How to Do xtabond2

David Roodman
Research Fellow
Center for Global Development

Center for Global Development

xtabond2 in a nutshell

- First ado version in 11/03, Mata version in 11/05.
- Extends built-in xtabond, to do system GMM, Windmeijer correction, revamped syntax
- Estimators designed for
  - Small-T, large-N panels
  - One dependent variable
  - Dynamic
  - Linear
  - Regressors endogenous and predetermined
  - Fixed individual effects
  - Arbitrary autocorrelation and het. within panels
  - General application

Outline of paper

- Introduction to linear GMM
- Motivation and design of difference and system GMM
- `xtabond2` syntax

Black box problem

- Canned & sophisticated procedure
- Dangers in hidden sophistication
  - finite sample $\neq$ asymptotic
- Users should understand motivation and limits of estimator

Linear GMM in one slide

- Instrument vector $\mathbf{z}$ such that $\mathrm{E}[\mathbf{z}\varepsilon]=\mathbf{0}$
- # instruments > # parameters so can't have $\mathrm{E}_N[\mathbf{z}\varepsilon]=\frac{1}{N}\mathbf{Z}'\hat{\mathbf{E}}=\mathbf{0}$
- Want to "minimize" $\frac{1}{N}\mathbf{Z}'\hat{\mathbf{E}}$ in some sense
- In what sense? By a pos-semi-def. quad. form given by $\mathbf{A}$:

$$\left\|\mathrm{E}_N[\mathbf{z}\varepsilon]\right\|_{\mathbf{A}}=\left\|\frac{1}{N}\mathbf{Z}'\hat{\mathbf{E}}\right\|_{\mathbf{A}}\equiv N\left(\frac{1}{N}\mathbf{Z}'\hat{\mathbf{E}}\right)'\mathbf{A}\left(\frac{1}{N}\mathbf{Z}'\hat{\mathbf{E}}\right)=\frac{1}{N}\hat{\mathbf{E}}'\mathbf{Z}\mathbf{A}\mathbf{Z}'\hat{\mathbf{E}}$$

- Given $\mathbf{A}$, minimizing leads to $\hat{\boldsymbol{\beta}}_{\mathbf{A}}=\left(\mathbf{X}'\mathbf{Z}\mathbf{A}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\mathbf{A}\mathbf{Z}'\mathbf{Y}$
- Always unbiased, but which $\mathbf{A}$ is efficient? Answer: $\mathbf{A}$ should weight moments $\mathbf{z}_i'\mathbf{E}$ inversely with their variances and covariances:

$$\mathbf{A}_{EGMM}=\mathrm{Var}\left[\mathbf{Z}'\mathbf{E}\middle|\mathbf{X},\mathbf{Z}\right]^{-1}=\left(\mathbf{Z}'\,\mathrm{Var}\left[\mathbf{E}\middle|\mathbf{X},\mathbf{Z}\right]\mathbf{Z}\right)^{-1}=\left(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}\right)^{-1}$$

- To make feasible, choose arbitrary proxy for $\boldsymbol{\Omega}$, call it $\mathbf{H}$. Do GMM (*one-step*). Use residuals to make robust sandwich estimator of $\left(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}\right)^{-1}$. Rerun. *Two-step* is feasible, theoretically efficient.

Linear GMM and 2SLS

$$\hat{\boldsymbol{\beta}}_{EGMM} = \left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{Y}$$

- If $\boldsymbol{\Omega} = \sigma^2\mathbf{I}$, reduces to

$$\hat{\boldsymbol{\beta}}_{2SLS} = \left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{Y}$$

- If errors i.i.d., efficient GMM *is* 2SLS
- If not, 2SLS inefficient

Linear GMM in another slide (Holtz-Eakin, Newey, and Rosen 1988)

(1) $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}$

OLS inconsistent: $\mathrm{E}\left[\mathbf{X}'\mathbf{E}\right] \neq 0$

(2) Take **Z**-moments: $\mathbf{Z}'\mathbf{Y} = \mathbf{Z}'\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}'\mathbf{E}$

