Generalized method of moments estimation of linear dynamic panel data models

Sebastian Kripfganz

University of Exeter Business School, Department of Economics, Exeter, UK

London Stata Conference

September 5, 2019

```
ssc install xtdpdgmm
net install xtdpdgmm, from(http://www.kripfganz.de/stata/)
```

GMM estimation of linear dynamic panel data models

- Panel data / longitudinal data allows to account for unobserved unit-specific heterogeneity and to model dynamic adjustment / feedback processes.
- Instrumental variables (IV) / generalized method of moments (GMM) estimation is the predominant estimation technique for models with endogenous variables, in particular lagged dependent variables, when the time horizon is short.
- This presentation introduces the community-contributed `xtdpdgmm` Stata command.

Some Stata milestones

- December 15, 2000: Stata 7 released with the new `xtabond` command for the Arellano and Bond (1991) *difference GMM* (diff-GMM) estimation.
- November 26, 2003: David Roodman announced the community-contributed `xtabond2` command for Arellano and Bover (1995) and Blundell and Bond (1998) *system GMM* (sys-GMM) estimation.
- June 25, 2007: Stata 10 released with the new `xtdpdsys` command for sys-GMM estimation. Both `xtabond` and `xtdpdsys` are wrappers for the `xtdpd` command.

Some Stata milestones

- March 2009: David Roodman's "How to do xtabond2" article appeared in the Stata Journal.
- July 13, 2009: Stata 11 released with the new `gmm` command for GMM estimation (not just of dynamic panel data models).
- December 2012: Stata Journal Editor's Prize for David Roodman.
- June 1, 2017: New community-contributed `xtdpdgmm` command for sys-GMM estimation and GMM estimation with the Ahn and Schmidt (1995) nonlinear moment conditions announced on Statalist.

Equivalent diff-GMM implementations in Stata[1]

```
. webuse abdata

. xtabond n, la(1) maxld(3) pre(w k) maxlag(3) nocons vce(r)

. xtdpd L(0/1).n w k, dgmm(L.n w k, lag(1 3)) nocons vce(r)

. xtabond2 L(0/1).n w k, gmm(L.n w k, lag(1 3) e(d)) nol r

. xtdpdgmm L(0/1).n w k, gmm(L.n w k, l(1 3) m(d)) nocons vce(r)

. gmm (D.n - {b1}*LD.n - {b2}*D.w - {b3}*D.k), ///
> xtinst(L.n w k, lags(1/3)) inst(, nocons) winit(xt D) one vce(r)
```

[1] Note: The examples in this presentation are oversimplified for expositional purposes. Throughout the presentation, the Arellano and Bond (1991) data set is used.

Equivalent system-GMM implementations in Stata

```
. xtdpdsys n, la(1) maxld(3) pre(w k) maxlag(3) two

. xtdpd L(0/1).n w k, dgmm(L.n w k, lag(1 3)) lgmm(L.n w k, lag(0)) two

. xtabond2 L(0/1).n w k, gmm(L.n w k, lag(1 3)) h(2) two

. xtdpdgmm L(0/1).n w k, gmm(L.n w k, l(1 3) m(d)) ///
> gmm(L.n w k, d l(0 0)) w(ind) two

. gmm (D.n - {b1}*LD.n - {b2}*D.w - {b3}*D.k) ///
> (n - {b1}*L.n - {b2}*w - {b3}*k - {c}), ///
> xtinst(1: L.n w k, lags(1/3)) inst(1:, nocons) ///
> xtinst(2: D.(L.n w k), lags(0)) winit(xt DL) wmat(r) vce(un) nocommonesample
```

GMM estimation

- $L \times 1$ vector of moment conditions:

$$E[\mathbf{m}_i(\boldsymbol{\theta})] = \mathbf{0}$$

as a function of a $K \times 1$ parameter vector $\boldsymbol{\theta}$, with $L \geq K$.

  - For example, linear regression model $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\theta} + \mathbf{e}_i$ with endogenous regressors $\mathbf{X}_i$ and instrumental variables $\mathbf{Z}_i$:

$$\mathbf{m}_i(\boldsymbol{\theta}) = \mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta}) = \mathbf{Z}_i'\mathbf{e}_i$$

- The GMM estimator minimizes a quadratic form:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\mathbf{b}} \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{m}_i(\mathbf{b})\right)' \mathbf{W} \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{m}_i(\mathbf{b})\right)$$

given a random sample of size N and weighting matrix $\mathbf{W}$.

GMM estimation

- When the model is overidentified, i.e. $L > K$, an asymptotically efficient estimator requires the weighting matrix to be *optimal*, i.e. a consistent estimate of the inverse of the asymptotic covariance matrix of $\mathbf{m}(\hat{\boldsymbol{\theta}})$:

$$\mathbf{W}(\hat{\boldsymbol{\theta}}) = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{m}_i(\hat{\boldsymbol{\theta}}) \mathbf{m}_i(\hat{\boldsymbol{\theta}})' \right)^{-1}$$

- $\mathbf{W}(\hat{\boldsymbol{\theta}})$ can be obtained from an inefficient initial GMM estimator based on some suboptimal choice of $\mathbf{W}$.
- The feasible efficient (two-step) GMM estimator is then

$$\hat{\hat{\boldsymbol{\theta}}} = \arg\min_{\mathbf{b}} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{m}_i(\mathbf{b}) \right)' \mathbf{W}(\hat{\boldsymbol{\theta}}) \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{m}_i(\mathbf{b}) \right)$$

Linear dynamic panel data model

- Autoregressive distributed lag (ARDL) panel data model:

$$y_{it} = \sum_{j=1}^{q_y} \lambda_j y_{i,t-j} + \sum_{j=0}^{q_x} \mathbf{x}'_{i,t-j} \boldsymbol{\beta}_j + \underbrace{\alpha_i + u_{it}}_{=e_{it}}$$

with many cross-sectional units $i = 1, 2, \ldots, N$ and few time periods $t = 1, 2, \ldots, T$.

  - The regressors $\mathbf{x}_{it}$ can be
    - strictly exogenous, $E[u_{it}|\mathbf{x}_{i0}, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}] = 0$,
    - weakly exogenous / predetermined, $E[u_{it}|\mathbf{x}_{i0}, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}] = 0$,
    - endogenous, $E[u_{it}|\mathbf{x}_{i0}, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{i,t-1}] = 0.$[2]
  - The idiosyncratic error term u_{it} shall be serially uncorrelated.
  - The unobserved unit-specific heterogeneity α_i can be correlated with the regressors $\mathbf{x}_{i,t-j}$. It is correlated by construction with the lagged dependent variables $y_{i,t-j}$.

[2]For simplicity, we exclude feedback from past regressors to current shocks.

Diff-GMM estimation: transformation and instruments

- First-difference transformation of the model:[3]

$$\Delta y_{it} = \sum_{j=1}^{q_y} \lambda_j \Delta y_{i,t-j} + \sum_{j=0}^{q_x} \Delta \mathbf{x}'_{i,t-j}\beta_j + \underbrace{\Delta u_{it}}_{=\Delta e_{it}}$$

  - $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$ and first differences of other predetermined variables are correlated with $\Delta u_{it} = u_{it} - u_{i,t-1}$.
  - Anderson and Hsiao (1981) propose an IV estimator with $\Delta y_{i,t-2}$ or $y_{i,t-2}$ as instruments for $\Delta y_{i,t-1}$.
  - Arellano and Bond (1991) suggest to use further lags of the levels as instruments. In particular, $y_{i,t-2}, y_{i,t-3}, \ldots$ are uncorrelated with Δu_{it} but (hopefully) correlated with $\Delta y_{i,t-1}$.
  - For endogenous regressors, the lagged levels $\mathbf{x}_{i,t-2}, \mathbf{x}_{i,t-3}, \ldots$ qualify as instruments. For predetermined regressors, $\mathbf{x}_{i,t-1}$ qualify as additional instruments.

[3] For simplicity, assume in the following that $q_y = 1$ and $q_x = 0$.

Diff-GMM estimation: moment conditions

- Moment conditions for the first-differenced model:
  - Lagged dependent variable:

$$E[y_{i,t-s}\Delta u_{it}] = 0, \quad s = 2, 3, \dots, t$$

  - Strictly exogenous regressors:

$$E[\mathbf{x}_{i,t-s}\Delta u_{it}] = \mathbf{0}, \quad t - s = 0, 1, \dots, T$$

  - Predetermined regressors:

$$E[\mathbf{x}_{i,t-s}\Delta u_{it}] = \mathbf{0}, \quad s = 1, 2, \dots, t$$

  - Endogenous regressors:

$$E[\mathbf{x}_{i,t-s}\Delta u_{it}] = \mathbf{0}, \quad s = 2, 3, \dots, t$$

with $t = s, \dots, T$.

Diff-GMM estimation: GMM-type instruments

- Stacked moment conditions:

$$E[\mathbf{m}_i(\boldsymbol{\theta})] = E\left[\mathbf{Z}_i^{D\prime}\Delta\mathbf{u}_i\right] = \mathbf{0}$$

where $\boldsymbol{\theta} = (\lambda, \beta)$, $\Delta\mathbf{u}_i = (\Delta u_{i2}, \Delta u_{i3}, \ldots, \Delta u_{iT})'$, and $\mathbf{Z}_i^D = (\mathbf{Z}_{yi}^D, \mathbf{Z}_{xi}^D)$, with *GMM-type* instruments

$$\mathbf{Z}_{yi}^D = \begin{pmatrix} y_{i0} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & y_{i0} & y_{i1} & \cdots & 0 & 0 & \cdots & 0 \\ & & & \ddots & & & & \\ 0 & 0 & 0 & \cdots & y_{i0} & y_{i1} & \cdots & y_{i,T-2} \end{pmatrix} \begin{array}{l} \leftarrow t = 2 \\ \leftarrow t = 3 \\ \quad\ \vdots \\ \leftarrow t = T \end{array}$$

and similarly for $\mathbf{Z}_{xi}^D$.

  - With `xtdpdgmm`, the option `model(difference)` creates instruments for the first-difference transformed model.

Diff-GMM estimation: initial weighting matrix

- When u_{it} is serially uncorrelated and homoskedastic, the optimal weighting matrix is independent of $\boldsymbol{\theta}$ such that we can use the one-step instead of the two-step estimator: $\mathbf{W} = \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{Z}_i^{D\prime}\mathbf{D}_i\mathbf{D}_i'\mathbf{Z}_i^D\right)^{-1}$, where $\mathbf{D}_i$ is the $T-1 \times T$ first-difference transformation matrix:

$$\mathbf{D}_i = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}$$

  such that $\Delta\mathbf{u}_i = \mathbf{D}_i\mathbf{u}_i$.
    - This weighting matrix accounts for the first-order serial correlation of Δu_{it}.

One-step diff-GMM estimation in Stata

- *GMM-type* instruments specified with the `gmmiv()` option, exemplarily for predetermined `w` and strictly exogenous `k`:

```
. xtdpdgmm L(0/1).n w k, model(diff) gmm(n, lag(2 .)) gmm(w, lag(1 .)) gmm(k, lag(. .)) nocons
note: standard errors may not be valid

Generalized method of moments estimation

Fitting full model:
Step 1         f(b) =  .01960406

Group variable: id                            Number of obs         =        891
Time variable: year                           Number of groups      =        140

Moment conditions:     linear =       126     Obs per group:    min =          6
                    nonlinear =         0                       avg =   6.364286
                        total =       126                       max =          8

------------------------------------------------------------------------------
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .4144164   .0341502    12.14   0.000     .3474833    .4813495
             |
           w |  -.8292293   .0588914   -14.08   0.000    -.9446543   -.7138042
           k |   .3929936   .0223829    17.56   0.000     .3491239    .4368634
------------------------------------------------------------------------------
```
(Continued on next page)

One-step diff-GMM estimation in Stata

```
Instruments corresponding to the linear moment conditions:
 1, model(diff):
   1978:L2.n 1979:L2.n 1980:L2.n 1981:L2.n 1982:L2.n 1983:L2.n 1984:L2.n
   1979:L3.n 1980:L3.n 1981:L3.n 1982:L3.n 1983:L3.n 1984:L3.n 1980:L4.n
   1981:L4.n 1982:L4.n 1983:L4.n 1984:L4.n 1981:L5.n 1982:L5.n 1983:L5.n
   1984:L5.n 1982:L6.n 1983:L6.n 1984:L6.n 1983:L7.n 1984:L7.n 1984:L8.n
 2, model(diff):
   1978:L1.w 1979:L1.w 1980:L1.w 1981:L1.w 1982:L1.w 1983:L1.w 1984:L1.w
   1978:L2.w 1979:L2.w 1980:L2.w 1981:L2.w 1982:L2.w 1983:L2.w 1984:L2.w
   1979:L3.w 1980:L3.w 1981:L3.w 1982:L3.w 1983:L3.w 1984:L3.w 1980:L4.w
   1981:L4.w 1982:L4.w 1983:L4.w 1984:L4.w 1981:L5.w 1982:L5.w 1983:L5.w
   1984:L5.w 1982:L6.w 1983:L6.w 1984:L6.w 1983:L7.w 1984:L7.w 1984:L8.w
 3, model(diff):
   1978:F6.k 1978:F5.k 1979:F5.k 1978:F4.k 1979:F4.k 1980:F4.k 1978:F3.k
   1979:F3.k 1980:F3.k 1981:F3.k 1978:F2.k 1979:F2.k 1980:F2.k 1981:F2.k
   1982:F2.k 1978:F1.k 1979:F1.k 1980:F1.k 1981:F1.k 1982:F1.k 1983:F1.k
   1978:k 1979:k 1980:k 1981:k 1982:k 1983:k 1984:k 1978:L1.k 1979:L1.k
   1980:L1.k 1981:L1.k 1982:L1.k 1983:L1.k 1984:L1.k 1978:L2.k 1979:L2.k
   1980:L2.k 1981:L2.k 1982:L2.k 1983:L2.k 1984:L2.k 1979:L3.k 1980:L3.k
   1981:L3.k 1982:L3.k 1983:L3.k 1984:L3.k 1980:L4.k 1981:L4.k 1982:L4.k
   1983:L4.k 1984:L4.k 1981:L5.k 1982:L5.k 1983:L5.k 1984:L5.k 1982:L6.k
   1983:L6.k 1984:L6.k 1983:L7.k 1984:L7.k 1984:L8.k
```

- `xtdpdgmm` has the options `nolog`, `noheader`, `notable`, and `nofootnote` to suppress undesired output.

