Using Stata to Replicate Table 1 in Bond (2002)

These notes refer to using either Stata/SE 13.1 or Stata/SE 14.1, in September 2016. They refer to version 03.06.00 of the xtabond2 command.

Preliminaries

Open the dataset invrat1. This can be done in several ways, for example,

Click File Open from within Stata

Use command (use filename) in the command window, e.g.

```
u invrat1
```

Double click the file in My Computer

Six variables are listed in the variables window

- year - calendar year (integer)
- date - exact date (end of accounting period; string)
- dscode - firm identifier (string)
- firm - firm identifier (integer)
- invrate - investment rate (real)
- noj - number of observations on this firm (integer)

Some useful commands to learn more about the variables

- describe [varlist] - lists variables and types
- summarize [varlist] - basic descriptive statistics
- summarize [varlist] , detail - more descriptive statistics
- list [varlist] - displays values
- list [varlist] if - displays specified values

Examples:

```
su
su invrate , d
l firm year invrate if firm == 9008282
```

Stata's help facility gives much more information on all these commands. For example,

Click Help Stata Command... and then type the name of the command

Click Help Search... and then the topic

Type help followed by the command name in the command window

To use Stata's commands for panel data, we need to tell Stata that this is a panel dataset.

In the command window, type

xtset firm year , yearly

This tells Stata that the variable firm identifies different cross-section units (like the subscript i), the variable year identifies different time periods (like the subscript t), and the observations are yearly. Firm and year should be integers. This will also sort the dataset, if it is not already sorted.

For some commands, we may need to create and store the first lag of the investment rate explicitly.

In the command window, type

generate invrate1 = l.invrate

This uses Stata's lag operator (l.) to create the first lag of the variable invrate. This requires that we have first used xtset to specify the panel nature of the dataset. The first observation on each firm is set to Stata's missing value code. The new variable invrate1 should appear in the variable list.

Other operators are also available. For example,

ll. or l2. create the second lag

d. creates the first-difference

ld. (or dl.) creates the first lag of the first-difference

Pooled OLS (OLS levels); Table 1, column (i)

In the command window, type

xi: regress invrate l.invrate i.year , vce(cluster firm)

Results for the pooled OLS levels regression should appear in the results window. Identical results are obtained with the abbreviated command

xi: reg invrate l.invrate i.year , vce(cl firm)

Suppressing the constant term changes the way in which the coefficients on the year-specific intercept terms is presented, but should have no effect on the coefficient on the lagged investment rate.

```
xi: reg invrate invrate1 i.year , vce(cl firm) noconstant
```

The prefix xi: allows the i.year shorthand to be used to include a set of year dummies. This prefix is no longer necessary, but take care if you omit the prefix and suppress the constant term.

The vce(cluster firm) option for variance-covariance estimation (vce) specifies standard errors that are asymptotically robust to both heteroskedasticity and serial correlation. Using instead the vce(robust) option would produce standard errors that are asymptotically robust to heteroskedasticity, but not to serial correlation (not usually recommended for panel data applications).

The estimated coefficient on the lagged dependent variable is identical to that in Table 1, column (i). The standard error is almost identical, but differs at the 4th decimal place. This is not a typo in the paper, but reflects slight differences between the ways in which Stata's vce(cluster firm) option and DPD98 for Gauss calculate the cluster-robust standard errors.

To store results to a file, first open a log file inside Stata. For example,

In the command window, type

```
log using results.log
```

for a plain text file.

Or click File Log Begin, or click the Log Begin button on the toolbar.

See help log for more options. Typing log close in the command window stops subsequent commands and results being stored. Or click File Log Close, or click the Log Close button on the toolbar.

If the abar command is installed (see below), typing

```
abar , lags(2)
```

will produce cluster-robust versions of the Arellano-Bond (1991) tests for the absence of first-order and second-order serial correlation in the residuals (similar but not identical to those reported in Table 1, column (i)).