OLS consistent $\left(\mathrm{E}\left[\mathbf{X}'\mathbf{Z}\mathbf{Z}'\mathbf{E}\right] = 0\right)$

but inefficient $\left(\mathrm{Var}\left[\mathbf{Z}'\mathbf{E}\right] = \mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z} \text{ not scalar}\right)$

Left-multiply by $\left(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}\right)^{-1/2}$:

$$\left(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}\right)^{-1/2}\mathbf{Z}'\mathbf{Y} = \left(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}\right)^{-1/2}\mathbf{Z}'\mathbf{X}\boldsymbol{\beta} + \left(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}\right)^{-1/2}\mathbf{Z}'\mathbf{E}$$

Let $\mathbf{X}^* = \left(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}\right)^{-\frac{1}{2}}\mathbf{Z}'\mathbf{X},\ \mathbf{Y}^* = \left(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}\right)^{-\frac{1}{2}}\mathbf{Z}'\mathbf{Y},\ \mathbf{E}^* = \left(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z}\right)^{-\frac{1}{2}}\mathbf{Z}'\mathbf{E}$

(3) $\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{E}^*$

OLS efficient $\left(\mathrm{Var}\left[\mathbf{E}^*\right] = \mathbf{I}\right)$

OLS on (3) = GLS on (2) = GMM on (1)

GMM = GLS on **Z**-moments

Difference and system GMM

Basic model:

$$y_{it} = \alpha y_{i,t-1} + \mathbf{x}_{it}^{'}\boldsymbol{\beta} + \nu_{it}$$

$$\nu_{it} = \mu_i + \varepsilon_{it}$$

$$\mathrm{E}\left[\mu_i\right] = \mathrm{E}\left[\varepsilon_{it}\right] = \mathrm{E}\left[\mu_i \varepsilon_{it}\right] = 0$$

Conceptual starting point: OLS

Problem: Dynamic Panel Bias (Nickell 1981)

- Fixed effects in disturbance term make $y_{i,t-1}$ endogenous
  - Example: Indonesia
- A problem of short panels
- Individual dummies (=Within Groups) don’t help
  - Transformed $y_{i,t-1}$ endogenous, as are deeper lags

Partial solution: OLS in differences

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta \mathbf{x}'_{it} \boldsymbol{\beta} + \Delta \varepsilon_{it}$$

- Purges fixed effects, doesn't spread endogeneity much
- Transformed $y_{i,t-1}$ still becomes endogenous since the $y_{i,t-1}$ in $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$ correlates with the $\varepsilon_{i,t-1}$ in $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$
- But deeper lags exogenous if no AR(), offering instruments

Problem: Other endogeneity

- Differencing eliminates endogeneity to fixed effects error component. But
  - $\Delta y_{i,t-1}$ now endogenous to $\Delta\varepsilon_{it}$
  - Other predetermined variables become endogenous in same way
  - Still other variables may be endogenous from the start
- For general application, assume no perfect instruments waiting in the wings

Solution: Instrument with lags (2SLS) (Anderson and Hsiao 1981)

- Assuming no AR() in ε_{it}, natural instruments for $\Delta y_{i,t-1}$ are $\Delta y_{i,t-2}$ and $y_{i,t-2}$
- Both mathematically related to $\Delta y_{i,t-1}$
- $y_{i,t-2}$ seems preferable: available at $t = 3$
- Again, small T influences
- Do same for other endogenous variables

Problem: Inefficiency

- Deeper lags available as instruments
  - But reduce sample in 2SLS
  - Problem for short panels
- In differences, errors not i.i.d.
  - $\Delta\varepsilon_{it}$ and $\Delta\varepsilon_{i,t-1}$ mathematically correlated
  - 2SLS not efficient

Solution: GMM & GMM-style instruments (Holtz-Eakin, Newey, and Rosen 1988)

- Use many lags, replacing missing with zero
- Generate separate instrument for each lag and time period instrumented

$$\text{IV-style: } \begin{bmatrix} . \\ . \\ y_{i1} \\ \vdots \\ y_{i,T-2} \end{bmatrix} \quad \text{GMM-style: } \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ y_{i1} & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & y_{i2} & y_{i1} & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & y_{i3} & y_{i2} & y_{i1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