Diff-GMM estimation: optimal weighting matrix

- When u_{it} is heteroskedastic, panel-robust or cluster-robust standard errors can be computed with options `vce(robust)` or `vce(cluster` *`clustvar`*`)`.
  - In general, cluster-robust standard errors are robust to serially correlated u_{it} as well. Yet, the instruments $y_{i,t-2}, y_{i,t-3}, \ldots$ would become invalid and the GMM estimator inconsistent.
  - The one-step GMM estimator remains consistent under heteroskedasticity but it is no longer efficient.
- The efficient two-step estimator uses optimal weighting matrix $\mathbf{W}(\hat{\boldsymbol{\theta}}) = \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{Z}_i^{D\prime} \Delta\hat{\mathbf{u}}_i \Delta\hat{\mathbf{u}}_i' \mathbf{Z}_i^D\right)^{-1}$ or its cluster-robust analogue (option `twostep` of `xtdpdgmm`).
  - The default two-step standard errors are biased in finite samples due to the neglected sampling error in $\mathbf{W}(\hat{\boldsymbol{\theta}})$. With options `vce(robust)` or `vce(cluster` *`clustvar`*`)`, the Windmeijer (2005) finite-sample correction is applied. (The corrected standard errors are still biased but less severely).

Two-step diff-GMM estimation in Stata

```
. xtdpdgmm L(0/1).n w k, model(diff) gmm(n, lag(2 .)) gmm(w, lag(1 .)) gmm(k, lag(. .)) nocons two ///
> vce(r) nofootnote

Generalized method of moments estimation

Fitting full model:
Step 1         f(b) =  .01960406
Step 2         f(b) =  .90967907

Group variable: id                           Number of obs         =       891
Time variable: year                          Number of groups      =       140

Moment conditions:     linear =      126     Obs per group:    min =         6
                    nonlinear =        0                       avg =  6.364286
                        total =      126                       max =         8

                                (Std. Err. adjusted for 140 clusters in id)
------------------------------------------------------------------------------
             |              WC-Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .4126102   .0740256     5.57   0.000     .2675228    .5576977
             |
           w |  -.8271943   .0944749    -8.76   0.000    -1.012362   -.6420268
           k |   .3931545   .0484993     8.11   0.000     .2980975    .4882115
------------------------------------------------------------------------------
```

Too-many-instruments problem

- The model is usually strongly overidentified, $L \gg K$.
- The number of instruments increases quickly with the number of regressors and the number of time periods.
- Too many instruments relative to the cross-sectional sample size can cause biased coefficient and standard error estimates and weakened specification tests (Roodman, 2009a).
  - Too many instruments can overfit the instrumented variables.
  - The optimal weighting matrix is of dimension $L \times L$ which becomes difficult to estimate when L is large relative to N.
  - Instrument proliferation can lead to substantial underrejection of overidentification tests, thus incorrectly signaling too often that the model is correctly specified when it is not.

Too-many-instruments problem: instrument reduction

- To reduce the number of instruments, two main approaches are typically used (Roodman, 2009a, 2009b; Kiviet, 2019):
  - Curtailing: Use only a limited number of lags as instruments, e.g. $y_{i,t-2}, y_{i,t-3}, \ldots, y_{i,t-l}$, with $t - l > 1$. For strictly exogenous regressors, it is common practice not to use leads $\mathbf{x}_{i,t-s}, s < 0$, as instruments.
  - Collapsing: Instead of the "GMM-type" instruments, use "standard" instruments, e.g.

$$
\mathbf{Z}_{yi}^{D} = \begin{array}{cc}
\begin{pmatrix}
y_{i0} & 0 & \cdots & 0 \\
y_{i1} & y_{i0} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
y_{i,T-2} & y_{i,T-3} & \cdots & y_{i0}
\end{pmatrix} &
\begin{array}{l}
\leftarrow \; t = 2 \\
\leftarrow \; t = 3 \\
\qquad \vdots \\
\leftarrow \; t = T
\end{array}
\end{array}
$$

The moment conditions $E[y_{i,t-s} \Delta u_{it}] = 0$ for individual time periods t are replaced by $E\left[\sum_{t=s}^{T} y_{i,t-s} \Delta u_{it}\right] = 0$.

Two-step diff-GMM estimation in Stata

- Combination of curtailed and collapsed instruments:

```
. xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w, lag(1 3)) gmm(k, lag(0 2)) ///
> nocons two vce(r) nolog

Generalized method of moments estimation

Moment conditions:     linear =          9      Obs per group:    min =         6
                    nonlinear =          0                        avg =  6.364286
                        total =          9                        max =         8

                                   (Std. Err. adjusted for 140 clusters in id)
------------------------------------------------------------------------------
             |              WC-Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .3564619   .1074848     3.32   0.001     .1457956    .5671281
             |
           w |  -1.432958   .2141048    -6.69   0.000    -1.852595   -1.01332
           k |   .2860594   .0541221     5.29   0.000     .1799821    .3921367
------------------------------------------------------------------------------
Instruments corresponding to the linear moment conditions:
 1, model(diff):
   L2.n L3.n L4.n
 2, model(diff):
   L1.w L2.w L3.w
 3, model(diff):
   k L1.k L2.k
```

Curtailed and collapsed GMM-type instruments

- The suboption lagrange() defines the first and last lag to be used, and a dot / missing value means to use all available lags.
- xtdpdgmm has a *global* option collapse that causes all *GMM-type* instruments to be collapsed.
  - The default set by this option can be overwritten for individual subsets of *GMM-type* instruments with the suboption [no]collapse.

```
. xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4) nocollapse) gmm(w, lag(1 3)) ///
> gmm(k, lag(0 2)) nocons two vce(r)
(Some output omitted)
Instruments corresponding to the linear moment conditions:
 1, model(diff):
   1978:L2.n 1979:L2.n 1980:L2.n 1981:L2.n 1982:L2.n 1983:L2.n 1984:L2.n
   1979:L3.n 1980:L3.n 1981:L3.n 1982:L3.n 1983:L3.n 1984:L3.n 1980:L4.n
   1981:L4.n 1982:L4.n 1983:L4.n 1984:L4.n
 2, model(diff):
   L1.w L2.w L3.w
 3, model(diff):
   k L1.k L2.k

. xtdpdgmm L(0/1).n w k, model(diff) gmm(n, lag(2 4)) gmm(w, lag(1 3) collapse) ///
> gmm(k, lag(0 2) collapse) nocons two vce(r)
(Output omitted)
```

GMM-type and standard instruments

- Collapsed *GMM-type* instruments, gmmiv() with option collapse, are equivalent to *standard* instruments, iv():

```
. xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w, lag(1 3)) gmm(k, lag(0 2)) ///
> nocons two vce(r)
(Output omitted)

. xtdpdgmm L(0/1).n w k, model(diff) iv(n, lag(2 4)) iv(w, lag(1 3)) iv(k, lag(0 2)) nocons two vce(r)
(Output omitted)
```

- Uncollapsed *GMM-type* instruments are *standard* instruments interacted with time dummies (Kiviet, 2019):

```
. xtdpdgmm L(0/1).n w k, model(diff) gmm(n, lag(2 4)) gmm(w, lag(1 3)) gmm(k, lag(0 2)) nocons two vce(r)
(Output omitted)

. xtdpdgmm L(0/1).n w k, model(diff) iv(i.year#cL(2/4).n) iv(i.year#cL(1/3).w) iv(i.year#cL(0/2).k) ///
> nocons two vce(r)
(Output omitted)
```

- In all cases, missing values in the instruments are replaced by zeros without dropping the observations.

Arellano-Bond serial-correlation test

- If u_{it} is serially uncorrelated, then Δu_{it} has negative first-order serial correlation, $Corr(\Delta u_{it}, \Delta u_{i,t-1}) = -0.5$, but no higher-order serial correlation.
- Absence of higher-order serial correlation of Δu_{it} is crucial for the validity of $y_{i,t-2}, y_{i,t-3}, \ldots$ as instruments, and similarly for the instruments of predetermined and endogenous $\mathbf{x}_{it}$.
- Arellano and Bond (1991) suggest an asymptotically $\mathcal{N}(0,1)$ distributed test statistic for the null hypothesis $H_0 : Corr(\Delta u_{it}, \Delta u_{i,t-j}) = 0$, $j > 0$.
    - The model passes this specification test if H_0 is rejected for $j = 1$ and not rejected for $j > 1$.
    - Not rejecting H_0 for $j = 1$ can be a sign of trouble (e.g. indicating that u_{it} follows a near-unit root process).
    - After `xtdpdgmm`, these tests are obtained with the postestimation command `estat serial`.

Sargan's overidentification tests

- In just-identified models, $L = K$, the validity of the instruments is an untested assumption.
  - $\sum_{i=1}^N \mathbf{m}_i(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^N \mathbf{Z}_i^{d\prime} \Delta \hat{\mathbf{u}}_i = \mathbf{0}$.
- In overidentified models, $L > K$, the validity of $L - K$ overidentifying restrictions can be tested, still assuming that at least K instruments are valid.
  - $\sum_{i=1}^N \mathbf{m}_i(\hat{\boldsymbol{\theta}}) \neq \mathbf{0}$ but close to zero if the model is correctly specified.
- After one-step estimation, the Sargan (1958) test statistic is asymptotically $\chi^2(df)$ distributed with $df = L - K$ degress of freedom, provided that $\mathbf{W}$ is an optimal weighting matrix:

$$J(\hat{\boldsymbol{\theta}}, \mathbf{W}) = \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{m}_i(\hat{\boldsymbol{\theta}})\right)' \mathbf{W} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{m}_i(\hat{\boldsymbol{\theta}})\right)$$

Hansen's overidentification tests

- After two-step estimation with optimal weighting matrix $\mathbf{W}(\hat{\theta})$, the Hansen (1982) test statistic is as well asymptotically $\chi^2(L-K)$ distributed:

$$J(\hat{\hat{\theta}}, \mathbf{W}(\hat{\theta})) = \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{m}_i(\hat{\hat{\theta}}) \right)' \mathbf{W}(\hat{\theta}) \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{m}_i(\hat{\hat{\theta}}) \right)$$

or with iterated weighting matrix:

$$J(\hat{\hat{\theta}}, \mathbf{W}(\hat{\hat{\theta}})) = \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{m}_i(\hat{\hat{\theta}}) \right)' \mathbf{W}(\hat{\hat{\theta}}) \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{m}_i(\hat{\hat{\theta}}) \right)$$

  - Under the null hypothesis, the overidentifying restrictions are valid, i.e. $E[\mathbf{m}_i(\theta)] = \mathbf{0}$.

Overidentification tests

- The `xtdpdgmm` postestimation command `estat overid` reports $J(\hat{\theta}, \mathbf{W})$ and $J(\hat{\theta}, \mathbf{W}(\hat{\theta}))$ after one-step estimation, and $J(\hat{\hat{\theta}}, \mathbf{W}(\hat{\theta}))$ and $J(\hat{\hat{\theta}}, \mathbf{W}(\hat{\hat{\theta}}))$ after two-step estimation.
  - If the initial weighting matrix $\mathbf{W}$ is not optimal, then both test statistics reported after one-step estimation are asymptotically invalid.
  - Both test statistics reported after two-step estimation are asymptotically equivalent. A large difference in finite samples indicates that the weighting matrix $\mathbf{W}(\hat{\theta})$ is imprecisely estimated.
  - If $\mathbf{W}$ is optimal, then all four test statistics are asymptotically equivalent but they might have different finite-sample properties.

Specification testing in Stata

```
. quietly xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w, lag(1 3)) ///
> gmm(k, lag(0 2)) nocons two vce(r)

. estat serial, ar(1/3)

Arellano-Bond test for autocorrelation of the first-differenced residuals
H0: no autocorrelation of order 1:     z =   -2.6865   Prob > |z|  =    0.0072
H0: no autocorrelation of order 2:     z =   -0.9414   Prob > |z|  =    0.3465
H0: no autocorrelation of order 3:     z =   -0.3256   Prob > |z|  =    0.7447

. estat overid

Sargan-Hansen test of the overidentifying restrictions
H0: overidentifying restrictions are valid

2-step moment functions, 2-step weighting matrix       chi2(6)      =   11.9878
                                                       Prob > chi2 =    0.0622

2-step moment functions, 3-step weighting matrix       chi2(6)      =   12.8283
                                                       Prob > chi2 =    0.0458
```

- The overidentification test does not provide confidence in the model specification.

Specification testing in Stata

- `k` classified as predetermined instead of strictly exogenous:

```
. xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) nocons two vce(r) nolog

Generalized method of moments estimation

Moment conditions:      linear =        9      Obs per group:    min =         6
                     nonlinear =        0                        avg =  6.364286
                         total =        9                        max =         8

                                   (Std. Err. adjusted for 140 clusters in id)
------------------------------------------------------------------------------
             |              WC-Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .5234179   .1316921     3.97   0.000     .2653061    .7815298
             |
           w |  -1.883857   .3499077    -5.38   0.000    -2.569663    -1.19805
           k |   -.020718   .1603249    -0.13   0.897    -.3349491    .2935131
------------------------------------------------------------------------------
Instruments corresponding to the linear moment conditions:
 1, model(diff):
   L2.n L3.n L4.n
 2, model(diff):
   L1.w L2.w L3.w L1.k L2.k L3.k
```

Specification testing in Stata

```
. estat serial, ar(1/3)

Arellano-Bond test for autocorrelation of the first-differenced residuals
H0: no autocorrelation of order 1:     z =   -2.7781   Prob > |z|  =    0.0055
H0: no autocorrelation of order 2:     z =   -1.1426   Prob > |z|  =    0.2532
H0: no autocorrelation of order 3:     z =   -0.1114   Prob > |z|  =    0.9113

. estat overid

Sargan-Hansen test of the overidentifying restrictions
H0: overidentifying restrictions are valid

2-step moment functions, 2-step weighting matrix       chi2(6)     =    4.9542
                                                       Prob > chi2 =    0.5497

2-step moment functions, 3-step weighting matrix       chi2(6)     =    4.5136
                                                       Prob > chi2 =    0.6075
```

- The specification tests provide more confidence in this new model specification.