Within Groups; Table 1, column (ii)

In the command window, type

xi: xtreg invrate l.invrate i.year , fe vce(cluster firm)

xtreg is Stata's command for classical panel data regression estimators.

The fe option specifies "fixed effects".

This command produces results which are similar but not identical to those in Table 1, column (ii). The Within Groups option in DPD98 for Gauss uses OLS on the model transformed to orthogonal deviations. This is identical to OLS on the Within transformed model only in the case of balanced panels.

Anderson-Hsiao First-Differenced 2SLS; Table 1, column (iii)

In the command window, type

xi: ivregress 2sls d.invrate (ld.invrate=ll.invrate) i.year , vce(cluster firm)

ivregress is Stata's command for instrumental variables regression. A more powerful alternative (ivreg2) can be installed if desired.

[For historical reasons, ivregress should not be abbreviated to ivreg. ivreg is an older and less powerful command which is now redundant.]

ivregress 2sls specifies that two-stage least squares is used here (other options are available with the ivregress command).

The dependent variable is specified in first-differences.

The second lag of invrate is specified as an instrumental variable for the lagged dependent variable in first-differences, which is treated as endogenous.

The year dummies are treated as (strictly) exogenous and also included in the instrument set.

The coefficient on the lagged dependent variable and its standard error are identical to those reported in Table 1, column (iii).

Arellano-Bond First-Differenced GMM: Table 1, column (iv)

In the command window, type

xi: xtabond invrate i.year , maxldep(2) vce(robust)

xtabond is Stata's original command for first-differenced GMM estimation of panel data models. Newer commands (xtdpd and xtdpdsys) are also available. The add-on command xtabond2 provides the most powerful and flexible alternative, and will be discussed below.

The dependent variable is specified in levels. This command automatically takes first-differences.

The lagged dependent variable is not specified. This command automatically includes a lagged dependent variable, and treats this as endogenous. There are options to include additional lags, or not to include a lagged dependent variable.

The year dummies are treated as (strictly) exogenous and also included in the instrument set.

There are options to include additional predetermined or endogenous explanatory variables. See help xtabond for details.

The maxldep(2) option specifies that lagged levels of the dependent variable dated t-2 and t-3 only should be included as instruments. That is, at most 2 lagged levels of the dependent variable are used as instruments in each of the first-differenced equations.

The vce(robust) option here produces standard errors that are robust to both heteroskedasticity and serial correlation (and are computed in an identical way to those obtained with DPD98 for Gauss).

This produces coefficients, standard errors and serial correlation tests that are identical to those in Table 1, column (iv). These are one-step GMM estimates.

To obtain the serial correlation tests, after estimation type:

estat abond

To obtain the test of overidentifying restrictions, use the twostep option to specify that the optimal two-step GMM estimator should be computed.

xi: xtabond invrate i.year , maxldep(2) two

Then type:

estat sargan

This gives standard errors for the two-step estimator which are asymptotically robust to heteroskedasticity and serial correlation, but which are known to have very poor finite sample properties. The corresponding test of overidentifying restrictions may have very low power in cases where a high number of overidentifying restrictions are specified relative to the (cross-section) sample size. [A symptom of this would be a p-value close to one, in which case if you are seriously interested in testing instrument validity you should consider reducing the total number of instruments specified.]

If the xtabond2 command has been installed, the same results can be obtained using the commands

```
xi: xtabond2 invrate l.invrate i.year , gmm(invrate, lag(2 3)) iv(i.year) robust noleveleq

xi: xtabond2 invrate l.invrate i.year , gmm(invrate, lag(2 3)) iv(i.year) robust twostep noleveleq
```

Notice that xtabond2 uses Stata's older syntax (,robust) to specify the option for cluster-robust standard errors, rather than the newer syntax (,vce(cluster varname)) that we have illustrated for earlier commands, or the ,vce(robust) syntax used with xtabond.

