Watson statistic (Durbin-Watson). For more information about the regression R^2 and the total R^2, see Chapter 8, "The AUTOREG Procedure."
8. the parameter estimates for the structural model (B), a standard error estimate, the ratio of estimate to standard error (t), and an approximation to the significance probability for the parameter being 0 ("Approx Pr > $|t|$")
9. a plot of the lag distribution (estimate of lag distribution)
10. the covariance matrix of the parameter estimates if the COVB option is specified

ODS Graphics

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, "Statistical Graphics Using ODS" (*SAS/STAT User's Guide*).

Before you create graphs, ODS Graphics must be enabled (for example, with the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section "Enabling and Disabling ODS Graphics" in that chapter.

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section "A Primer on ODS Statistical Graphics" in that chapter.

PROC PDLREG assigns a name to each table it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in Table 27.2.

Table 27.2 ODS Tables Produced in PROC PDLREG

ODS Table Name	Description	Option
ODS Tables Created by the MODEL Statement		
ARParameterEstimates	Estimates of autoregressive parameters	NLAG=
CholeskyFactor	Cholesky root of gamma	NLAG= and ALL
Coefficients	Coefficients for first NLAG observations	NLAG= and (COEF or ALL)
ConvergenceStatus	Convergence status table	Default
CorrB	Correlation of parameter estimates	CORRB
CorrGraph	Estimates of autocorrelations	NLAG=
CovB	Covariance of parameter estimates	COVB
DependenceEquations	Linear dependence equation	
Dependent	Dependent variable	Default
DWTest	Durbin-Watson statistics	DW=
DWTestProb	Durbin-Watson statistics and p-values	DW= DWPROB
ExpAutocorr	Expected autocorrelations	{NLAG= and (COEF or ALL)} or {NLAG=($l_1 \dots l_m$) where $l_m > m$}
FitSummary	Summary of regression	Default
GammaInverse	Gamma inverse	NLAG= and (GINV or ALL)
IterHistory	Iteration history	ITPRINT
LagDist	Lag distribution	Default
ParameterEstimates	Parameter estimates	Default
ParameterEstimatesGivenAR	Parameter estimates assuming AR parameters are given	NLAG=
PartialAutoCorr	Partial autocorrelation	PARTIAL

Table 27.2 *continued*

ODS Table Name	Description	Option
PreMSE	Preliminary MSE	NLAG=
XPXIMatrix	$(\mathbf{X}'\mathbf{X})^{-1}$ matrix	XPX
XPXMatrix	$\mathbf{X}'\mathbf{X}$ matrix	XPX
YWIterSSE	Yule-Walker iteration sum of squared error	METHOD=ITYW
ODS Tables Created by the RESTRICT Statement		
Restrict	Restriction table	Default

Examples: PDLREG Procedure

Example 27.1: Industrial Conference Board Data

In this example, a second-degree Almon polynomial lag model is fit to a model with a five-period lag, and dummy variables are used for quarter effects. The PDL model is estimated using capital appropriations data series for the period 1952 to 1967. The estimation model is written

$$CE_t = a_0 + b_1 Q1_t + b_2 Q2_t + b_3 Q3_t + c_0 CA_t + c_1 CA_{t-1} + \cdots + c_5 CA_{t-5}$$

where *CE* represents capital expenditures and *CA* represents capital appropriations.

```
title 'National Industrial Conference Board Data';
title2 'Quarterly Series - 1952Q1 to 1967Q4';

data a;
   input ce ca @@;
   qtr = mod( _n_-1, 4 ) + 1;
   q1  = qtr=1;
   q2  = qtr=2;
   q3  = qtr=3;
datalines;
   2072 1660 2077 1926 2078 2181 2043 1897 2062 1695

   ... more lines ...


proc pdlreg data=a;
   model ce = q1 q2 q3 ca(5,2) / dwprob;
run;
```

The printed output produced by the PDLREG procedure is shown in Output 27.1.1. The small Durbin-Watson test indicates autoregressive errors.