Sys-GMM estimation: initial-conditions assumption

- The instruments $y_{i,t-2}, y_{i,t-3}, \ldots$ are weakly correlated with the first-differenced lagged dependent variable $\Delta y_{i,t-1}$ when $\lambda \rightarrow 1$.[4] In particular when T is small, the diff-GMM estimator could be substantially biased.
  - Blundell and Bond (1998) show that under the initial-conditions assumption $E[\Delta y_{i1} \alpha_i] = 0$, the first differences $\Delta y_{i,t-1}$ become available as instruments for $y_{i,t-1}$. A sufficient but not necessary condition is joint mean stationarity of the y_{it} and $\mathbf{x}_{it}$ processes (Blundell, Bond, and Windmeijer, 2001).
  - Under the assumption that the predetermined variables $\mathbf{x}_t$ have constant correlation over time with α_i, Arellano and Bover (1995) already proposed to use first differences $\Delta \mathbf{x}_t$ as instruments.

[4] See Gørgens, Han, and Xue (2019) for a recent discussion of potential diff-GMM identification failures even for any value of $\lambda \in [0, 1]$.

Sys-GMM estimation: moment conditions

- Additional moment conditions for the level model:
  - Lagged dependent variable:

$$E[\Delta y_{i,t-1} \underbrace{(\alpha_i + u_{it})}_{e_{it}}] = 0, \quad t = 2, 3, \ldots, T$$

  - Strictly exogenous or predetermined regressors:

$$E[\Delta \mathbf{x}_{it} \underbrace{(\alpha_i + u_{it})}_{=e_{it}}] = \mathbf{0}, \quad t = 1, 2, \ldots, T$$

  - Endogenous regressors:

$$E[\Delta \mathbf{x}_{i,t-1} \underbrace{(\alpha_i + u_{it})}_{=e_{it}}] = \mathbf{0}, \quad t = 2, 3, \ldots, T$$

- In combination with the moment conditions for the differenced model, further lags for the level model are redundant.

Sys-GMM estimation: stacked moment conditions

- Stacked moment conditions:

$$E[\mathbf{m}_i(\theta)] = E\left[\begin{pmatrix} \mathbf{Z}_i^{D\prime} \Delta \mathbf{u}_i \\ \mathbf{Z}_i^{L\prime} \mathbf{e}_i \end{pmatrix}\right] = \mathbf{0}$$

where $\mathbf{e}_i = (e_{i2}, e_{i3}, \dots, e_{iT})'$, and $\mathbf{Z}_i^L = (\mathbf{Z}_{yi}^L, \mathbf{Z}_{xi}^L)$, with *GMM-type* instruments

$$\mathbf{Z}_{yi}^L = \begin{array}{cc} \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \Delta y_{i1} & 0 & \cdots & 0 \\ 0 & \Delta y_{i2} & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \Delta y_{i,T-1} \end{pmatrix} & \begin{array}{l} \leftarrow \; t = 1 \\ \leftarrow \; t = 2 \\ \leftarrow \; t = 3 \\ \qquad \vdots \\ \leftarrow \; t = T \end{array} \end{array}$$

and similarly for $\mathbf{Z}_{xi}^L$.

Sys-GMM as level GMM

- Alternative formulation of the stacked moment conditions, recalling that $\Delta\mathbf{u}_i = \mathbf{D}_i\mathbf{u}_i = \mathbf{D}_i\mathbf{e}_i$:

$$E\left[\begin{pmatrix} \mathbf{Z}_i^{D\prime}\mathbf{D}_i\mathbf{e}_i \\ \mathbf{Z}_i^{L\prime}\mathbf{e}_i \end{pmatrix}\right] = E\left[\begin{pmatrix} \mathbf{Z}_i^{D\prime}\mathbf{D}_i \\ \mathbf{Z}_i^{L\prime} \end{pmatrix}\mathbf{e}_i\right] = E[\mathbf{Z}_i'\mathbf{e}_i] = \mathbf{0}$$

  where $\mathbf{Z}_i = (\tilde{\mathbf{Z}}_i^D, \mathbf{Z}_i^L)$ is a set of instruments for the level model with transformed instruments $\tilde{\mathbf{Z}}_i^D = \mathbf{D}_i'\mathbf{Z}_i^D$.
  - The sys-GMM estimator can be written as a *level GMM* estimator (Arellano and Bover, 1995).
  - Internally, this is how `xtdpdgmm` is implemented.

Sys-GMM estimation: optimal weighting matrix

- When u_{it} is serially uncorrelated and both u_{it} and α_i are homoskedastic, an optimal weighting matrix would be a function of the unknown variance ratio $\tau = \sigma_\alpha^2/\sigma_u^2$:

$$\mathbf{W}(\tau) = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_i'(\tau \boldsymbol{\iota}_T \boldsymbol{\iota}_T' + \mathbf{I}_T)\mathbf{Z}_i \right)^{-1}$$

  where $\boldsymbol{\iota}_T$ is a $T \times 1$ vector of ones and $\mathbf{I}_T$ is the $T \times T$ identity matrix.
  - Efficient one-step GMM estimation is infeasible, unless all moment conditions refer to the transformed model (because $\mathbf{D}_i \boldsymbol{\iota}_T = \mathbf{0}$) or τ is known. (A value for τ can be specified with the `wmatrix()` suboption `ratio(#)`).
- Optimal weighting matrix $\mathbf{W}(\hat{\boldsymbol{\theta}}) = (\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_i{}'\hat{\mathbf{e}}_i\hat{\mathbf{e}}_i'\mathbf{Z}_i)^{-1}$ requires initial consistent estimates.

Sys-GMM estimation: initial weighting matrix

- Candidates for an initial weighting matrix:
  - `xtdpdgmm` default option `wmatrix(unadjusted)` (Windmeijer, 2000), identical to initial two-stage least squares estimation:

$$\mathbf{W} = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{Z}_i \right)^{-1} = \left(\frac{1}{N} \sum_{i=1}^{N} \begin{pmatrix} \mathbf{Z}_i^{D\prime} \mathbf{D}_i \mathbf{D}_i' \mathbf{Z}_i^D & \mathbf{Z}_i^{D\prime} \mathbf{D}_i \mathbf{Z}_i^L \\ \mathbf{Z}_i^{L\prime} \mathbf{D}_i' \mathbf{Z}_i^D & \mathbf{Z}_i^{L\prime} \mathbf{Z}_i^L \end{pmatrix} \right)^{-1}$$

  - `xtdpdgmm` option `wmatrix(independent)` (Blundell, Bond, and Windmeijer, 2001):

$$\mathbf{W} = \left(\frac{1}{N} \sum_{i=1}^{N} \begin{pmatrix} \mathbf{Z}_i^{D\prime} \mathbf{D}_i \mathbf{D}_i' \mathbf{Z}_i^D & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_i^{L\prime} \mathbf{Z}_i^L \end{pmatrix} \right)^{-1}$$

  - `xtdpdgmm` option `wmatrix(separate)` (Arellano and Bover, 1995; Blundell and Bond, 1998):

$$\mathbf{W} = \left(\frac{1}{N} \sum_{i=1}^{N} \begin{pmatrix} \mathbf{Z}_i^{D\prime} \mathbf{Z}_i^D & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_i^{L\prime} \mathbf{Z}_i^L \end{pmatrix} \right)^{-1}$$

Two-step sys-GMM estimation in Stata

```
. xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) ///
> gmm(n, lag(1 1) diff model(level)) gmm(w k, lag(0 0) diff model(level)) two vce(r)

Generalized method of moments estimation

Fitting full model:
Step 1         f(b) =  .00285146
Step 2         f(b) =  .11568719

Group variable: id                           Number of obs         =        891
Time variable: year                          Number of groups      =        140

Moment conditions:     linear =       13     Obs per group:    min =          6
                    nonlinear =        0                       avg =   6.364286
                        total =       13                       max =          8

                                   (Std. Err. adjusted for 140 clusters in id)
------------------------------------------------------------------------------
             |              WC-Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .5117523   .1208484     4.23   0.000     .2748937    .7486109
             |
           w |  -1.323125   .2383451    -5.55   0.000    -1.790273    -.855977
           k |   .1931365   .0941343     2.05   0.040     .0086367    .3776363
       _cons |   4.698425   .7943584     5.91   0.000     3.141511    6.255339
------------------------------------------------------------------------------
```

(Continued on next page)

Two-step sys-GMM estimation in Stata

```
Instruments corresponding to the linear moment conditions:
 1, model(diff):
   L2.n L3.n L4.n
 2, model(diff):
   L1.w L2.w L3.w L1.k L2.k L3.k
 3, model(level):
   L1.D.n
 4, model(level):
   D.w D.k
 5, model(level):
   _cons

. estat serial, ar(1/3)

Arellano-Bond test for autocorrelation of the first-differenced residuals
H0: no autocorrelation of order 1:     z =   -3.3341   Prob > |z|  =    0.0009
H0: no autocorrelation of order 2:     z =   -1.2436   Prob > |z|  =    0.2136
H0: no autocorrelation of order 3:     z =   -0.1939   Prob > |z|  =    0.8462

. estat overid

Sargan-Hansen test of the overidentifying restrictions
H0: overidentifying restrictions are valid

2-step moment functions, 2-step weighting matrix       chi2(9)      =   16.1962
                                                       Prob > chi2 =    0.0629

2-step moment functions, 3-step weighting matrix       chi2(9)      =   13.8077
                                                       Prob > chi2 =    0.1293
```

Sys-GMM estimation: transformations

- The *global* option `model()` of the `xtdpdgmm` command sets the default model transformation for all instrument subsets, which is the level model unless specified otherwise.
  - The default set by this option can be overwritten for individual subsets of *GMM-type* and *standard* instruments with the suboption `model()`, e.g. `model(difference)` or `model(level)`.

```
. xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) ///
> gmm(n, lag(1 1) diff model(level)) gmm(w k, lag(0 0) diff model(level)) two vce(r)
(Output omitted)

. xtdpdgmm L(0/1).n w k, collapse gmm(n, lag(2 4) model(diff)) gmm(w k, lag(1 3) model(diff)) ///
> gmm(n, lag(1 1) diff) gmm(w k, lag(0 0) diff) two vce(r)
(Output omitted)
```

- The suboption `difference` of the `gmmiv()` and `iv()` options requests a first-difference transformation of the instruments (not the model).

Sys-GMM estimation: transformed instruments

- After the estimation with `xtdpdgmm`, the postestimation command `predict` with option `iv` generates the transformed instruments for the level model, $\mathbf{Z}_i = (\tilde{\mathbf{Z}}_i^D, \mathbf{Z}_i^L)$ (excluding the intercept), as new variables.
    - These new variables can be used subsequently to replicate the results (besides the Windmeijer correction of the standard errors) with Stata's `ivregress` command or the community-contributed `ivreg2` command (Baum, Schaffer, and Stillman, 2003, 2007).
    - This provides easy access to the additional options and postestimation statistics of these commands, e.g. the underidentification test based on the Kleibergen and Paap (2006) rank statistic reported by `ivreg2`.

Two-step sys-GMM estimation in Stata

```
. quietly predict iv*, iv

. ivregress gmm n (L.n w k = iv*), wmat(cluster id)

Instrumental variables (GMM) regression            Number of obs   =        891
                                                   Wald chi2(3)    =     485.45
                                                   Prob > chi2     =     0.0000
                                                   R-squared       =     0.8545
GMM weight matrix: Cluster (id)                    Root MSE        =     .51125

                                   (Std. Err. adjusted for 140 clusters in id)
------------------------------------------------------------------------------
             |               Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .5117523    .098918     5.17   0.000     .3178765    .7056281
             |
           w |  -1.323125   .2031404    -6.51   0.000    -1.721273   -.924977
           k |   .1931365   .0873607     2.21   0.027     .0219126    .3643604
       _cons |   4.698425   .6369462     7.38   0.000     3.450034    5.946817
------------------------------------------------------------------------------
Instrumented:  L.n w k
Instruments:   iv1 iv2 iv3 iv4 iv5 iv6 iv7 iv8 iv9 iv10 iv11 iv12

. estat overid

  Test of overidentifying restriction:

  Hansen's J chi2(9) = 16.1962 (p = 0.0629)
```

Two-step sys-GMM estimation in Stata

```
. ivreg2 n (L.n w k = iv*), gmm2s cluster(id)

2-Step GMM estimation
---------------------

Estimates efficient for arbitrary heteroskedasticity and clustering on id
Statistics robust to heteroskedasticity and clustering on id

Number of clusters (id) =          140                Number of obs =      891
                                                      F(  3,    139) =   230.77
                                                      Prob > F      =   0.0000
Total (centered) SS     =  1601.042507                Centered R2   =   0.8545
Total (uncentered) SS   =  2564.249196                Uncentered R2 =   0.9092
Residual SS             =  232.8868955                Root MSE      =    .5113

------------------------------------------------------------------------------
             |               Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .5117523   .0822341     6.22   0.000     .3505763    .6729282
             |
           w |  -1.323125   .1621898    -8.16   0.000    -1.641011   -1.005239
           k |   .1931365   .0660458     2.92   0.003     .0636892    .3225838
       _cons |   4.698425   .5321653     8.83   0.000     3.655401     5.74145
------------------------------------------------------------------------------
```

(Continued on next page)

Two-step sys-GMM estimation in Stata

```
Underidentification test (Kleibergen-Paap rk LM statistic):              30.312
                                                   Chi-sq(10) P-val =    0.0008
------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):                  0.376
                         (Kleibergen-Paap rk Wald F statistic):            5.128
Stock-Yogo weak ID test critical values:  5% maximal IV relative bias     17.80
                                         10% maximal IV relative bias     10.01
                                         20% maximal IV relative bias      5.90
                                         30% maximal IV relative bias      4.42
Source: Stock-Yogo (2005).  Reproduced by permission.
NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.
------------------------------------------------------------------------------
Hansen J statistic (overidentification test of all instruments):         16.196
                                                   Chi-sq(9) P-val =     0.0629
------------------------------------------------------------------------------
Instrumented:         L.n w k
Excluded instruments: iv1 iv2 iv3 iv4 iv5 iv6 iv7 iv8 iv9 iv10 iv11 iv12
------------------------------------------------------------------------------
```

Underidentification tests

- While it is standard practice to test for overidentification, the potential problem of underidentification is largely ignored in the empirical practice of estimating dynamic panel data models.
- Underidentification tests based on (robust) versions of the Cragg and Donald (1993) and Kleibergen and Paap (2006) statistics test the null hypothesis $H_0 : \mathrm{rk}(E[\mathbf{Z}_i'\mathbf{X}_i]) = K - 1$, i.e. the model is underidentified, versus the alternative hypothesis $H_1 : \mathrm{rk}(E[\mathbf{Z}_i'\mathbf{X}_i]) = K$, where $\mathbf{X}_i$ is the matrix of regressors (including the lagged dependent variable).

