POLS571 - Longitudinal Data Analysis
October 2, 2001

Distributed Lag Models

1 Basics

One of the most common motivations for using time-series methods is in order to account for lagged effects of covariates. Lots of theories predict that "X will affect Y at a lag" (though they usually don't tell us at *what* lag).

Consider a model of the form:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + ... + \beta_k X_{t-k} + u_t \tag{1}$$

this is what is known as a *distributed lag model*: the effects of X on Y occur both contemporaneously and at k lags. As $k \rightarrow \infty$, we have:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + ... + u_t \tag{2}$$

which is an infinite distributed lag model (and which isn't identified for a single time series...).

One can estimate a basic distributed lag model using plain-vanilla OLS. Often, one will estimate several models to determine the appropriate lag length k to use. That is, estimate:

$$\begin{aligned} Y_t &= \alpha + \beta_0 X_t + u_t \\ Y_t &= \alpha + \beta_0 X_t + \beta_1 X_{t-1} + u_t \\ Y_t &= \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t \end{aligned}$$

etc., until some estimate of $\hat{\beta}_{k+1}$ isn't significantly different from zero.

Advantages

- The model is very flexible; lagged effects can be positive or negative, and change discretely at different lags (e.g., positive at $t-1$, negative at $t-2$, etc.).

- Each β_k is estimated holding $\beta_0, \beta_1, ...$ constant; this allows us to assess period effects in the relationship between X and Y.

Disadvantages:

- The process is very ad hoc; we're letting the data, rather than theory, determine the right lag length.
- If the Xs are autocorrelated, multicollinearity can be a serious problem.
  - Can yield badly inflated standard error estimates.
  - May wrongly fail to reject the null $\hat{\beta} = 0$.

2 The Kocyk Distributed Lag Model

The Kocyk model is:

$$Y_t = \alpha + \beta X_t + \beta\lambda X_{t-1} + \beta\lambda^2 X_{t-2} + ... + \beta\lambda^k X_{t-k} + e_t \tag{3}$$

with the restriction that $0 < \lambda < 1$. This model thus assumes that the influence of X on Y:

- Remains positive or negative over time, and
- declines *geometrically* as k increases.

Note that we still don't necessarily know what the "correct" lag length k is. However, we can rewrite (3) as follows: First, lag both sides of the equation, and multiply both sides by λ:

$$\lambda Y_{t-1} = \alpha\lambda + \beta\lambda X_{t-1} + \beta\lambda^2 X_{t-2} + ... + \lambda e_{t-1} \tag{4}$$

Next, subtract (4) from (3):

$$Y_t - \lambda Y_{t-1} = \alpha(1 - \lambda) + \beta X_t + (e_t - \lambda e_{t-1}) \tag{5}$$

And then add λY_{t-1} to both sides, to give:

$$Y_t = \alpha(1 - \lambda) + \beta X_t + \lambda Y_{t-1} + u_t \tag{6}$$

where $u_t = (e_t - \lambda e_{t-1})$ is the compound error term. This transformation:

- Converts the Kocyk distributed lag model into a more standard AR(1) model.
- This limits the number of parameters to three (α, β, λ), BUT
- Still allows the researcher to get estimates of all the parameters in (3).

This seems like a really nice model. However, it has several minor and one major problem(s):

- We have an endogenous / stochastic independent variable (Y_{t-1}),
- There is serial correlation in the u_ts (how much depends on the value of λ), and
- Worst of all, Y_{t-1} is correlated with the u_ts...

This last one is a big problem. See that:

$$
\begin{aligned}
Cov(Y_{t-1}, u_t) &= Cov(Y_{t-1}, (e_t - \lambda e_{t-1}) \\
&= -\lambda[Var(e_t)] \\
&= -\lambda\sigma^2
\end{aligned}
$$

This means that if we use OLS to estimate the model in (3), we'll get biased and inconsistent results...

This model is usually estimated using instrumental variables. Specifically, the researcher creates an "instrument" for Y_{t-1} which is

- Highly correlated with Y_{t-1}, but
- uncorrelated with u_t.

One option might be, for example, to substitute X_{t-1} for Y_{t-1} in (6).

In political science, we almost always have lousy instruments for everything; this is why you see very little use of the Kocyk model...