Output 27.1.1 Printed Output Produced by PROC PDLREG

National Industrial Conference Board Data
Quarterly Series - 1952Q1 to 1967Q4

The PDLREG Procedure

Dependent Variable	ce

Ordinary Least Squares Estimates

SSE	1205186.4	**DFE**	48
MSE	25108	**Root MSE**	158.45520
SBC	733.84921	**AIC**	719.797878
MAE	107.777378	**AICC**	722.180856
MAPE	3.71653891	**HQC**	725.231641
Durbin-Watson	0.6157	**Total R-Square**	0.9834

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	210.0109	73.2524	2.87	0.0061
q1	1	-10.5515	61.0634	-0.17	0.8635
q2	1	-20.9887	59.9386	-0.35	0.7277
q3	1	-30.4337	59.9004	-0.51	0.6137
ca0**	1	0.3760	0.007318	51.38	<.0001
ca1**	1	0.1297	0.0251	5.16	<.0001
ca2**	1	0.0247	0.0593	0.42	0.6794

Estimate of Lag Distribution

Variable	Estimate	Standard Error	t Value	Approx Pr > |t|	0 0.2444
ca(0)	0.089467	0.0360	2.49	0.0165	***************
ca(1)	0.104317	0.0109	9.56	<.0001	*****************
ca(2)	0.127237	0.0255	5.00	<.0001	*********************
ca(3)	0.158230	0.0254	6.24	<.0001	***************************
ca(4)	0.197294	0.0112	17.69	<.0001	*********************************
ca(5)	0.244429	0.0370	6.60	<.0001	***

The following statements use the REG procedure to fit the same polynomial distributed lag model. A DATA step computes lagged values of the regressor X, and RESTRICT statements are used to impose the polynomial lag distribution. For the restricted least squares estimation of the Almon distributed lag model, see Judge et al. (1985, pp. 357–359).

```
data b;
   set a;
   ca_1 = lag( ca );
   ca_2 = lag2( ca );
   ca_3 = lag3( ca );
   ca_4 = lag4( ca );
   ca_5 = lag5( ca );
run;
```

```
proc reg data=b;
   model  ce = q1 q2 q3 ca ca_1 ca_2 ca_3 ca_4 ca_5;
   restrict   - ca + 5*ca_1 - 10*ca_2 + 10*ca_3 - 5*ca_4 +   ca_5;
   restrict     ca - 3*ca_1 +  2*ca_2 +  2*ca_3 - 3*ca_4 +   ca_5;
   restrict  -5*ca + 7*ca_1 +  4*ca_2 -  4*ca_3 - 7*ca_4 + 5*ca_5;
run;
```

The REG procedure output is shown in Output 27.1.2.

Output 27.1.2 Printed Output Produced by PROC REG

National Industrial Conference Board Data
Quarterly Series - 1952Q1 to 1967Q4

The REG Procedure
Model: MODEL1
Dependent Variable: ce

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	71343377	11890563	473.58	<.0001
Error	48	1205186	25108		
Corrected Total	54	72548564			

Root MSE	158.45520	**R-Square**	0.9834
Dependent Mean	3185.69091	**Adj R-Sq**	0.9813
Coeff Var	4.97397		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	210.01094	73.25236	2.87	0.0061
q1	1	-10.55151	61.06341	-0.17	0.8635
q2	1	-20.98869	59.93860	-0.35	0.7277
q3	1	-30.43374	59.90045	-0.51	0.6137
ca	1	0.08947	0.03599	2.49	0.0165
ca_1	1	0.10432	0.01091	9.56	<.0001
ca_2	1	0.12724	0.02547	5.00	<.0001
ca_3	1	0.15823	0.02537	6.24	<.0001
ca_4	1	0.19729	0.01115	17.69	<.0001
ca_5	1	0.24443	0.03704	6.60	<.0001
RESTRICT	-1	623.63242	12697	0.05	0.9614*
RESTRICT	-1	18933	44803	0.42	0.6772*
RESTRICT	-1	10303	18422	0.56	0.5814*