Underidentification tests

- Windmeijer (2018) highlights that the underidentification tests are overidentification tests in an auxiliary regression of any endogenous variable on the remaining regressors, e.g.

$$y_{i,t-1} = \sum_{j=2}^{q_y} \varphi_j y_{i,t-j} + \sum_{j=0}^{q_x} \mathbf{x}'_{i,t-j} \psi_j + v_{it}$$

  using the same instruments $\mathbf{Z}_i$ as before.
- Windmeijer (2018) shows that a robust Cragg-Donald statistic is the Hansen J-statistic based on the continuously updating GMM estimator, and that the robust Kleibergen-Paap statistic is a J-statistic based on the limited information maximum likelihood (LIML) estimator. Both are invariant to the choice of the left-hand side variable in the auxiliary regression.

Underidentification tests

- Sanderson and Windmeijer (2016) use the above auxiliary regressions to compute weak-identification tests. Their robust version is the Hansen J-statistic based on the two-step GMM estimator. As it is not invariant to the choice of the left-hand side variable, it can inform about the particular endogenous variables that are poorly predicted by the instruments (Windmeijer, 2018).
- The forthcoming `underid` command by Mark Schaffer and Frank Windmeijer presents both overidentification and underidentification statistics after internally reestimating the model with the `ivreg2` command, using the instruments generated by `xtdpdgmm`. From the users' perspective, `underid` works as a postestimation command for `xtdpdgmm`.

Underidentification tests in Stata

```
. quietly xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) ///
> gmm(n, lag(1 1) diff model(level)) gmm(w k, lag(0 0) diff model(level)) two vce(r)

. underid, overid jgmm2s
Number of obs:      891
Number of panels:   140
Dep var:            n
Endog Xs (3):       L.n w k
Exog Xs (1):        _cons
Excl IVs (12):      __alliv_1 __alliv_2 __alliv_3 __alliv_4 __alliv_5 __alliv_6
                    __alliv_7 __alliv_8 __alliv_9 __alliv_10 __alliv_11
                    __alliv_12

Overidentification test: 2-step-GMM-based (LM version)
  Test statistic robust to heteroskedasticity and clustering on id
j=   16.20  Chi-sq(  9) p-value=0.0629

. underid, overid underid jcue noreport

Overidentification test: Cragg-Donald robust CUE-based (LM version)
  Test statistic robust to heteroskedasticity and clustering on id
j=    8.17  Chi-sq(  9) p-value=0.5168

Underidentification test: Cragg-Donald robust CUE-based (LM version)
  Test statistic robust to heteroskedasticity and clustering on id
j=   26.92  Chi-sq( 10) p-value=0.0027
```

Underidentification tests in Stata

```
. underid, overid underid kp sw noreport

Overidentification test: Kleibergen-Paap robust LIML-based (LM version)
  Test statistic robust to heteroskedasticity and clustering on id
j=    9.98  Chi-sq(  9) p-value=0.3520

Underidentification test: Kleibergen-Paap robust LIML-based (LM version)
  Test statistic robust to heteroskedasticity and clustering on id
j=   30.31  Chi-sq( 10) p-value=0.0008

2-step GMM J underidentification stats by regressor:
j=   30.00  Chi-sq( 10) p-value=0.0009 L.n
j=   29.07  Chi-sq( 10) p-value=0.0012 w
j=   26.01  Chi-sq( 10) p-value=0.0037 k
```

- The tests would raise concerns if the overidentification tests were rejected or the underidentification tests were not rejected.
  - Note that the robust Cragg-Donald and Kleibergen-Paap overidentification tests have no power to detect a violation if the model is underidentified (Windmeijer, 2018).

Incremental overidentification tests

- Under the assumption that the diff-GMM estimator is correctly specified, we can test the validity of the additional moment conditions for the level model.
- Incremental overidentification tests / *difference Sargan-Hansen* tests are asymptotically $\chi^2(df_f - df_r)$ distributed, where df_f and df_r are the degrees of freedom of the *full-model* and the *reduced-model* overidentification tests, respectively (Eichenbaum, Hansen, and Singleton, 1988), e.g.:

$$J(\hat{\hat{\theta}}_f, \mathbf{W}(\hat{\theta}_f)) - J(\hat{\hat{\theta}}_r, \mathbf{W}(\hat{\theta}_r))$$

  - Incremental overidentifications tests are only meaningful if the reduced model already passed the overidentification test.

Incremental overidentification tests in Stata

- The `xtdpdgmm` postestimation command `estat overid` allows to compute the difference of two nested overidentification test statistics.

```
. quietly xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) nocons two ///
> vce(r)

. estimates store diff

. quietly xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) ///
> gmm(n, lag(1 1) diff model(level)) gmm(w k, lag(0 0) diff model(level)) two vce(r)

. estat overid diff

Sargan-Hansen difference test of the overidentifying restrictions
H0: additional overidentifying restrictions are valid

2-step moment functions, 2-step weighting matrix       chi2(3)      =   11.2420
                                                       Prob > chi2 =    0.0105

2-step moment functions, 3-step weighting matrix       chi2(3)      =    9.2942
                                                       Prob > chi2 =    0.0256
```

- The incremental overidentification test rejects the validity of the additional moment conditions for the level model.

Incremental overidentification tests

- In finite samples, the incremental overidentification test statistic can become negative because $\mathbf{W}(\hat{\theta}_f)$ and $\mathbf{W}(\hat{\theta}_r)$ are estimated separately.
- As an alternative that is guaranteed to be nonnegative, the relevant partition of the weighting matrix from the full model can be used to evaluate the test statistic for the reduced model (Newey, 1985):

$$J(\hat{\hat{\theta}}_f, \mathbf{W}(\hat{\theta}_f)) - J(\hat{\hat{\theta}}_r, \mathbf{W}(\hat{\theta}_f))$$

  - `xtdpdgmm` specified with option `overid` computes incremental overidentification tests for each set of `gmmiv()` or `iv()` instruments, and jointly for all moment conditions refering to the same model transformation.
  - The postestimation command `estat overid` displays the incremental tests when called with option `difference`.

Incremental overidentification tests in Stata

```
. xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) ///
> gmm(n, lag(1 1) diff model(level)) gmm(w k, lag(0 0) diff model(level)) two vce(r) overid

Generalized method of moments estimation

Fitting full model:
Step 1         f(b) =  .00285146
Step 2         f(b) =  .11568719

Fitting reduced model 1:
Step 1         f(b) =  .10476123

Fitting reduced model 2:
Step 1         f(b) =  .02873833

Fitting reduced model 3:
Step 1         f(b) =   .1131458

Fitting reduced model 4:
Step 1         f(b) =  .08632894

Fitting no-diff model:
Step 1         f(b) =  8.476e-19

Fitting no-level model:
Step 1         f(b) =  .05779984
(Some output omitted)
(Continued on next page)
```

Incremental overidentification tests in Stata

```
Instruments corresponding to the linear moment conditions:
 1, model(diff):
   L2.n L3.n L4.n
 2, model(diff):
   L1.w L2.w L3.w L1.k L2.k L3.k
 3, model(level):
   L1.D.n
 4, model(level):
   D.w D.k
 5, model(level):
   _cons

. estat overid, difference

Sargan-Hansen (difference) test of the overidentifying restrictions
H0: (additional) overidentifying restrictions are valid

2-step weighting matrix from full model

                  | Excluding                    | Difference
Moment conditions |       chi2     df          p |       chi2     df          p
------------------+------------------------------+-----------------------------
   1, model(diff) |    14.6666      6     0.0230 |     1.5296      3     0.6754
   2, model(diff) |     4.0234      3     0.2590 |    12.1728      6     0.0582
  3, model(level) |    15.8404      8     0.0447 |     0.3558      1     0.5509
  4, model(level) |    12.0861      7     0.0978 |     4.1102      2     0.1281
      model(diff) |     0.0000      0          . |    16.1962      9     0.0629
     model(level) |     8.0920      6     0.2314 |     8.1042      3     0.0439
```

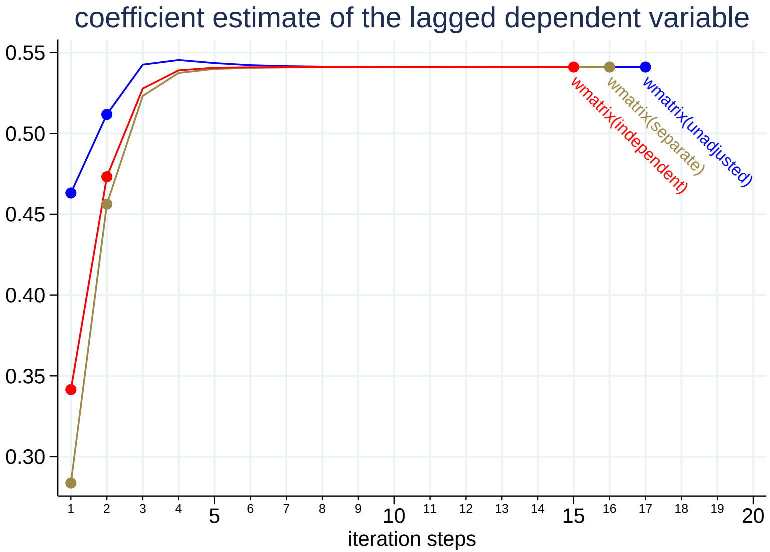
Iterated GMM estimation

- While the two-step estimator is asymptotically efficient (for a given set of instruments), in finite samples the estimation of the optimal weighting matrix might be sensitive to the chosen initial weighting matrix.
  - The resulting lack of robustness of the coefficient estimates and the overidentification test results to the choice of **W** has the undesired consequence that empiricists might be tempted to select the "most favorable" results.
- Hansen, Heaton, and Yaron (1996) suggest to use an iterated GMM estimator that updates the weighting matrix and coefficient estimates until convergence.
  - The iterated GMM estimator removes the arbitrariness in the choice of the initial weighting matrix (Hansen and Lee, 2019).
  - Similar to Stata's `gmm` or `ivregress` command, `xtdpdgmm` provides the option `igmm` as alternatives to `onestep` and `twostep`.

Iterated sys-GMM estimation in Stata

```
. xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) ///
> gmm(n, lag(1 1) diff model(level)) gmm(w k, lag(0 0) diff model(level)) igmm vce(r) nofootnote

Generalized method of moments estimation

Fitting full model:
Steps
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
.................                                  17

Group variable: id                           Number of obs         =        891
Time variable: year                          Number of groups      =        140

Moment conditions:     linear =        13    Obs per group:    min =          6
                    nonlinear =         0                      avg =   6.364286
                        total =        13                      max =          8

                                  (Std. Err. adjusted for 140 clusters in id)
------------------------------------------------------------------------------
             |              WC-Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |    .541044   .1265822     4.27   0.000     .2929474    .7891406
             |
           w |  -1.527984    .304707    -5.01   0.000    -2.125199   -.9307697
           k |   .1075032   .1115814     0.96   0.335    -.1111923    .3261986
       _cons |   5.275027   .9736502     5.42   0.000     3.366707    7.183346
------------------------------------------------------------------------------
```

Iterated sys-GMM estimation: initial weighting matrices



Continuously updated GMM estimation

- As an alternative to the iterated GMM estimator, Hansen, Heaton, and Yaron (1996) also suggest a continuously updated GMM estimator that numerically minimizes

$$\tilde{\boldsymbol{\theta}} = \arg\min_{\mathbf{b}} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{m}_i(\mathbf{b}) \right)' \mathbf{W}(\mathbf{b}) \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{m}_i(\mathbf{b}) \right)$$

  where the optimal weighting matrix $\mathbf{W}(\tilde{\boldsymbol{\theta}})$ is obtained directly as part of the minimization process.
  - This estimator is not currently implemented in `xtdpdgmm` but the `ivreg2` command can be used with the instruments previously generated from `xtdpdgmm`.

Continuously updated sys-GMM estimation in Stata

```
. ivreg2 n (L.n w k = iv*), cue cluster(id)
Iteration 0:   f(p) =  24.858945  (not concave)
(Some output omitted)
Iteration 21:  f(p) =  8.2335574

CUE estimation
--------------

Estimates efficient for arbitrary heteroskedasticity and clustering on id
Statistics robust to heteroskedasticity and clustering on id
(Some output omitted)
------------------------------------------------------------------------------
             |               Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .5239428   .1138624     4.60   0.000     .3007766    .7471089
             |
           w |  -2.025771   .2810169    -7.21   0.000    -2.576555   -1.474988
           k |  -.0193789   .1221278    -0.16   0.874    -.2587449    .2199872
       _cons |   6.781101   .8346986     8.12   0.000     5.145122     8.41708
------------------------------------------------------------------------------
(Some output omitted)
Hansen J statistic (overidentification test of all instruments):         8.234
                                                   Chi-sq(9) P-val =    0.5108
------------------------------------------------------------------------------
Instrumented:         L.n w k
Excluded instruments: iv1 iv2 iv3 iv4 iv5 iv6 iv7 iv8 iv9 iv10 iv11 iv12
------------------------------------------------------------------------------
```

Nonlinear moment conditions: no serial correlation

- Absence of serial correlation in u_{it} is a necessary condition for the validity of $y_{i,t-2}, y_{i,t-3}, \dots$ as instruments for the first-differenced model.
- Ahn and Schmidt (1995) suggest to exploit additional nonlinear (quadratic) moment conditions:

$$E[\underbrace{(\alpha_i + u_{iT})}_{e_{iT}} \Delta u_{it}] = 0, \quad t = 1, 2, \dots, T-1$$

  - These nonlinear moment conditions are redundant when added to the sys-GMM moment conditions (Blundell and Bond, 1998) but improve efficiency when added to the diff-GMM moment conditions. Furthermore, they may provide identification when the diff-GMM estimator does not (Gørgens, Han, and Xue, 2019).
  - The nonlinear moment conditions remain valid even when the sys-GMM moment conditions for the level model are not.