- Result: Arellano-Bond (1991) difference GMM

Problem: Autocorrelation

- E.g., if ε_{it} are AR(1), then $y_{i,t-2} \sim \varepsilon_{i,t-2} \sim \varepsilon_{i,t-1} \sim \Delta\varepsilon_{it}$
- Must assume $y_{i,t-2}$ is invalid instrument in *i*,*t*

Solution: Restrict to deeper lags

- If we find AR(l) in ε_{it}, use lags $l + 1$ and deeper

Arellano-Bond AR() test

- Expect AR() in $v_{it} = \mu_i + \varepsilon_{it}$
- To check for AR(1) in ε_{it}, test for AR(2) in Δe_{it}
- E.g., compare $e_{it} - e_{i,t-1}$ and $e_{i,t-2} - e_{i,t-3}$ to detect $e_{i,t-1} \sim e_{i,t-2}$
- Test statistic for AR(*l*) in differences: $\sum_{i,t} \Delta e_{it} \Delta e_{i,t-l}$
- Normal under null of no AR(*l*)
- Arellano and Bond calculate its standard deviation
- *z* test for AR()
- More general than other AR() tests in Stata.
- `abar`: post-estimation command for `regress`, `ivreg`, `ivreg2`

Problem: Weak instruments

If y is nearly a random walk, $y_{i,t-1}$ is a poor instrument for Δy_{it}, mathematical relationship notwithstanding

Solution: Instead of purging fixed effects, find instruments orthogonal to them (Arellano and Bover 1995)

- If $\mathrm{E}[y_{it}\mu_i]$ stationary, then $\mathrm{E}[\Delta y_{it}\mu_i] = 0$
- $\Delta y_{i,t-1}$ uncorrelated with fixed effects, thus with $\nu_{it} = \mu$ good instrument in *levels* (if no AR)
- Make system of difference and levels equations
- Concretely, make a stacked data set, with differenc up top, levels below. Treat as single estimation pro
- Instrument differences with levels and v.v.
- "System GMM" (Blundell and Bond 1998)

Relationship among moments (Tue Gorgens)

$$
\begin{array}{ccccccc}
\mathrm{E}\left[w_{i1}\upsilon_{i1}\right] & \underline{\underline{D}} & \mathrm{E}\left[w_{i1}\upsilon_{i2}\right] & \underline{\underline{D}} & \mathrm{E}\left[w_{i1}\upsilon_{i3}\right] & \underline{\underline{D}} & \mathrm{E}\left[w_{i1}\upsilon_{i4}\right] \\
 & & \| L & & \| L & & \| L \\
\mathrm{E}\left[w_{i2}\upsilon_{i1}\right] & & \mathrm{E}\left[w_{i2}\upsilon_{i2}\right] & \underline{\underline{D}} & \mathrm{E}\left[w_{i2}\upsilon_{i3}\right] & \underline{\underline{D}} & \mathrm{E}\left[w_{i2}\upsilon_{i4}\right] \\
 & & & & \| L & & \| L \\
\mathrm{E}\left[w_{i3}\upsilon_{i1}\right] & & \mathrm{E}\left[w_{i3}\upsilon_{i2}\right] & & \mathrm{E}\left[w_{i3}\upsilon_{i3}\right] & \underline{\underline{D}} & \mathrm{E}\left[w_{i3}\upsilon_{i4}\right] \\
 & & & & & & \| L \\
\mathrm{E}\left[w_{i4}\upsilon_{i1}\right] & & \mathrm{E}\left[w_{i4}\upsilon_{i2}\right] & & \mathrm{E}\left[w_{i4}\upsilon_{i3}\right] & & \mathrm{E}\left[w_{i4}\upsilon_{i4}\right]
\end{array}
$$

