Coding Almon Polynomial Lags in Econometrics Programs

Brad R. Humphreys

June 19, 2002

Review of Theory

Almon or polynomial distributed lags are used to reduce the effects of collinearity in distributed lag settings. They impose some particular shape on lag coefficients. Suppose that you are interested in estimating the effect of lags of a variable x on y_t and you select a second-order polynomial to represent the lag weights. In this case, the effect of a change in x_{t-i} on $E(y_t)$ represented by a second order polynomial is

$$\frac{\partial E(y_t)}{\partial x_{t-i}} = \beta_i = \gamma_0 + \gamma_1 i + \gamma_2 i^2 \quad 1 = 0, 1, \ldots, n \tag{1}$$

Suppose that you believe that the effect of changes in x will be distributed over 4 periods (a distributed lag model with $n = 4$ periods). The finite lag model for this is

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \beta_4 x_{t-4} + e_t \tag{2}$$

In this case, the relationship between the polynomial lag terms and the distributed lag coefficients can be found by substituting $i = 1, 2, 3, 4$ into equation (1) to get

$$\begin{array}{ll}
\beta_0 = \gamma_0 & i = 0 \\
\beta_1 = \gamma_0 + \gamma_1 + \gamma_2 & i = 1 \\
\beta_2 = \gamma_0 + 2\gamma_1 + 4\gamma_2 & i = 2 \\
\beta_3 = \gamma_0 + 3\gamma_1 + 9\gamma_2 & i = 3 \\
\beta_4 = \gamma_0 + 4\gamma_1 + 16\gamma_2 & i = 4
\end{array}$$

Now substitute the γ_i into equation (2), to get an estimable equation

$$
\begin{aligned}
y_t = \alpha + \gamma_0 x_t + (\gamma_0 + \gamma_1 + \gamma_2) x_{t-1} + (\gamma_0 + 2\gamma_1 + 4\gamma_2) x_{t-2} \\
+ (\gamma_0 + 3\gamma_1 + 9\gamma_2) x_{t-3} + (\gamma_0 + 4\gamma_1 + 16\gamma_2) x_{t-4} + e_t
\end{aligned} \tag{3}
$$

or alternatively, solve this for the γ_is

$$
\begin{aligned}
y_t = \alpha + \gamma_0(x_t + x_{t-1} + x_{t-2} + x_{t-3} + x_{t-4}) \\
+ \gamma_1(x_{t-1} + 2x_{t-2} + 3x_{t-3} + 4x_{t-4}) \\
+ \gamma_2(x_{t-1} + 4x_{t-2} + 9x_{t-3} + 16x_{t-4}) + e_t
\end{aligned} \tag{4}
$$

This equation can be estimated and the estimated parameters from this equation (the $\hat{\gamma}_i$s) can be used to calculate the parameters from the distributed lag model (the $\hat{\beta}_i$s) from

$$
\hat{\beta}_i = \hat{\gamma}_0 + \hat{\gamma}_1 i + \hat{\gamma}_2 i^2 \quad 1 = 0, 1, 2, 3, 4. \tag{5}
$$

An Example in STATA

Suppose you wanted to estimate the above polynomial distributed lag model. The data file `table15_1.dat` has 88 observations on capital investment (y) and output (x) for US manufacturing firms. The variables on the file are time (t), investment (y) and output (x). See the accompanying do file for details.

Equation (4) can be estimated using OLS. After reading the data, use `tsset` to tell STATA that t is the time variable.

```
tsset t
```

Next, create three new variables for the right hand side of equation (4), using the STATA lag function. for a variable `z` in a data set that has a time variable defined, you can create the lag of this variable using `L.z`, a second lag using `L2.z`, and so on. The three variables on the right hand side of (4) can be created by

```
gen z1 = x + L.x + L2.x + L3.x + L4.x
gen z2 = L.x + 2*L2.x + 3*L3.x + 4*L4.x
gen z3 = L.x + 4*L2.x + 9*L3.x + 16*L4.x
```