Nonlinear moment conditions: no serial correlation

- `xtdpdgmm` with option `nl(noserial)` adds these moment conditions. They can be collapsed into the single moment condition $E[e_{iT} \sum_{t=1}^{T} \Delta u_{it}] = 0$ with *global* option `collapse` or suboption `[no]collapse`, similar to other instruments.
  - Due to the presence of the level error term e_{iT}, an intercept should generally be included in the estimation even if all other moment conditions refer to the first-differenced model.
- While GMM estimators with only linear moment conditions have a closed-form solution, this is no longer the case with nonlinear moment conditions.
  - `xtdpdgmm` minimizes the GMM criterion function numerically with Stata's Gauss-Newton algorithm.
- A feasible efficient one-step GMM estimator does not exist.
  - `xtdpdgmm` uses a block-diagonal initial weighting matrix.

Estimation with nonlinear moment conditions in Stata

```
. xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) nl(noserial) igmm vce(r

Generalized method of moments estimation

Fitting full model:

Steps
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..........                                              10

Group variable: id                          Number of obs         =        891
Time variable: year                         Number of groups      =        140

Moment conditions:     linear =        10   Obs per group:    min =          6
                    nonlinear =         1                     avg =   6.364286
                        total =        11                     max =          8

                                   (Std. Err. adjusted for 140 clusters in id)
------------------------------------------------------------------------------
             |              WC-Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .5206104   .1226228     4.25   0.000     .2802741    .7609466
             |
           w |  -1.700205    .255932    -6.64   0.000    -2.201823   -1.198588
           k |   .0508781    .109654     0.46   0.643    -.1640397     .265796
       _cons |   5.824618   .8009101     7.27   0.000     4.254863    7.394373
------------------------------------------------------------------------------
```

(Continued on next page)

Estimation with nonlinear moment conditions in Stata

```
Instruments corresponding to the linear moment conditions:
 1, model(diff):
   L2.n L3.n L4.n
 2, model(diff):
   L1.w L2.w L3.w L1.k L2.k L3.k
 3, model(level):
   _cons

. estat serial, ar(1/3)

Arellano-Bond test for autocorrelation of the first-differenced residuals
H0: no autocorrelation of order 1:     z =   -3.0815   Prob > |z|  =    0.0021
H0: no autocorrelation of order 2:     z =   -1.1802   Prob > |z|  =    0.2379
H0: no autocorrelation of order 3:     z =   -0.1635   Prob > |z|  =    0.8701

. estat overid

Sargan-Hansen test of the overidentifying restrictions
H0: overidentifying restrictions are valid

10-step moment functions, 10-step weighting matrix     chi2(7)     =    6.2103
                                                       Prob > chi2 =    0.5154

10-step moment functions, 11-step weighting matrix     chi2(7)     =    6.2103
                                                       Prob > chi2 =    0.5154
```

Nonlinear moment conditions: homoskedasticity

- Under the assumption of homoskedasticity, the previous nonlinear moment conditions can be replaced by

$$E[\bar{e}_i \Delta u_{it}] = 0, \quad t = 2, 3, \dots, T$$

and the additional linear moment conditions

$$E[y_{i,t-2}\Delta u_{i,t-1} - y_{i,t-1}\Delta u_{it}] = 0, \quad t = 3, 4, \dots, T$$

- `xtdpdgmm` with option `nl(iid)` implements a variation of these moment conditions where $\bar{e}_i = \frac{1}{T}\sum_{t=1}^{T} e_{it}$ is multiplied by the factor $\sqrt{T}$, unless *global* option `norescale` or suboption `[no]rescale` is specified. Collapsing of both nonlinear and linear moment conditions is possible as before.

Generalized Hausman test

- When the homoskedasticity assumption is satisfied, the GMM estimator using the additional moment conditions is more efficient. Otherwise, it becomes inconsistent.
- This motivates a generalized Hausman (1978) test for the statistical difference between the two estimators. The test statistic is asymptotically $\chi^2(df)$ distributed with $df = \min(df_f - df_r, K)$ degrees of freedom.
  - `xtdpdgmm` provides the postestimation command `estat hausman` to carry out the generalized Hausman test. A robust estimate of the covariance matrix is used that does not require one of the estimators to be fully efficient (White, 1982).
- When the number of additional overidentifying restrictions, $df_f - df_r$, is not larger than the number of contrasted coefficients, K, then the generalized Hausman test is asymptotically equivalent to incremental Sargan-Hansen tests.

Generalized Hausman test in Stata

```
. xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) nl(iid) igmm vce(r)

Generalized method of moments estimation

Fitting full model:

Steps
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
.........                                             9

Group variable: id                           Number of obs      =       891
Time variable: year                          Number of groups   =       140

Moment conditions:     linear =         11   Obs per group:    min =         6
                    nonlinear =          1                     avg =  6.364286
                        total =         12                     max =         8

                                   (Std. Err. adjusted for 140 clusters in id)
------------------------------------------------------------------------------
             |              WC-Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |    .543599   .1347044     4.04   0.000     .2795833    .8076148
             |
           w |  -2.011612   .4641684    -4.33   0.000    -2.921365   -1.101859
           k |  -.1157727   .1900186    -0.61   0.542    -.4882024     .256657
       _cons |   6.720082   1.339408     5.02   0.000     4.094891    9.345273
------------------------------------------------------------------------------
```

(Continued on next page)

Generalized Hausman test

```
Instruments corresponding to the linear moment conditions:
 1, model(iid):
   L.n
 2, model(diff):
   L2.n L3.n L4.n
 3, model(diff):
   L1.w L2.w L3.w L1.k L2.k L3.k
 4, model(level):
   _cons

. estimates store iid

. quietly xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) ///
> nl(noserial) igmm vce(r)

. estat hausman iid

Generalized Hausman test                               chi2(1)      =     7.1129
H0: coefficients do not systematically differ          Prob > chi2 =     0.0077
```

- The generalized Hausman test rejects the additional overidentifying restriction from the homoskedasticity assumption.

Forward-orthogonal deviations: transformation

- Assuming no serial correlation in u_{it}, the first-difference transformation creates first-order serial correlation in Δu_{it}.
- Arellano and Bover (1995) propose to use forward-orthogonal deviations (FOD) instead that remain serially uncorrelated:

$$\tilde{\Delta}_t y_{it} = \sum_{j=1}^{q_y} \lambda_j \tilde{\Delta}_t y_{i,t-j} + \sum_{j=0}^{q_x} \tilde{\Delta}_t \mathbf{x}'_{i,t-j} \beta_j + \underbrace{\tilde{\Delta}_t u_{it}}_{=\tilde{\Delta}_t e_{it}}$$

where $\tilde{\Delta}_t u_{it} = \sqrt{\frac{T-t+1}{T-t}} \left(u_{it} - \frac{1}{T-t+1} \sum_{s=0}^{T-t} u_{i,t+s} \right)$, with $Corr(\tilde{\Delta}_t u_{it}, \tilde{\Delta}_t u_{i,t-1}) = 0$.

  - By subtracting the forward mean, the unit-specific effects α_i (and all other time-invariant variables) are again eliminated.
  - The factor $\sqrt{\frac{T-t+1}{T-t}}$ ensures that the variance remains unchanged if u_{it} is homoskedastic. It can be suppressed with option `norescale`.

Forward-orthogonal deviations: moment conditions

- Moment conditions for the FOD-transformed model:
  - Lagged dependent variable:

$$E[y_{i,t-s}\tilde{\Delta}_t u_{it}] = 0, \quad s = 1, 2, \ldots, t$$

  - Strictly exogenous regressors:

$$E[\mathbf{x}_{i,t-s}\tilde{\Delta}_t u_{it}] = \mathbf{0}, \quad t - s = 0, 1, \ldots, T$$

  - Predetermined regressors:

$$E[\mathbf{x}_{i,t-s}\tilde{\Delta}_t u_{it}] = \mathbf{0}, \quad s = 0, 1, \ldots, t$$

  - Endogenous regressors:

$$E[\mathbf{x}_{i,t-s}\tilde{\Delta}_t u_{it}] = \mathbf{0}, \quad s = 1, 2, \ldots, t$$

with $t = s, \ldots, T - 1$.

Forward-orthogonal deviations: transformation matrix

- Stacked moment conditions:

$$E[\mathbf{m}_i(\theta)] = E\left[\mathbf{Z}_i^{FOD\prime}\mathbf{H}_i\mathbf{u}_i\right] = \mathbf{0}$$

where $\mathbf{H}_i\mathbf{u}_i = (\tilde{\Delta}_1 u_{i1}, \tilde{\Delta}_2 u_{i2}, \dots, \tilde{\Delta}_{T-1} u_{i,T-1})'$ with $T-1 \times T$ FOD-transformation matrix

$$\mathbf{H}_i = \text{diag}\left(\sqrt{\frac{T}{T-1}}, \sqrt{\frac{T-1}{T-2}}, \dots, \sqrt{\frac{2}{1}}\right) \times \begin{pmatrix} \frac{T-1}{T} & -\frac{1}{T} & -\frac{1}{T} & \cdots & -\frac{1}{T} & -\frac{1}{T} \\ 0 & \frac{T-2}{T-1} & -\frac{1}{T-1} & \cdots & -\frac{1}{T-1} & -\frac{1}{T-1} \\ & & & \ddots & & \\ 0 & 0 & 0 & \cdots & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

- With `xtdpdgmm`, the option `model(fodev)` creates instruments for the FOD-transformed model.

Forward-orthogonal deviations versus first differences

- With balanced panel data, the diff-GMM estimator and the FOD-GMM estimator are identical if the default weighting matrix and all available *GMM-type* instruments (non-curtailed and non-collapsed) are used (Arellano and Bover, 1995):

```
. preserve

. keep if year > 1977 & year < 1983
(331 observations deleted)

. xtdpdgmm L(0/1).n w k, model(diff) gmm(n, lag(2 .)) gmm(w k, lag(1 .)) nocons vce(r)
(Output omitted)

. xtdpdgmm L(0/1).n w k, model(fodev) gmm(n, lag(1 .)) gmm(w k, lag(0 .)) nocons vce(r)
(Output omitted)

. restore
```

- When the panel data set is unbalanced with interior gaps, the FOD-GMM estimator retains more information than the diff-GMM estimator.

Forward-orthogonal deviations: other Stata commands

- In contrast to xtdpdgmm, the FOD implementation in xtabond2 is problematic. xtabond2 (and likewise xtdpd) internally shifts the FOD model by one time period.
  - For example, the first lag of an instrument must be specified as if it was the second lag.

```
. xtdpdgmm L(0/1).n w k, model(fodev) collapse gmm(n, lag(1 3)) gmm(w k, lag(0 2)) nocons vce(r)
(Some output omitted)
-------------+----------------------------------------------------------------
           n |
         L1. |   .4432348   .1368918     3.24   0.001     .1749319    .7115377
             |
           w |   -1.92711   .3610225    -5.34   0.000    -2.634701   -1.219518
           k |   .0511631   .1908062     0.27   0.789    -.3228102    .4251363
------------------------------------------------------------------------------
(Some output omitted)

. xtabond2 L(0/1).n w k, orthogonal gmm(n, lag(2 4) collapse) gmm(w k, lag(1 3) collapse) nolevel r
(Some output omitted)
-------------+----------------------------------------------------------------
           n |
         L1. |   .4432348   .1368918     3.24   0.001     .1749319    .7115377
             |
           w |   -1.92711   .3610225    -5.34   0.000    -2.634701   -1.219518
           k |   .0511631   .1908062     0.27   0.789    -.3228102    .4251363
------------------------------------------------------------------------------
(Some output omitted)
```

Forward-orthogonal deviations: other Stata commands

- The `xtabond2` and `xtdpd` implementations lead to incorrect results when combined with *standard* instruments.
  - The following two specifications are supposed to be equivalent to the previous two but the second is not. Bug!

```
. xtdpdgmm L(0/1).n w k, model(fodev) iv(n, lag(1 3)) iv(w k, lag(0 2)) nocons vce(r)
(Some output omitted)
-------------+----------------------------------------------------------------
           n |
         L1. |   .4432348   .1368918     3.24   0.001     .1749319    .7115377
             |
           w |   -1.92711   .3610225    -5.34   0.000    -2.634701   -1.219518
           k |   .0511631   .1908062     0.27   0.789    -.3228102    .4251363
------------------------------------------------------------------------------
(Some output omitted)

. xtabond2 L(0/1).n w k, orthogonal iv(L(2/4).n, passthru mz) iv(L(1/3).(w k), passthru mz) nolevel r
(Some output omitted)
-------------+----------------------------------------------------------------
           n |
         L1. |   .4254774   .1369818     3.11   0.002     .1569979    .6939569
             |
           w |  -1.860978   .3532973    -5.27   0.000    -2.553428   -1.168528
           k |   .1301844   .1844341     0.71   0.480    -.2312997    .4916686
------------------------------------------------------------------------------
(Some output omitted)
```

Double-filter GMM estimation

- For models with predetermined variables (and motivated for samples with large T), Hayakawa, Qi, and Breitung (2019) suggest a *double-filter* IV / GMM estimator that combines forward-orthogonal deviations of the error term with backward-orthogonal deviations of the instruments.
- While taking lags and differencing are interchangeable time series operations, the same is not true for lags and backward-orthogonal deviations.
  - The option `iv(L.n, bodev model(fodev))` takes backward-orthogonal deviations of the lagged dependent variable, while `iv(n, bodev lags(1 1) model(fodev))` takes the lag of the backward-orthogonally deviated dependent variable. Hayakawa, Qi, and Breitung (2019) suggest the former.

Double-filter GMM estimation in Stata

```
. xtdpdgmm L(0/1).n w k, model(fodev) collapse gmm(L.n, bodev lag(0 2)) gmm(w k, bodev lag(0 2)) ///
> nocons igmm vce(r) noheader

Generalized method of moments estimation

Fitting full model:
Steps
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
...............                                      15

                                   (Std. Err. adjusted for 140 clusters in id)
------------------------------------------------------------------------------
             |              WC-Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |    .205428   .1676214     1.23   0.220    -.1231038    .5339598
             |
           w |  -.8464892   .3586161    -2.36   0.018    -1.549364   -.1436145
           k |   .4751495   .2757519     1.72   0.085    -.0653143    1.015613
------------------------------------------------------------------------------
Instruments corresponding to the linear moment conditions:
 1, model(fodev):
   B.L.n L1.B.L.n L2.B.L.n
 2, model(fodev):
   B.w L1.B.w L2.B.w B.k L1.B.k L2.B.k
```

Time effects

- To account for global shocks, it is common practice to include a set of time dummies in the regression model:

$$y_{it} = \sum_{j=1}^{q_y} \lambda_j y_{i,t-j} + \sum_{j=0}^{q_x} \mathbf{x}'_{i,t-j} \boldsymbol{\beta}_j + \delta_t + \underbrace{\alpha_i + u_{it}}_{=e_{it}}$$

  - Without loss of generality, time dummies δ_t can be treated as strictly exogenous and uncorrelated with the unit-specific effects α_i. Hence, time dummies can be instrumented by themselves.
  - When the model contains an intercept, only $T - 1$ time dummies can be included to avoid the *dummy trap*.