3 The Almon Distributed Lag Model

3.1 The Math

The Almon model is more flexible than the Kocyk model, in that it allows the effect of X on Y over time to change (e.g., increase, decrease, etc.). We can write a general finite distributed lag model as:

$$Y_t = \alpha + \sum_{i=0}^{k} \beta_i X_{t-i} + u_t \tag{7}$$

where i indexes the k lag lengths. Now, we might write:

$$\beta_i = a_0 + a_1 i + a_2 i^2 + ... + a_m i^m \tag{8}$$

so that β_i is now an mth-order polynomial. Note that we need to restrict the model so that $m < k$. Plug (8) into (7):

$$\begin{aligned} Y_t &= \alpha + \sum_{i=0}^{k}(a_0 + a_1 i + a_2 i^2 + ... + a_m i^m) X_{t-i} + u_t \\ &= \alpha + a_0 \sum_{i=0}^{k} X_{t-i} + a_1 \sum_{i=0}^{k} i X_{t-i} + ... + a_m \sum_{i=0}^{k} i^m X_{t-i} + u_t \end{aligned} \tag{9}$$

Next, write:

$$\begin{aligned} Z_{0t} &= \sum_{i=0}^{k} X_{t-i} \\ Z_{1t} &= \sum_{i=0}^{k} i X_{t-i} \\ Z_{2t} &= \sum_{i=0}^{k} i^2 X_{t-i} \\ &\dots \\ Z_{mt} &= \sum_{i=0}^{k} i^m X_{t-i} \end{aligned}$$

and we can rewrite (9) as:

$$Y_t = \alpha + a_0 Z_{0t} + a_1 Z_{1t} + a_2 Z_{2t} + ... + a_m Z_{mt} + u_t \tag{10}$$

which we can estimate via OLS.

3.2 How do you do it?

The answer seems pretty simple:

1. Create Z_{0t}, Z_{1t}, etc...
2. Regress Y_t on Z_{0t}, Z_{1t}, etc... using OLS.

The hardest part is creating the Zs. At each time point, and for each covariate X, we need to calculate $m+1$ different, new Z variables. Here's an example of what they would look like for $k = 3$ and $m = 2$:

t	X_t	Z_{0t}	Z_{1t}		Z_{2t}	
0	10	-	-		-	
1	15	-	-		-	
2	20	-	-		-	
3	15	60	100	(0+15+40+45)	230	(0+15+80+135)
4	10	60	80	(0+20+30+30)	170	(0+20+60+90)
5	15	60	80	(0+15+20+45)	190	(0+15+40+135)
6	25	65	100	(0+10+30+60)	295	(0+10+60+225)
7	10	60	95	(0+15+50+30)	205	(0+15+100+90)
8	15	65	90	(0+25+20+45)	200	(0+25+40+135)

Unfortunately, there's no canned Stata^{TM} routine for estimating these models, so you have to do them "by hand".

Once we've estimated (10), we can get back estimates of the $\hat{\beta}$s easily enough:

$$\hat{\beta}_k = \hat{a}_0 + k\hat{a}_1 + k^2\hat{a}_2 + ... + k^m\hat{a}_m \tag{11}$$

This means that:

$$
\begin{aligned}
\hat{\beta}_0 &= \hat{a}_0 \\
\hat{\beta}_1 &= \hat{a}_0 + \hat{a}_1 + \hat{a}_2 + ... + \hat{a}_m \\
\hat{\beta}_2 &= \hat{a}_0 + 2\hat{a}_1 + 4\hat{a}_2 + ... + 2^m \hat{a}_m \\
\hat{\beta}_3 &= \hat{a}_0 + 3\hat{a}_1 + 9\hat{a}_2 + ... + 3^m \hat{a}_m \\
& etc.
\end{aligned}
$$

Advantages of the Almon model:

- Very flexible with respect to how lags of X influence Y.
- Estimable using OLS (because there's no autoregressive component).

Potential disadvantages:

- Must specify both k (lag length) and m (polynomial degree), and
- the model can be sensitive to the specification of both...

One strategy is to add degrees of polynomials successively, each time testing for the significance of $\hat{a}_m$; a nonsignificant coefficient indicates the correct polynomial degree,

HOWEVER

bear in mind that the Z_ts are, by construction, multicollinear; take this into account when conducting significance tests on the $\hat{a}_m$s.

For a political science example of Almon lags, see the Box-Steffensmeier and Lin (1996) article...