With xtabond2 the robust option can be specified together with the twostep option. This produces standard errors for the two-step estimator that are asymptotically robust to both heteroskedasticity and serial correlation, and which use the finite-sample correction proposed by Windmeijer (2005). This is one advantage of using the xtabond2 command.

The syntax of xtabond2 is quite different from that of xtabond. The dependent variable and the list of explanatory variables are specified before the comma. The instrument set is specified after the comma.

The option gmm(invrate, lag(2 3)) includes the second and third lags of the level of invrate in the instrument set. These instruments are not "stacked", so that coefficients on these instruments in the first-stage regressions are allowed to vary over time.

The option iv(i.year) includes the year dummies in the instrument set. These instruments are "stacked", so that coefficients on these instruments in the first-stage regressions are restricted to be constant over time.

The option noleveleq specifies that only the equations in first-differences are used in estimation. Without this option, the xtabond2 command uses both equations in first-differences combined with equations in levels. The option to compute these "system GMM" estimators is another important advantage of using the xtabond2 command.

The latest version of xtabond2 uses the Mata programming language now available in Stata. To use xtabond2 with (much) earlier versions of Stata, also include the nomata option (after the comma).

If xtabond2 is not already installed, and your computer is connected to the internet, it can be installed easily from within Stata.

Click Help Search...

Check Search net resources and type xtabond2

Click on the url for the xtabond2 command

Click on the download/install option

The abar, ivreg2 and many other commands can be installed in the same way.

Note that the xtabond2 command is written by David Roodman. Please include the requested citation if you use this command in your work. Details can be found using help xtabond2 after the command is installed. Please also direct any questions on this command to david@davidroodman.com

Arellano-Bond First-Differenced GMM: Table 1, column (v)

In the command window, type

```
xi: xtabond invrate i.year , vce(robust)
estat abond
```

The only difference from the xtabond command used in the previous section is that we do not use the maxldep option to restrict the number of lagged levels of invrate that are used as instruments in each of the first-differenced equations. The default, used here, is to include all available lags.

These produce identical results to those in Table 1, column (v).

The corresponding two-step GMM estimator and test of overidentifying restrictions are obtained using

```
xi: xtabond invrate i.year , two
estat sargan
```

To obtain the same results using xtabond2, the commands are

```
xi: xtabond2 invrate l.invrate i.year , gmm(invrate, lag(2 .)) iv(i.year) r nol
xi: xtabond2 invrate l.invrate i.year , gmm(invrate, lag(2 .)) iv(i.year) two nol
```

Here the option gmm(invrate, lag(2 .)) specifies that all available lags of invrate dated t-2 and earlier should be used as instruments for the first-differenced equations.

To obtain the two-step estimates with finite-sample corrected standard errors

```
xi: xtabond2 invrate l.invrate i.year , gmm(invrate, lag(2 .)) iv(i.year) r two nol
```

This produces identical results to those discussed on p.150 of Bond (2002), or pp.14-15 of the Cemmap working paper version.

Do files

While we have introduced these Stata commands as interactive commands, for serious use it is recommended that commands are saved and executed from .do files, which allows results to be replicated easily at a later date. A simple example would be:

```
set more off
u c:\foldername\invrat1, clear
xtset firm year, yearly
log using c:\foldername\invrat1, replace
xi: reg invrate l.invrate i.year, vce(cl firm)
xi: xtreg invrate l.invrate i.year, fe vce(cl firm)
xi: ivregress 2sls d.invrate (ld.invrate=ll.invrate) i.year, vce(cl firm)
xi: xtabond2 invrate l.invrate i.year, gmm(invrate, lag(2 3)) iv(i.year) r nol
xi: xtabond2 invrate l.invrate i.year, gmm(invrate, lag(2 .)) iv(i.year) r nol
log close
```

Saved as a text file (for example, invrat1.do) and executed from within Stata, this produces the main results in all five columns of Table 1, saved in the log file invrat1.smcl.

Do files can be executed either by clicking File Do, or by using Stata's do file editor (click the Do File Editor button on Stata's toolbar). See help do for more information.