Time effects: instruments

- With balanced panel data, instrumenting the time dummies in the level model or the transformed model yields identical estimates (with the default initial weighting matrix):

```
. preserve

. keep if year > 1977 & year < 1983
(331 observations deleted)

. xtdpdgmm L(0/1).n w k yr1980-yr1982, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) ///
> iv(yr1980-yr1982, model(level)) two vce(r)
(Output omitted)

. xtdpdgmm L(0/1).n w k yr1980-yr1982, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) ///
> iv(yr1980-yr1982, diff) two vce(r)
(Output omitted)

. restore
```

- Even in unbalanced panel data sets, instruments for time dummies should not be specified for both the level and the transformed model because one of them is asymptotically redundant.

Time effects: multicollinear instruments

- xtdpdgmm and xtabond2 differ in the way they treat perfectly collinear instruments which might lead to different estimates (if another than the default initial weighting matrix is used).
  - xtdpdgmm detects and removes perfectly collinear instruments from the transformed level instruments $\mathbf{Z}_i = (\tilde{\mathbf{Z}}_i^D, \mathbf{Z}_i^L)$, while xtabond2 does not remove them and effectively only detects perfect collinearity separately within each group of instruments $\mathbf{Z}_i^D$ and $\mathbf{Z}_i^L$ (and likewise with the FOD transformation).
  - As a consequence, xtabond2 might report a number of instruments that is too large and hence also too many degrees of freedom for the overidentification tests. The reported p-values in this case are too large.

Time effects: multicollinear instruments

```
. preserve

. keep if year > 1977 & year < 1983
(331 observations deleted)

. xtdpdgmm L(0/1).n w k yr1980-yr1982, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) ///
> iv(yr1980-yr1982, diff) iv(yr1980-yr1982, model(level)) two vce(r)
(Output omitted)

. xtabond2 L(0/1).n w k yr1980-yr1982, gmm(n, lag(2 4) collapse eq(diff)) ///
> gmm(w k, lag(1 3) collapse eq(diff)) iv(yr1980-yr1982, eq(diff)) iv(yr1980-yr1982, eq(level)) two r
(Output omitted)

. xtabond2 L(0/1).n w k yr1980-yr1982, gmm(n, lag(2 4) collapse eq(diff)) ///
> gmm(w k, lag(1 3) collapse eq(diff)) iv(yr1980-yr1982, eq(diff)) iv(yr1980-yr1982, eq(level)) h(1) ///
> two r
(Output omitted)

. restore
```

- With the default weighting matrix, the first two specifications correctly detect the perfect collinearity among the instruments for the time dummies. The last specification with weighting matrix `h(1)` reports 3 instruments too many.

Time effects: other Stata commands

- When time dummies (or other variables) are specified with the factor variable notation and some of them are omitted due to perfect collinearity, `xtabond2` reports too few degrees of freedom for the overidentification tests. The reported *p*-values in this case are too small. Bug!

```
. quietly xtdpdgmm L(0/1).n w k yr1978-yr1984, model(diff) collapse gmm(n, lag(2 4)) ///
> gmm(w k, lag(1 3)) iv(yr1978-yr1984, model(level)) two vce(r)

. estat overid
(Some output omitted)
2-step moment functions, 2-step weighting matrix       chi2(6)     =    8.8841
                                                       Prob > chi2 =    0.1802
(Some output omitted)

. xtabond2 L(0/1).n w k yr1978-yr1984, gmm(n, lag(2 4) collapse eq(diff)) ///
> gmm(w k, lag(1 3) collapse eq(diff)) iv(yr1978-yr1984, eq(level)) two r
(Some output omitted)
Hansen test of overid. restrictions: chi2(6)    =   8.88  Prob > chi2 =  0.180
(Some output omitted)

. xtabond2 L(0/1).n w k i.year, gmm(n, lag(2 4) collapse eq(diff)) ///
> gmm(w k, lag(1 3) collapse eq(diff)) iv(i.year, eq(level)) two r
(Some output omitted)
Hansen test of overid. restrictions: chi2(4)    =   8.88  Prob > chi2 =  0.064
(Some output omitted)
```

Time effects: other Stata commands

- Stata's xtdpd command (and xtabond and xtdpdsys) drops one time dummy too many. Bug!

```
. xtdpd L(0/1).n w k yr1978-yr1984, dgmm(n, lag(2 4)) dgmm(w k, lag(1 3)) liv(yr1978-yr1984) two vce(r)
note: D.yr1984 dropped because of collinearity
(Some output omitted)
------------------------------------------------------------------------------
             |              WC-Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .5362071   .1327262     4.04   0.000     .2760684    .7963458
             |
           w |  -.7354218   .1342332    -5.48   0.000    -.9985139   -.4723296
           k |   .4675843   .0979644     4.77   0.000     .2755775     .659591
      yr1978 |  -.0304008   .0149698    -2.03   0.042    -.0597409   -.0010606
      yr1979 |  -.0444556   .0191132    -2.33   0.020    -.0819168   -.0069944
      yr1980 |  -.0650701   .0199986    -3.25   0.001    -.1042666   -.0258737
      yr1981 |  -.0944965   .0204774    -4.61   0.000    -.1346314   -.0543615
      yr1982 |  -.0389697   .0192286    -2.03   0.043     -.076657   -.0012824
      yr1983 |   .0037684   .0225635     0.17   0.867    -.0404553    .0479921
       _cons |   3.030333   .5184783     5.84   0.000     2.014134    4.046532
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/4).n L(1/3).w L(1/3).k
Instruments for level equation
        Standard: yr1978 yr1979 yr1980 yr1981 yr1982 yr1983 yr1984 _cons
```

GMM estimation with time effects in Stata

- xtdpdgmm has the option teffects that automatically adds the correct number of time dummies and corresponding instruments:

```
. xtdpdgmm L(0/1).n w k, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) nl(noserial) ///
> teffects igmm vce(r)

Generalized method of moments estimation

Fitting full model:
Steps
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
...................................                  35

Group variable: id                            Number of obs         =        891
Time variable: year                           Number of groups      =        140

Moment conditions:     linear =       17      Obs per group:    min =          6
                    nonlinear =        1                        avg =   6.364286
                        total =       18                        max =          8

                                   (Std. Err. adjusted for 140 clusters in id)
(Continued on next page)
```

GMM estimation with time effects in Stata

```
------------------------------------------------------------------------------
             |              WC-Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |    .715963   .2630756     2.72   0.006     .2003442    1.231582
             |
           w |  -.7645527   .6235711    -1.23   0.220     -1.98673    .4576242
           k |   .4043948    .270444     1.50   0.135    -.1256657    .9344553
             |
        year |
        1978 |  -.0656579   .0317356    -2.07   0.039    -.1278586   -.0034572
        1979 |  -.0825628   .0346171    -2.39   0.017    -.1504111   -.0147145
        1980 |  -.1035026   .0263053    -3.93   0.000      -.15506   -.0519452
        1981 |  -.1335986   .0313492    -4.26   0.000    -.1950419   -.0721553
        1982 |  -.0661445   .0574973    -1.15   0.250    -.1788372    .0465482
        1983 |   .0033487   .0685548     0.05   0.961    -.1310163    .1377137
        1984 |   .0538893   .1010754     0.53   0.594    -.1442148    .2519933
             |
       _cons |   2.932618   2.345137     1.25   0.211    -1.663767    7.529002
------------------------------------------------------------------------------
Instruments corresponding to the linear moment conditions:
 1, model(diff):
   L2.n L3.n L4.n
 2, model(diff):
   L1.w L2.w L3.w L1.k L2.k L3.k
 3, model(level):
   1978bn.year 1979.year 1980.year 1981.year 1982.year 1983.year 1984.year
 4, model(level):
   _cons
```

Time-invariant regressors

- Unless the effects of observed time-invariant variables are of particular interest, there is usually no need to explicitly include them in the regression model as they can simply be subsumed under the unit-specific effects:

$$y_{it} = \sum_{j=1}^{q_y} \lambda_j y_{i,t-j} + \sum_{j=0}^{q_x} \mathbf{x}'_{i,t-j}\boldsymbol{\beta}_j + \delta_t + \underbrace{\mathbf{f}'_i\gamma + \alpha_i}_{\tilde{\alpha}_i} + u_{it}$$

- If we still want to estimate the coefficients γ, the transformed instruments $\tilde{\mathbf{Z}}_i^D = \mathbf{D}'_i\mathbf{Z}_i^D$ or $\tilde{\mathbf{Z}}_i^{FOD} = \mathbf{H}'_i\mathbf{Z}_i^{FOD}$ are not useful because they are orthogonal to all time-invariant variables.
  - Appropriate instruments for the level model are needed.

Time-invariant regressors: Hausman-Taylor instruments

- The sys-GMM estimator with first-differenced instruments $\Delta y_{i,t-1}$ and $\Delta \mathbf{x}_{it}$ as the only instruments for the level model produces spurious estimates for the coefficients of time-invariant regressors.
  - These instruments are assumed to be uncorrelated with time-invariant variables. The estimates for the coefficients of time-invariant regressors are then driven by spurious correlation in finite samples (Kripfganz and Schwarz, 2019).
- Instruments can be found in the spirit of Hausman and Taylor (1981), assuming that some time-varying regressors $\mathbf{x}_{it}$ are uncorrelated with the unobserved effects α_i (and sufficiently correlated with the endogenous time-invariant regressors $\mathbf{f}_i$).
  - These regressors (or their within-group averages $\bar{\mathbf{x}}_i$ if they are strictly exogenous) can serve as instruments for the level model if they are uncorrelated with α_i.

Time-invariant regressors: overidentification test

- Excluded instruments in the traditional sense can also be used.
- To identify γ, the number of all relevant level instruments must be at least as large as the number of time-invariant regressors. If it is strictly larger, incremental overidentification tests can be used (Kripfganz and Schwarz, 2019).
  - As a word of caution, if the coefficients γ of the time-invariant regressors are overidentified, incorrect exogeneity assumptions about the additional instruments can cause inconsistency of all coefficient estimates (not just those of the time-invariant regressors).[5]

[5] To avoid this problem, the Kripfganz and Schwarz (2019) two-stage procedure might be useful.

Time-invariant regressors: Mundlak approach

- As an alternative to the Hausman and Taylor (1981) assumption, a correlated random-effects (CRE) approach (Mundlak, 1978) could be used, assuming that the unobserved effects α_i are uncorrelated with the observed time-invariant regressors $\mathbf{f}_i$ after adding the within-group averages $\bar{\mathbf{x}}_i$ (or the initial observations $\mathbf{x}_{i0}$ in the case of predetermined variables, with or without y_{i0}) as exogenous time-invariant regressors (Kripfganz and Schwarz, 2019).
  - Once it is reasonable to assume that all time-invariant regressors $\mathbf{f}_i$ are uncorrelated with α_i, they can serve as their own level instruments.
  - The CRE assumption is untestable.

Estimation with time-invariant regressors in Stata

- Estimation with exogenous industry dummy variables:

```
. xtdpdgmm L(0/1).n w k i.ind, model(diff) collapse gmm(n, lag(2 4)) gmm(w k, lag(1 3)) ///
> iv(i.ind, model(level)) nl(noserial) teffects igmm vce(r)
(Some output omitted)
Instruments corresponding to the linear moment conditions:
 1, model(diff):
   L2.n L3.n L4.n
 2, model(diff):
   L1.w L2.w L3.w L1.k L2.k L3.k
 3, model(level):
   2bn.ind 3.ind 4.ind 5.ind 6.ind 7.ind 8.ind 9.ind
 4, model(level):
   1978bn.year 1979.year 1980.year 1981.year 1982.year 1983.year 1984.year
 5, model(level):
   _cons
```

- In this case, the exogeneity assumption for the industry dummies cannot be tested because their coefficients are no longer identified when the respective instruments / identifying restrictions are excluded.

Small-sample test statistics

- By default, `xtdpdgmm` reports asymptotically standard-normally distributed z-statistics, and the postestimation `test` command for linear hypotheses reports the asymptotically χ^2-distributed Wald statistic.
- In small samples, the t-distribution or the F-distribution might have better coverage. `xtdpdgmm` reports the t-statistic (and the F-statistic with the `test` command) if the option `small` is specified.
  - Stata's usual small-sample degrees-of-freedom correction is applied to the covariance matrix in that case: $\frac{NT}{NT-K}$, or $\frac{M}{M-1}\frac{NT-1}{NT-K}$ with panel-robust or cluster-robust standard errors, where M denotes the number of groups / clusters.

Deviations from within-group means

- For strictly exogenous regressors $\mathbf{x}_{it}$, the following moment conditions for the model in deviations from within-group means, option `model(mdev)`, are valid:

$$E[\mathbf{x}_{it}\ddot{\Delta} u_{it}] = 0, \quad t = 1, 2, \ldots, T$$

where $\ddot{\Delta} u_{it} = \sqrt{\frac{T}{T-1}} \underbrace{(u_{it} - \bar{u}_i)}_{(e_{it}-\bar{e}_i)}$.

  - Unless the option `norescale` is specified, `xtdpdgmm` applies the factor $\sqrt{\frac{T}{T-1}}$, analogously to forward-orthogonal deviations. In unbalanced panels, the factor ensures that groups with different numbers of observations receive proportionate weights. In balanced panels, it is irrelevant.
  - The collapsed version of the (unweighted) moment conditions, $E\left[\sum_{t=1}^{T} \mathbf{x}_{it}(u_{it} - \bar{u}_i)\right] = 0$, corresponds to those utilized by the conventional fixed-effects estimator.