Problem: Two-step errors too small

Regression for Arellano-Bond (1991) column (a1), Table 4

Arellano-Bond dynamic panel-data estimation, **one-step** difference GMM results

	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
n						
L1.	.6862261	.1445943	4.75	0.000	.4003376	.9721147
L2.	-.0853582	.0560155	-1.52	0.130	-.1961109	.0253944
w						
--.	-.6078208	.1782055	-3.41	0.001	-.9601647	-.2554769
L1.	.3926237	.1679931	2.34	0.021	.0604714	.7247759
k						
--.	.3568456	.0590203	6.05	0.000	.240152	.4735392

Arellano-Bond dynamic panel-data estimation, **two-step** difference GMM results

	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
n						
L1.	.6287089	.0904543	6.95	0.000	.4498646	.8075531
L2.	-.0651882	.0265009	-2.46	0.015	-.1175852	-.0127912
w						
--.	-.5257597	.0537692	-9.78	0.000	-.6320709	-.4194485
L1.	.3112899	.0940116	3.31	0.001	.1254122	.4971675
k						
--.	.2783619	.0449083	6.20	0.000	.1895702	.3671537

Problem, cont'd

- Problem appears to be one of overfitting
  - Efficient GMM deemphasizes moments with high variance (high second moments)
  - Feasible efficient GMM in small samples may deemphasize outliers (high first moments)
  - Spurious precision

Solution: finite-sample correction (Windmeijer 2005)

- One-step estimate: $\hat{\boldsymbol{\beta}}_1 = f(\mathbf{Y})$ (conditioning on **X**, **Z**)
- One-step residuals used to construct $\hat{\boldsymbol{\Omega}}$:

$$\hat{\boldsymbol{\beta}}_2 = \left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\hat{\boldsymbol{\Omega}}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\hat{\boldsymbol{\Omega}}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{Y} \equiv g(\mathbf{Y},\hat{\boldsymbol{\Omega}}) \equiv g(\mathbf{Y}, f(\mathbf{Y}))$$

- Standard estimate of $\mathrm{Var}\left[\hat{\boldsymbol{\beta}}_2\right]$ treats $\hat{\boldsymbol{\Omega}}$ as constant, observed, precise—despite dependence on random **Y**
- Taylor expansion of g around true $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}}_2 = g\left(\mathbf{Y},\hat{\boldsymbol{\Omega}}_{\hat{\beta}_1}\right) \approx g\left(\mathbf{Y},\hat{\boldsymbol{\Omega}}_{\beta}\right) + \frac{\partial}{\partial\hat{\boldsymbol{\beta}}} g\left(\mathbf{Y},\hat{\boldsymbol{\Omega}}_{\hat{\beta}}\right)\Big|_{\hat{\boldsymbol{\beta}}=\boldsymbol{\beta}} \left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}\right)$$

- "Correction" comes from second term
  - $\mathrm{E}\left[\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}\right] = 0$ so no effect on $\mathrm{E}\left[\hat{\boldsymbol{\beta}}_2\right]$—no coefficient bias
  - Affects variance

Arellano-Bond dynamic panel-data estimation, **one-step** difference GMM results

	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
n						
L1.	.6862261	.1445943	4.75	0.000	.4003376	.9721147
L2.	-.0853582	.0560155	-1.52	0.130	-.1961109	.0253944
w						
--.	-.6078208	.1782055	-3.41	0.001	-.9601647	-.2554769
L1.	.3926237	.1679931	2.34	0.021	.0604714	.7247759
k						
--.	.3568456	.0590203	6.05	0.000	.240152	.4735392

Arellano-Bond dynamic panel-data estimation, **two-step** difference GMM results

	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
n						
L1.	.6287089	.0904543	6.95	0.000	.4498646	.8075531
L2.	-.0651882	.0265009	-2.46	0.015	-.1175852	-.0127912
w						
--.	-.5257597	.0537692	-9.78	0.000	-.6320709	-.4194485
L1.	.3112899	.0940116	3.31	0.001	.1254122	.4971675
k						
--.	.2783619	.0449083	6.20	0.000	.1895702	.3671537