*** Probability computed using beta distribution.**

Example 27.2: Money Demand Model

This example estimates the demand for money by using the dynamic specification

$$m_t = a_0 + b_0 m_{t-1} + \sum_{i=0}^{5} c_i y_{t-i} + \sum_{i=0}^{2} d_i r_{t-i} + \sum_{i=0}^{3} f_i p_{t-i} + u_t$$

where

m_t = log of real money stock (M1)

y_t = log of real GNP

r_t = interest rate (commercial paper rate)

p_t = inflation rate

c_i, d_i, and f_i ($i > 0$) are coefficients for the lagged variables

The following DATA step reads the data and transforms the real money and real GNP variables using the natural logarithm. For a description of the data, see Balke and Gordon (1986).

```
data a;
   input m1 gnp gdf r @@;
   m    = log( 100 * m1 / gdf );
   lagm = lag( m );
   y    = log( gnp );
   p    = log( gdf / lag( gdf ) );
   date = intnx( 'qtr', '1jan1968'd, _n_-1 );
   format date yyqc6.;
   label m    = 'Real Money Stock (M1)'
         lagm = 'Lagged Real Money Stock'
         y    = 'Real GNP'
         r    = 'Commercial Paper Rate'
         p    = 'Inflation Rate';
datalines;
187.15  1036.22   81.18    5.58

   ... more lines ...
```

Output 27.2.1 shows a partial list of the data set.

Output 27.2.1 Partial List of the Data Set A

National Industrial Conference Board Data
Quarterly Series - 1952Q1 to 1967Q4

Obs	date	m	lagm	y	r	p
1	1968:1	5.44041	.	6.94333	5.58	.
2	1968:2	5.44732	5.44041	6.96226	6.08	0.011513
3	1968:3	5.45815	5.44732	6.97422	5.96	0.008246
4	1968:4	5.46492	5.45815	6.97661	5.96	0.014865
5	1969:1	5.46980	5.46492	6.98855	6.66	0.011005

The regression model is written for the PDLREG procedure with a MODEL statement. The LAGDEP= option is specified to test for the serial correlation in disturbances since regressors contain the lagged dependent variable LAGM.

```
title 'Money Demand Estimation using Distributed Lag Model';
title2 'Quarterly Data - 1968Q2 to 1983Q4';

proc pdlreg data=a;
   model m = lagm y(5,3) r(2, , ,first) p(3,2) / lagdep=lagm;
run;
```

The estimated model is shown in Output 27.2.2 and Output 27.2.3.

Output 27.2.2 Parameter Estimates

Money Demand Estimation using Distributed Lag Model
Quarterly Data - 1968Q2 to 1983Q4

The PDLREG Procedure

Dependent Variable	m
	Real Money Stock (M1)

Ordinary Least Squares Estimates			
SSE	0.00169815	**DFE**	48
MSE	0.0000354	**Root MSE**	0.00595
SBC	-404.60169	**AIC**	-427.4546
MAE	0.00383648	**AICC**	-421.83758
MAPE	0.07051345	**HQC**	-418.53375
		Total R-Square	0.9712

Output 27.2.2 *continued*

Parameter Estimates					
Variable	**DF**	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**
Intercept	1	-0.1407	0.2625	-0.54	0.5943
lagm	1	0.9875	0.0425	23.21	<.0001
y0**	1	0.0132	0.004531	2.91	0.0055
y1**	1	-0.0704	0.0528	-1.33	0.1891
y2**	1	0.1261	0.0786	1.60	0.1154
y3**	1	-0.4089	0.1265	-3.23	0.0022
r0**	1	-0.000186	0.000336	-0.55	0.5816
r1**	1	0.002200	0.000774	2.84	0.0065
r2**	1	0.000788	0.000249	3.16	0.0027
p0**	1	-0.6602	0.1132	-5.83	<.0001
p1**	1	0.4036	0.2321	1.74	0.0885
p2**	1	-1.0064	0.2288	-4.40	<.0001

Restriction	DF	L Value	Standard Error	t Value	Approx Pr > |t|
r(-1)	-1	0.0164	0.007275	2.26	0.0223