Deviations from within-group means: static model

- Static fixed-effects estimator:

```
. xtdpdgmm n w k, model(mdev) iv(w k, norescale) vce(r) small
(Some output omitted)
------------------------------------------------------------------------------
             |               Robust
           n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |   -.367774   .1163345    -3.16   0.002    -.5977879   -.1377601
           k |   .6403675   .0449394    14.25   0.000     .5515144    .7292206
       _cons |   2.494684   .3566839     6.99   0.000     1.789456    3.199911
------------------------------------------------------------------------------
(Some output omitted)

. xtreg n w k, fe vce(r)
(Some output omitted)
------------------------------------------------------------------------------
             |               Robust
           n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           w |   -.367774   .1163345    -3.16   0.002    -.5977879   -.1377601
           k |   .6403675   .0449394    14.25   0.000     .5515144    .7292206
       _cons |   2.494684   .3557261     7.01   0.000      1.79135    3.198017
-------------+----------------------------------------------------------------
(Some output omitted)
```

Model selection: specification search

- Unless (economic) theory gives a clear prescription of the model to be estimated, a specification search might be necessary as part of the empirical analysis (Kiviet, 2019).
  - Higher-order lags of the dependent variable, $y_{i,t-2}, y_{i,t-3}, \dots$, and the other regressors, $\mathbf{x}_{i,t-1}, \mathbf{x}_{i,t-2}, \dots$, might have predictive power and could help to prevent serial correlation of the error term u_{it} when included as regressors.
  - Time dummies should be included by default unless there is sufficient evidence against them.
  - Interaction effects among the explanatory variables (possibly including lags of the variables and time dummies) might be necessary to allow for heterogeneity in the dynamic impact multipliers.
  - The regressors $\mathbf{x}_{it}$ need to be classified correctly as strictly exogenous, predetermined, or endogenous.

Model and moment selection criteria

- Omitted variables (such as higher-order lags of already included variables as well as other excluded variables) can cause correlation of the instruments with the error term.
  - Rather than dropping seemingly invalid instruments, it is sometimes a better idea to augment the regression model with additional lags or excluded variables.
- The Andrews and Lu (2001) model and moment selection criteria (MMSC) can support the specification search. These criteria subtract a bonus term from the overidentification test statistic that rewards fewer coefficients for a given number of moment conditions (or more overidentifying restrictions for a given number of coefficients).
  - The `xtdpdgmm` postestimation command `estat mmsc` computes the Akaike (AIC), Bayesian (BIC), and Hannan-Quinn (HQIC) versions of the Andrews-Lu MMSC.
  - Models with lower values of the criteria are preferred.

Model and moment selection criteria in Stata

```
. quietly xtdpdgmm L(0/1).n L(0/1).(w k), model(diff) gmm(n, lag(2 .) collapse) ///
> gmm(w k, lag(1 .) collapse) nl(noserial, collapse) teffects igmm vce(r)

. estat overid
(Some output omitted)
16-step moment functions, 16-step weighting matrix     chi2(19)     =   28.5871
                                                       Prob > chi2 =    0.0728

(Some output omitted)

. estimates store xlags

. quietly xtdpdgmm L(0/1).n w k, model(diff) gmm(n, lag(2 .) collapse) gmm(w k, lag(1 .) collapse) ///
> nl(noserial, collapse) teffects igmm vce(r)

. estat overid
(Some output omitted)
18-step moment functions, 18-step weighting matrix     chi2(21)     =   30.2297
                                                       Prob > chi2 =    0.0875

(Some output omitted)

. estat mmsc xlags

Andrews-Lu model and moment selection criteria

       Model | ngroups          J   nmom   npar   MMSC-AIC   MMSC-BIC  MMSC-HQIC
-------------+-----------------------------------------------------------------
           . |     140    30.2297     32     11   -11.7703   -73.5448   -37.5447
       xlags |     140    28.5871     32     13    -9.4129   -65.3042   -32.7326
```

Sequential model selection process

- The following sequential selection process is adapted from Kiviet (2019), with some modifications.

1. Specify an initial candidate "maintained statistical model" (MSM).
   - An initial candidate MSM should avoid the omission of relevant regressors, include sufficient lags and time dummies, and treat variables $\mathbf{x}_{it}$ as endogenous (unless there is opposing theory or evidence), but it should also avoid an overparametrization.
   - If the sample size permits, use all available instruments for the first-differenced or FOD-transformed model. In small samples, collapse and/or curtail the instruments. As a (somewhat arbitrary) rule of thumb, Kiviet (2019) suggests:

$$K + 4 \leq L < \min\left(h_K K, \frac{1}{h_L}(NT - K)\right)$$

where $4 < h_k < h_L < 10$.

Sequential model selection process

2. Compute the two-step GMM estimator with Windmeijer-corrected standard errors for the initial candidate MSM, and check whether it passes the specification tests.[6]
    - If there are concerns about an imprecisely estimated optimal weighting matrix, the one-step GMM estimator with robust standard errors might be used instead.
    - Check the serial correlation tests at least up to order 2.
    - Check the overall overidentification test and the incremental overidentification tests for each subset of instruments.
    - If any of the tests is not satisfied, go back to step 1 and amend the initial candidate MSM.

[6] See Kiviet (2019) for a discussion of reasonable p-value ranges.

Sequential model selection process in Stata

- Initial candidate MSM with time dummies and 3 lags for all variables, treating w, k, and ys as endogenous with collapsed but non-curtailed instruments for the FOD-transformed model:

```
. xtdpdgmm L(0/3).n L(0/3).(w k ys), model(fod) collapse gmm(n, lag(1 .)) gmm(w, lag(1 .)) ///
> gmm(k, lag(1 .)) gmm(ys, lag(1 .)) teffects two vce(r) overid
(Some output omitted)
Instruments corresponding to the linear moment conditions:
 1, model(fodev):
   L1.n L2.n L3.n L4.n L5.n L6.n L7.n
 2, model(fodev):
   L1.w L2.w L3.w L4.w L5.w L6.w L7.w
 3, model(fodev):
   L1.k L2.k L3.k L4.k L5.k L6.k L7.k
 4, model(fodev):
   L1.ys L2.ys L3.ys L4.ys L5.ys L6.ys L7.ys
 5, model(level):
   1980bn.year 1981.year 1982.year 1983.year 1984.year
 6, model(level):
   _cons

. estat serial, ar(1/3)

Arellano-Bond test for autocorrelation of the first-differenced residuals
H0: no autocorrelation of order 1:     z =   -4.4534   Prob > |z|  =    0.0000
H0: no autocorrelation of order 2:     z =   -0.1300   Prob > |z|  =    0.8966
H0: no autocorrelation of order 3:     z =   -0.3777   Prob > |z|  =    0.7057
```

Sequential model selection process in Stata

```
. estat overid

Sargan-Hansen test of the overidentifying restrictions
H0: overidentifying restrictions are valid

2-step moment functions, 2-step weighting matrix       chi2(13)     =   12.6823
                                                       Prob > chi2 =     0.4726

2-step moment functions, 3-step weighting matrix       chi2(13)     =   15.3271
                                                       Prob > chi2 =     0.2874

. estat overid, difference

Sargan-Hansen (difference) test of the overidentifying restrictions
H0: (additional) overidentifying restrictions are valid

2-step weighting matrix from full model

                  | Excluding                     | Difference
Moment conditions |       chi2     df           p |        chi2     df           p
------------------+-------------------------------+------------------------------
  1, model(fodev) |     8.9323      6      0.1774 |      3.7500      7      0.8081
  2, model(fodev) |     9.8897      6      0.1294 |      2.7926      7      0.9035
  3, model(fodev) |     9.2784      6      0.1585 |      3.4039      7      0.8453
  4, model(fodev) |     6.2261      6      0.3983 |      6.4561      7      0.4876
  5, model(level) |     9.6163      8      0.2930 |      3.0659      5      0.6898
     model(fodev) |          .    -15           . |           .      .           .

. estimates store model1
```

Sequential model selection process

3. Remove lags or interaction effects with (very) high p-values in individual or joint significance tests, and/or check whether further lags or interaction effects improve the model fit, adjusted for the degrees of freedom.
   - Reduce the model sequentially, i.e. remove the longest lag or interaction effect with the highest p-value first and reestimate the model. Repeat the procedure until none of the longest lags has (very) high p-values any more.
   - Keep in mind that increasing the lag orders q_y and/or q_x reduces the sample size which can be costly when T is small.
   - For every new candidate model, carry out the specification tests as in step 2.
   - Use the MMSC to compare the candidate models that pass the specification tests.
   - Check whether the results for the preferred model are robust to the estimation with the iterated GMM estimator and to alternative ways of instrument reduction.

Sequential model selection process in Stata

```
. testparm L3.k

 ( 1)  L3.k = 0

           chi2(  1) =    0.02
         Prob > chi2 =    0.9011

. xtdpdgmm L(0/3).n L(0/3).w L(0/2).k L(0/3).ys, model(fod) collapse gmm(n, lag(1 .)) ///
> gmm(w, lag(1 .)) gmm(k, lag(1 .)) gmm(ys, lag(1 .)) teffects two vce(r) overid
(Output omitted)

. estat serial, ar(1/3)

Arellano-Bond test for autocorrelation of the first-differenced residuals
H0: no autocorrelation of order 1:     z =   -4.5960   Prob > |z|  =    0.0000
H0: no autocorrelation of order 2:     z =   -0.2258   Prob > |z|  =    0.8213
H0: no autocorrelation of order 3:     z =   -0.3713   Prob > |z|  =    0.7104

. estat overid

Sargan-Hansen test of the overidentifying restrictions
H0: overidentifying restrictions are valid

2-step moment functions, 2-step weighting matrix       chi2(14)    =   12.2034
                                                       Prob > chi2 =    0.5900
(Some output omitted)

. estat overid, difference
(Output omitted)

. estimates store model2
```

Sequential model selection process in Stata

```
. testparm L3.n

 ( 1)  L3.n = 0

           chi2(  1) =    0.20
         Prob > chi2 =    0.6520

. xtdpdgmm L(0/2).n L(0/3).w L(0/2).k L(0/3).ys, model(fod) collapse gmm(n, lag(1 .)) ///
> gmm(w, lag(1 .)) gmm(k, lag(1 .)) gmm(ys, lag(1 .)) teffects two vce(r) overid
(Output omitted)

. estat serial, ar(1/3)

Arellano-Bond test for autocorrelation of the first-differenced residuals
H0: no autocorrelation of order 1:     z =   -4.5016   Prob > |z|  =    0.0000
H0: no autocorrelation of order 2:     z =   -0.1957   Prob > |z|  =    0.8448
H0: no autocorrelation of order 3:     z =   -0.2132   Prob > |z|  =    0.8312

. estat overid

Sargan-Hansen test of the overidentifying restrictions
H0: overidentifying restrictions are valid

2-step moment functions, 2-step weighting matrix       chi2(15)    =   12.1648
                                                       Prob > chi2 =    0.6665
(Some output omitted)

. estat overid, difference
(Output omitted)

. estimates store model3
```

Sequential model selection process in Stata

```
. testparm L2.k

 ( 1)  L2.k = 0

           chi2(  1) =    0.20
         Prob > chi2 =    0.6520

. xtdpdgmm L(0/2).n L(0/3).w L(0/1).k L(0/3).ys, model(fod) collapse gmm(n, lag(1 .)) ///
> gmm(w, lag(1 .)) gmm(k, lag(1 .)) gmm(ys, lag(1 .)) teffects two vce(r) overid
(Output omitted)

. estat serial, ar(1/3)

Arellano-Bond test for autocorrelation of the first-differenced residuals
H0: no autocorrelation of order 1:     z =   -4.2569   Prob > |z|  =    0.0000
H0: no autocorrelation of order 2:     z =    0.0883   Prob > |z|  =    0.9296
H0: no autocorrelation of order 3:     z =   -0.1340   Prob > |z|  =    0.8934

. estat overid

Sargan-Hansen test of the overidentifying restrictions
H0: overidentifying restrictions are valid

2-step moment functions, 2-step weighting matrix       chi2(16)    =   12.0198
                                                       Prob > chi2 =    0.7426

(Some output omitted)

. estat overid, difference
(Output omitted)

. estimates store model4
```

Sequential model selection process in Stata

```
. testparm L3.w

 ( 1)  L3.w = 0

           chi2(  1) =    0.65
         Prob > chi2 =    0.4189

. xtdpdgmm L(0/2).n L(0/2).w L(0/1).k L(0/3).ys, model(fod) collapse gmm(n, lag(1 .)) ///
> gmm(w, lag(1 .)) gmm(k, lag(1 .)) gmm(ys, lag(1 .)) teffects two vce(r) overid
(Output omitted)

. estat serial, ar(1/3)

Arellano-Bond test for autocorrelation of the first-differenced residuals
H0: no autocorrelation of order 1:     z =   -4.3570   Prob > |z|  =    0.0000
H0: no autocorrelation of order 2:     z =   -0.0999   Prob > |z|  =    0.9205
H0: no autocorrelation of order 3:     z =   -0.0464   Prob > |z|  =    0.9630

. estat overid

Sargan-Hansen test of the overidentifying restrictions
H0: overidentifying restrictions are valid

2-step moment functions, 2-step weighting matrix       chi2(17)     =   12.9399
                                                       Prob > chi2 =    0.7402
(Some output omitted)

. estat overid, difference
(Output omitted)

. estimates store model5
```

Sequential model selection process in Stata

```
. testparm L.k

 ( 1)  L.k = 0

           chi2(  1) =    0.65
         Prob > chi2 =    0.4216

. xtdpdgmm L(0/2).n L(0/2).w k L(0/3).ys, model(fod) collapse gmm(n, lag(1 .)) gmm(w, lag(1 .)) ///
> gmm(k, lag(1 .)) gmm(ys, lag(1 .)) teffects two vce(r) overid
(Output omitted)

. estat serial, ar(1/3)

Arellano-Bond test for autocorrelation of the first-differenced residuals
H0: no autocorrelation of order 1:     z =   -4.7944   Prob > |z|  =    0.0000
H0: no autocorrelation of order 2:     z =   -0.4182   Prob > |z|  =    0.6758
H0: no autocorrelation of order 3:     z =   -0.4924   Prob > |z|  =    0.6225

. estat overid

Sargan-Hansen test of the overidentifying restrictions
H0: overidentifying restrictions are valid

2-step moment functions, 2-step weighting matrix       chi2(18)     =   13.6173
                                                       Prob > chi2 =    0.7537
(Some output omitted)

. estat overid, difference
(Output omitted)

. estimates store model6
```