Arellano-Bond dynamic panel-data estimation, **two-step** difference GMM results

	Coef.	**Corrected** Std. Err.	t	P>|t|	[95% Conf. Interval]	
n						
L1.	.6287089	.1934138	3.25	0.001	.2462954	1.011122
L2.	-.0651882	.0450501	-1.45	0.150	-.1542602	.0238838
w						
--.	-.5257597	.1546107	-3.40	0.001	-.8314524	-.2200669
L1.	.3112899	.2030006	1.53	0.127	-.0900784	.7126582
k						
--.	.2783619	.0728019	3.82	0.000	.1344196	.4223043

Problem: too many instruments

- In difference and system GMM, # instruments (j) quadratic in T
- Analogy:
  - In 2SLS, if j = # of regressors, first-stage $R2$'s=1.0 and 2SLS=OLS (biased)
  - Too many instruments overfit endogenous variables
- And # of cross-moments in $\mathrm{Var}\left[\mathbf{Z}'\mathbf{E}\middle|\mathbf{X},\mathbf{Z}\right]^{-1}$ to be estimated for efficient GMM quadratic in j—quartic in T!
- Estimate of $\mathrm{Var}\left[\mathbf{Z}'\mathbf{E}\middle|\mathbf{X},\mathbf{Z}\right]^{-1}$ degrades
- Hansen test very weak—p values of 1.000 not uncommon
- Little guidance on how many is too many
- `xtabond2` warns if $j > N$

Solution: consider limiting instruments

- Limit number of lags of variables used as instruments
- Or "collapse" instruments:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ y_{i1} & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & y_{i2} & y_{i1} & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & y_{i3} & y_{i2} & y_{i1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \Rightarrow \begin{bmatrix} 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ y_{i1} & 0 & 0 & \cdots \\ y_{i2} & y_{i1} & 0 & \cdots \\ y_{i3} & y_{i2} & y_{i1} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

$$\sum_i y_{i,t-2}\Delta\hat{e}_{it} = 0 \text{ for each } t \geq 3 \Rightarrow \sum_{i,t} y_{i,t-2}\Delta e_{it} = 0.$$

xtabond2 syntax

Y X Z

xtabond2 ***depvar varlist*** **[if** ***exp*****] [in** ***range*****]**
[, level(#) twostep robust noconstant small noleveleq
artests(#) arlevels h(#) nomata]
ivopt [ivopt ...] gmmopt [gmmopt ...]]

where **gmmopt** is

“GMM-style”

gmmstyle(***varlist*** **[, laglimits(# #) collapse**
equation({diff | level | both}) passthru])

and **ivopt** is

Classic

ivstyle(***varlist*** **[, equation({diff | level | both})**
passthru mz])

Examples

- Classic one-step difference GMM with no controls except time dummies
```
xi: xtabond2 y L.y i.t, gmm(y, laglim(2 .))
  iv(i.t) robust noleveleq
```

- Equivalents:
```
xi: xtabond2 y L.y i.t, gmm(L.y, laglim(1 .))
  iv(i.t) robust noleveleq
xi: xtabond2 y L.y i.t, gmm(L.y)
  iv(i.t) robust noleveleq
```

- System GMM, two-step, Windmeijer correction, w1 exogenous, w2 predetermined, w3 exogenous:
```
xi: xtabond2 y L.y w1 w2 w3 i.t,
  gmm(L.y w2 L.w3) iv(i.t w1) two robust
```

Examples, cont'd

If conditions imposed only on levels, difference equation effectively discarded.

Equivalent pairs:

```
regress n w k
xtabond2 n w k, iv(w k, eq(level)) small

ivreg2 n cap (w = k ys)
xtabond2 n w cap, iv(cap k ys, eq(level))

ivreg2 n cap (w = k ys), cluster(id) gmm
xtabond2 n w cap, iv(cap k ys, eq(level)) two
```

Or even:

```
regress n w k
abar, lags(2)

xtabond2 n w k, iv(w k, eq(level)) small arlevel
```

Run times for bbest (seconds) 700 MHz PC

`xtabond2` ado	57
`xtabond2` Mata, favoring space	14
`xtabond2` Mata, favoring speed	11
DPD for Ox	3