and the γ_is estimated by

```
reg y z1 z2 z3
```

which produces

```
. reg y z1 z2 z3
  Source |       SS       df       MS                  Number of obs =      84
---------+------------------------------               F(  3,    80) = 1937.06
   Model |   393448269     3   131149423               Prob > F      =  0.0000
Residual |  5416428.17    80  67705.3521               R-squared     =  0.9864
---------+------------------------------               Adj R-squared =  0.9859
   Total |   398864697    83  4805598.76               Root MSE      =  260.20

------------------------------------------------------------------------------
       y |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      z1 |   .0442266   .0380279      1.163   0.248      -.0314512    .1199045
      z2 |  -.0531724   .0717424     -0.741   0.461      -.1959444    .0895995
      z3 |   .0399124   .0183027      2.181   0.032       .0034888    .0763359
   _cons |   217.4883   64.48597      3.373   0.001       89.15712    345.8195
------------------------------------------------------------------------------
```

To calculate estimates of the β_is, just plug these into equation (5), calculated in STATA by

```
matrix myb = e(b)
matrix list myb

scalar b0 = myb[1,1]
scalar b1 = myb[1,1]+ myb[1,2] + myb[1,3]
scalar b2 = myb[1,1]+ 2*myb[1,2] + 4*myb[1,3]
scalar b3 = myb[1,1]+ 3*myb[1,2] + 9*myb[1,3]
scalar b4 = myb[1,1]+ 4*myb[1,2] + 16*myb[1,3]

scalar list
        b4 =  .47013456
        b3 =  .24392052
        b2 =  .09753118
        b1 =  .03096656
        b0 =  .04422664
```

Alternatively, suppose you wanted to estimate a second order polynomial for a distributed lag model with a lag length of 8 periods. Again the polynomial function is

$$\frac{\partial E(y_t)}{\partial x_{t-i}} = \beta_i = \gamma_0 + \gamma_1 i + \gamma_2 i^2 \quad 1 = 0, 1, \ldots, 8$$

and the finite lag model for this example is

$$y_t = \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \beta_4 x_{t-4} + \beta_5 x_{t-5}$$
$$+\beta_6 x_{t-6} + \beta_7 x_{t-7} + \beta_8 x_{t-8} + e_t.$$

The relationship between the polynomials and the distributed lag parameters is

$$\begin{array}{ll}
\beta_0 = \gamma_0 & i = 0 \\
\beta_1 = \gamma_0 + \gamma_1 + \gamma_2 & i = 1 \\
\beta_2 = \gamma_0 + 2\gamma_1 + 4\gamma_2 & i = 2 \\
\beta_3 = \gamma_0 + 3\gamma_1 + 9\gamma_2 & i = 3 \\
\beta_4 = \gamma_0 + 4\gamma_1 + 16\gamma_2 & i = 4 \\
\beta_5 = \gamma_0 + 5\gamma_1 + 25\gamma_2 & i = 5 \\
\beta_6 = \gamma_0 + 6\gamma_1 + 36\gamma_2 & i = 6 \\
\beta_7 = \gamma_0 + 7\gamma_1 + 49\gamma_2 & i = 7 \\
\beta_8 = \gamma_0 + 8\gamma_1 + 64\gamma_2 & i = 8
\end{array}$$

And substituting into the distributed lag equation gives

$$y_t = \alpha + \gamma_0 x_t + (\gamma_0 + \gamma_1 + \gamma_2)x_{t-1} + (\gamma_0 + 2\gamma_1 + 4\gamma_2)x_{t-2}$$
$$+(\gamma_0 + 3\gamma_1 + 9\gamma_2)x_{t-3} + (\gamma_0 + 4\gamma_1 + 16\gamma_2)x_{t-4}$$
$$+(\gamma_0 + 5\gamma_1 + 25\gamma_2)x_{t-5} + (\gamma_0 + 6\gamma_1 + 36\gamma_2)x_{t-6}$$
$$+(\gamma_0 + 7\gamma_1 + 49\gamma_2)x_{t-7} + (\gamma_0 + 8\gamma_1 + 64\gamma_2)x_{t-8} + e_t$$

This equation could be solved for the γ_is to get an estimable equation and then the β_is could be calculated from these estimates.