Output 27.2.3 Estimates for Lagged Variables

<table>
<tr><th colspan="6">Estimate of Lag Distribution</th></tr>
<tr><th>Variable</th><th>Estimate</th><th>Standard Error</th><th>t Value</th><th>Approx Pr > |t|</th><th><code>-0.196   0   0.2686</code></th></tr>
<tr><td>y(0)</td><td>0.268619</td><td>0.0910</td><td>2.95</td><td>0.0049</td><td><code>| |************************|</code></td></tr>
<tr><td>y(1)</td><td>-0.196484</td><td>0.0612</td><td>-3.21</td><td>0.0024</td><td><code>|****************| |</code></td></tr>
<tr><td>y(2)</td><td>-0.163148</td><td>0.0537</td><td>-3.04</td><td>0.0038</td><td><code>| *************| |</code></td></tr>
<tr><td>y(3)</td><td>0.063850</td><td>0.0451</td><td>1.42</td><td>0.1632</td><td><code>| |****** |</code></td></tr>
<tr><td>y(4)</td><td>0.179733</td><td>0.0588</td><td>3.06</td><td>0.0036</td><td><code>| |**************** |</code></td></tr>
<tr><td>y(5)</td><td>-0.120276</td><td>0.0679</td><td>-1.77</td><td>0.0827</td><td><code>| *********| |</code></td></tr>
</table>

<table>
<tr><th colspan="6">Estimate of Lag Distribution</th></tr>
<tr><th>Variable</th><th>Estimate</th><th>Standard Error</th><th>t Value</th><th>Approx Pr > |t|</th><th><code>-0.001   0   0.0018</code></th></tr>
<tr><td>r(0)</td><td>-0.001341</td><td>0.000388</td><td>-3.45</td><td>0.0012</td><td><code>|*****************| |</code></td></tr>
<tr><td>r(1)</td><td>-0.000751</td><td>0.000234</td><td>-3.22</td><td>0.0023</td><td><code>| *********| |</code></td></tr>
<tr><td>r(2)</td><td>0.001770</td><td>0.000754</td><td>2.35</td><td>0.0230</td><td><code>| |***********************|</code></td></tr>
</table>

<table>
<tr><th colspan="6">Estimate of Lag Distribution</th></tr>
<tr><th>Variable</th><th>Estimate</th><th>Standard Error</th><th>t Value</th><th>Approx Pr > |t|</th><th><code>-1.104   0   0.2634</code></th></tr>
<tr><td>p(0)</td><td>-1.104051</td><td>0.2027</td><td>-5.45</td><td><.0001</td><td><code>|********************************| |</code></td></tr>
<tr><td>p(1)</td><td>0.082892</td><td>0.1257</td><td>0.66</td><td>0.5128</td><td><code>| |*** |</code></td></tr>
<tr><td>p(2)</td><td>0.263391</td><td>0.1381</td><td>1.91</td><td>0.0624</td><td><code>| |********|</code></td></tr>
<tr><td>p(3)</td><td>-0.562556</td><td>0.2076</td><td>-2.71</td><td>0.0093</td><td><code>| ****************| |</code></td></tr>
</table>

References

Balke, N. S., and Gordon, R. J. (1986). "Historical Data." In *The American Business Cycle*, edited by R. J. Gordon, 781–850. Chicago: University of Chicago Press.

Emerson, P. L. (1968). "Numerical Construction of Orthogonal Polynomials from a General Recurrence Formula." *Biometrics* 24:695–701.

Gallant, A. R., and Goebel, J. J. (1976). "Nonlinear Regression with Autoregressive Errors." *Journal of the American Statistical Association* 71:961–967.

Harvey, A. C. (1981). *The Econometric Analysis of Time Series*. New York: John Wiley & Sons.

Johnston, J. (1972). *Econometric Methods*. 2nd ed. New York: McGraw-Hill.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., and Lee, T.-C. (1985). *The Theory and Practice of Econometrics*. 2nd ed. New York: John Wiley & Sons.

Park, R. E., and Mitchell, B. M. (1980). "Estimating the Autocorrelated Error Model with Trended Data." *Journal of Econometrics* 13:185–201.

Pringle, R. M., and Rayner, A. A. (1971). *Generalized Inverse Matrices with Applications to Statistics*. New York: Hafner Publishing.

Subject Index

Syntax Index