Using Stata to Replicate Table 4 in Bond (2002)

These notes refer to using Stata/SE 9.1, in March 2006.

Preliminaries

Open the dataset usbal89.

The main variables are:

id - firm identifier
year - year
y - log sales
n - log employment
k - log capital stock

Other variables have been derived from these. E.g.

y_1 - first lag of y
yk - (y - k), log of sales-capital ratio

Set panel format

```
tsset id year , yearly
```

Pooled OLS (OLS levels); Table 4, column (i)

```
xi: regress y n l.n k l.k l.y i.year , robust cluster(id)
```

Within Groups; Table 4, column (ii)

```
xi: xtreg y n l.n k l.k l.y i.year , fe robust cluster(id)
```

or

```
xi: areg y n n_1 k k_1 y_1 i.year , absorb(id) robust cluster(id)
```

Here the coefficients are identical to those in Table 4, column (i), since this is a balanced panel.

First-Differenced GMM; Table 4, column (iii)

xi: xtabond2 y n l.n k l.k l.y i.year , gmm(y n k, lag(2 .)) iv(i.year) robust noleveleq

First-Differenced GMM; Table 4, column (iv)

xi: xtabond2 y n l.n k l.k l.y i.year , gmm(y n k, lag(3 .)) iv(i.year) robust noleveleq

This omits the levels of the variables dated t-2 from the set of instruments. The serial correlation tests reported in Table 4, column(iv) are slightly different from those produced by Stata.

System GMM; Table 4, column (v)

xi: xtabond2 y n l.n k l.k l.y i.year , gmm(y n k, lag(2 .)) iv(i.year, equation(level)) robust h(1)

The noleveleq option is not specified. This uses a "system" combining equations in first-differences with equations in levels.

The h(1) option uses 2SLS as the one-step estimator. This was also the case in the Blundell-Bond (2000) production function estimates, which are reproduced in Table 4. This is *not* the one-step weight matrix used in DPD98 for Gauss, or the DPD package in PC Give and OX. The h(2) option in Stata uses the same one-step weight matrix as these programs. Neither of these is the default option in Stata, which corresponds to the h(3) option. See help xtabond2 for further details.

The gmm(y n k, lag(2 .)) option uses the lagged levels of y, n and k dated t-2 and earlier as instruments for the equations in first-differences; and (correspondingly) the lagged first-differences of y, n and k dated t-1 (only) as instruments for the equations in levels. This is the default specification of gmm-style instruments for the levels equations. Other options are available; see help xtabond2 for further details.

The iv(i.year, equation(level)) option uses the year dummies as instruments for the equations in levels *only*. This treatment is specific to year dummies, and ensures that the correct number of moment conditions of the form $E[u_{it} - c_t] = 0$ are used. For other strictly exogenous variables used as iv-style instruments, the equation(level) restriction would not normally be used.

The results are similar but not identical to those in Table 4, column (v).

System GMM; Table 4, column (vi)

xi: xtabond2 y n l.n k l.k l.y i.year , gmm(y n k, lag(3 .)) iv(i.year, equation(level)) robust h(1)

The gmm(y n k, lag(3 .)) option uses the lagged levels of y, n and k dated t-3 and earlier as instruments for the equations in first-differences; and (correspondingly) the lagged first-differences of y, n and k dated t-2 (only) as instruments for the equations in levels.

The results are similar but not identical to those in Table 4, column (vi).

To obtain the Difference-Sargan test in column (vi):

```
g dsar = 75.80 - 53.66
g df = 55 - 40
g pval = chi2tail(df, dsar)
su pval
```

This gives a p-value of 0.104.

To test the "common factor" restrictions in column (vi):

```
testnl (_b[l.y]*_b[n] = -_b[l.n]) (_b[l.y]*_b[k] = -_b[l.k])
```

This gives a p-value of 0.796.

A useful feature of xtabond2 is that different assumptions can be made about the validity of different instruments. For example, suppose we do not want to use lagged first-differences of capital as instruments for the equations in levels, but we do want to use lagged first-differences of sales and employment. This requires two separate uses of the gmm(.) option:

```
xi: xtabond2 y n l.n k l.k l.y i.year, gmm(k, lag(3 .) equation(diff))
gmm(y n, lag(3 .)) iv(i.year, equation(level)) robust h(1)
```

The gmm(k, lag(3 .) equation(diff)) option specifies the use of lagged levels of k dated t-3 and earlier as instruments for the equations in first-differences; with no lagged differences of k used as instruments for the equations in levels.

This is particularly useful when we expect first-differences of some but not all of the variables to be uncorrelated with the individual-specific effects.