Sequential model selection process in Stata

- Square of w and interaction effect between w and k added:

```
. xtdpdgmm L(0/2).n L(0/2).w k L(0/3).ys c.w#c.w c.w#c.k, model(fod) collapse gmm(n, lag(1 .)) ///
> gmm(w, lag(1 .)) gmm(k, lag(1 .)) gmm(ys, lag(1 .)) gmm(c.w#c.w, lag(1 .)) gmm(c.w#c.k, lag(1 .)) ///
> teffects two vce(r) overid
(Output omitted)

. estat serial, ar(1/3)

Arellano-Bond test for autocorrelation of the first-differenced residuals
H0: no autocorrelation of order 1:     z =   -3.3178   Prob > |z|  =    0.0009
H0: no autocorrelation of order 2:     z =    0.2324   Prob > |z|  =    0.8162
H0: no autocorrelation of order 3:     z =   -0.8583   Prob > |z|  =    0.3907

. estat overid

Sargan-Hansen test of the overidentifying restrictions
H0: overidentifying restrictions are valid

2-step moment functions, 2-step weighting matrix       chi2(30)    =   22.2653
                                                       Prob > chi2 =    0.8442
(Some output omitted)

. estat overid, difference
(Output omitted)

. estimates store model7
```

Sequential model selection process in Stata

```
. testparm i.year

 ( 1)  1980bn.year = 0
 ( 2)  1981.year = 0
 ( 3)  1982.year = 0
 ( 4)  1983.year = 0
 ( 5)  1984.year = 0

           chi2(  5) =    3.46
         Prob > chi2 =    0.6297

. xtdpdgmm L(0/2).n L(0/2).w k L(0/3).ys c.w#c.w c.w#c.k, model(fod) collapse gmm(n, lag(1 .)) ///
> gmm(w, lag(1 .)) gmm(k, lag(1 .)) gmm(ys, lag(1 .)) gmm(c.w#c.w, lag(1 .)) gmm(c.w#c.k, lag(1 .)) ///
> two vce(r) overid
(Output omitted)

. estat serial, ar(1/3)
(Output omitted)

. estat overid

Sargan-Hansen test of the overidentifying restrictions
H0: overidentifying restrictions are valid

2-step moment functions, 2-step weighting matrix       chi2(30)     =   27.5377
                                                       Prob > chi2 =    0.5949

(Some output omitted)

. estat overid, difference
(Output omitted)
```

Sequential model selection process in Stata

```
. estat mmsc model7 model6 model5 model4 model3 model2 model1

Andrews-Lu model and moment selection criteria

       Model | ngroups          J   nmom   npar   MMSC-AIC   MMSC-BIC  MMSC-HQIC
-------------+------------------------------------------------------------------
           . |     140    27.5377     43     13   -32.4623  -120.7116   -69.2828
      model7 |     140    22.2653     48     18   -37.7347  -125.9840   -74.5552
      model6 |     140    13.6173     34     16   -22.3827   -75.3323   -44.4750
      model5 |     140    12.9399     34     17   -21.0601   -71.0680   -41.9250
      model4 |     140    12.0198     34     18   -19.9802   -67.0465   -39.6178
      model3 |     140    12.1648     34     19   -17.8352   -61.9598   -36.2454
      model2 |     140    12.2034     34     20   -15.7966   -56.9796   -32.9795
      model1 |     140    12.6823     34     21   -13.3177   -51.5591   -29.2733
```

- Among the considered candidates, the MMSC select `model7`.
    - Despite their joint statistical insignificance with large *p*-value, omitting the time dummies is not supported by the MMSC.
    - Other models with further interaction terms or lags of interaction terms might be worth taking into consideration.
    - Another sequential selection strategy might be to add interaction terms first before reducing lag orders, i.e. an *inductive bottom-up discovery phase* followed by a *deductive top-down specialization phase* (Kiviet, 2019).

Sequential model selection process

4. Separately for all regressors classified as endogenous, add the extra instruments that become valid if the regressors were predetermined (unless theory clearly indicates that a variable should be endogenous), and check the corresponding incremental overidentification tests.
    - Keep an eye on other specification tests and MMSC as well.
    - Treat the variable with the highest acceptable *p*-value of the incremental overidentification tests as predetermined, and repeat the procedure for the remaining variables until no more variable can be confidently classified as predetermined.
5. Separately for all regressors classified as predetermined, add the extra instruments that become valid if the regressors were strictly exogenous, and follow the procedure of step 4.
    - Have a look at underidentification tests as well. Passing the underidentification tests might require stronger exogeneity assumptions, possibly creating a conflict with overidentification tests.

Sequential model selection process in Stata

```
. estimates restore model7
(results model7 are active now)

. underid, underid kp sw noreport

Underidentification test: Kleibergen-Paap robust LIML-based (LM version)
  Test statistic robust to heteroskedasticity and clustering on id
j=   36.33  Chi-sq( 31) p-value=0.2342

2-step GMM J underidentification stats by regressor:
j=   40.36  Chi-sq( 31) p-value=0.1212 L.n
j=   40.77  Chi-sq( 31) p-value=0.1127 L2.n
j=   40.99  Chi-sq( 31) p-value=0.1082 w
j=   36.37  Chi-sq( 31) p-value=0.2328 L.w
j=   55.29  Chi-sq( 31) p-value=0.0046 L2.w
j=   37.38  Chi-sq( 31) p-value=0.1993 k
j=   59.63  Chi-sq( 31) p-value=0.0015 ys
j=   66.14  Chi-sq( 31) p-value=0.0002 L.ys
j=   75.12  Chi-sq( 31) p-value=0.0000 L2.ys
j=   64.30  Chi-sq( 31) p-value=0.0004 L3.ys
j=   41.91  Chi-sq( 31) p-value=0.0914 c.w#c.w
j=   34.58  Chi-sq( 31) p-value=0.3007 c.w#c.k
j=   92.43  Chi-sq( 31) p-value=0.0000 1980bn.year
j=   92.43  Chi-sq( 31) p-value=0.0000 1981.year
j=   92.43  Chi-sq( 31) p-value=0.0000 1982.year
j=   92.43  Chi-sq( 31) p-value=0.0000 1983.year
j=   92.43  Chi-sq( 31) p-value=0.0000 1984.year
```

- The underidentification test is not yet satisfying.

Sequential model selection process in Stata

- Treating w as predetermined with collapsed instruments, adds one more moment condition:

```
. xtdpdgmm L(0/2).n L(0/2).w k L(0/3).ys c.w#c.w c.w#c.k, model(fod) collapse gmm(n, lag(1 .)) ///
> gmm(w, lag(1 .)) gmm(k, lag(1 .)) gmm(ys, lag(1 .)) gmm(c.w#c.w, lag(1 .)) gmm(c.w#c.k, lag(1 .)) ///
> gmm(w, lag(0 0)) teffects two vce(r) overid
(Some output omitted)
Instruments corresponding to the linear moment conditions:
 1, model(fodev):
   L1.n L2.n L3.n L4.n L5.n L6.n L7.n
 2, model(fodev):
   L1.w L2.w L3.w L4.w L5.w L6.w L7.w
 3, model(fodev):
   L1.k L2.k L3.k L4.k L5.k L6.k L7.k
 4, model(fodev):
   L1.ys L2.ys L3.ys L4.ys L5.ys L6.ys L7.ys
 5, model(fodev):
   L1.c.w#c.w L2.c.w#c.w L3.c.w#c.w L4.c.w#c.w L5.c.w#c.w L6.c.w#c.w
   L7.c.w#c.w
 6, model(fodev):
   L1.c.w#c.k L2.c.w#c.k L3.c.w#c.k L4.c.w#c.k L5.c.w#c.k L6.c.w#c.k
   L7.c.w#c.k
 7, model(fodev):
   w
 8, model(level):
   1980bn.year 1981.year 1982.year 1983.year 1984.year
 9, model(level):
   _cons
```

Sequential model selection process in Stata

```
. estat serial, ar(1/3)
(Output omitted)

. estat overid
(Output omitted)

. estat overid, difference

Sargan-Hansen (difference) test of the overidentifying restrictions
H0: (additional) overidentifying restrictions are valid

2-step weighting matrix from full model

                  | Excluding                     | Difference
Moment conditions |       chi2     df          p |        chi2     df          p
------------------+-----------------------------+-----------------------------
   1, model(fodev) |    18.0364     24     0.8012 |      5.2684      7     0.6273
   2, model(fodev) |    19.5489     24     0.7221 |      3.7559      7     0.8074
   3, model(fodev) |    16.3453     24     0.8752 |      6.9595      7     0.4331
   4, model(fodev) |    20.9307     24     0.6428 |      2.3740      7     0.9363
   5, model(fodev) |    18.2849     24     0.7890 |      5.0198      7     0.6575
   6, model(fodev) |    16.2789     24     0.8777 |      7.0259      7     0.4262
   7, model(fodev) |    22.2441     30     0.8450 |      1.0607      1     0.3031
   8, model(level) |    23.0013     26     0.6329 |      0.3035      5     0.9976
      model(fodev) |          .    -12          . |           .      .          .
```

- The *p*-value of the incremental overidentification test might be acceptable in order to reduce the risk of underidentification.

Sequential model selection process in Stata

- Skipping some intermediate steps, we arrive at a model with `w` and `k` (as well as the interaction terms) treated as predetermined:

```
. xtdpdgmm L(0/2).n L(0/2).w k L(0/3).ys c.w#c.w c.w#c.k, model(fod) collapse gmm(n, lag(1 .)) ///
> gmm(w, lag(1 .)) gmm(k, lag(1 .)) gmm(ys, lag(1 .)) gmm(c.w#c.w, lag(1 .)) gmm(c.w#c.k, lag(1 .)) ///
> gmm(w k c.w#c.w c.w#c.k, lag(0 0)) teffects two vce(r) overid
(Output omitted)

. estat serial, ar(1/3)
(Output omitted)

. estat overid
(Output omitted)

. estat overid, difference
(Output omitted)

. estat mmsc model7

Andrews-Lu model and moment selection criteria

       Model | ngroups          J   nmom   npar   MMSC-AIC   MMSC-BIC  MMSC-HQIC
-------------+------------------------------------------------------------------
           . |     140    24.6025     52     18   -43.3975  -143.4134   -85.1274
      model7 |     140    22.2653     48     18   -37.7347  -125.9840   -74.5552

. estimates store model7pre
```

Sequential model selection process in Stata

```
. underid, underid kp sw noreport

Underidentification test: Kleibergen-Paap robust LIML-based (LM version)
  Test statistic robust to heteroskedasticity and clustering on id
j=   42.32  Chi-sq( 35) p-value=0.1844

2-step GMM J underidentification stats by regressor:
j=   46.05  Chi-sq( 35) p-value=0.1002 L.n
j=   47.31  Chi-sq( 35) p-value=0.0801 L2.n
j=   45.78  Chi-sq( 35) p-value=0.1049 w
j=   40.58  Chi-sq( 35) p-value=0.2377 L.w
j=   64.82  Chi-sq( 35) p-value=0.0016 L2.w
j=   44.40  Chi-sq( 35) p-value=0.1326 k
j=   64.09  Chi-sq( 35) p-value=0.0019 ys
j=   78.26  Chi-sq( 35) p-value=0.0000 L.ys
j=   84.90  Chi-sq( 35) p-value=0.0000 L2.ys
j=   81.45  Chi-sq( 35) p-value=0.0000 L3.ys
j=   45.70  Chi-sq( 35) p-value=0.1065 c.w#c.w
j=   56.93  Chi-sq( 35) p-value=0.0110 c.w#c.k
j=   97.78  Chi-sq( 35) p-value=0.0000 1980bn.year
j=   97.78  Chi-sq( 35) p-value=0.0000 1981.year
j=   97.78  Chi-sq( 35) p-value=0.0000 1982.year
j=   97.78  Chi-sq( 35) p-value=0.0000 1983.year
j=   97.78  Chi-sq( 35) p-value=0.0000 1984.year
```

- The underidentification test is still unsatisfying.

Sequential model selection process in Stata

- Again skipping some intermediate steps, we might be willing to treat k as strictly exogenous, using its contemporaneous term as an instrument for the model in mean deviations:

```
. xtdpdgmm L(0/2).n L(0/2).w k L(0/3).ys c.w#c.w c.w#c.k, model(fod) collapse gmm(n, lag(1 .)) ///
> gmm(w, lag(0 .)) gmm(k, lag(0 .)) gmm(ys, lag(1 .)) gmm(c.w#c.w, lag(0 .)) gmm(c.w#c.k, lag(0 .)) ///
> gmm(k, lag(0 0) model(md)) teffects two vce(r) overid
(Some output omitted)
Instruments corresponding to the linear moment conditions:
 1, model(fodev):
   L1.n L2.n L3.n L4.n L5.n L6.n L7.n
 2, model(fodev):
   w L1.w L2.w L3.w L4.w L5.w L6.w L7.w
 3, model(fodev):
   k L1.k L2.k L3.k L4.k L5.k L6.k L7.k
 4, model(fodev):
   L1.ys L2.ys L3.ys L4.ys L5.ys L6.ys L7.ys
 5, model(fodev):
   c.w#c.w L1.c.w#c.w L2.c.w#c.w L3.c.w#c.w L4.c.w#c.w L5.c.w#c.w L6.c.w#c.w
   L7.c.w#c.w
 6, model(fodev):
   c.w#c.k L1.c.w#c.k L2.c.w#c.k L3.c.w#c.k L4.c.w#c.k L5.c.w#c.k L6.c.w#c.k
   L7.c.w#c.k
 7, model(mdev):
   k
 8, model(level):
   1980bn.year 1981.year 1982.year 1983.year 1984.year
 9, model(level):
   _cons
```

Sequential model selection process in Stata

```
. estat serial, ar(1/3)
(Output omitted)

. estat overid

Sargan-Hansen test of the overidentifying restrictions
H0: overidentifying restrictions are valid

2-step moment functions, 2-step weighting matrix       chi2(35)     =   27.1733
                                                       Prob > chi2 =     0.8250

(Some output omitted)

. estat overid, difference

Sargan-Hansen (difference) test of the overidentifying restrictions
H0: (additional) overidentifying restrictions are valid

2-step weighting matrix from full model

                  | Excluding                      | Difference
Moment conditions |       chi2     df            p |        chi2     df            p
------------------+--------------------------------+--------------------------------
  1, model(fodev) |    25.0233     28       0.6266 |      2.1499      7       0.9511
  2, model(fodev) |    22.7133     27       0.7003 |      4.4600      8       0.8134
  3, model(fodev) |    22.0626     27       0.7342 |      5.1107      8       0.7457
  4, model(fodev) |    26.2077     28       0.5616 |      0.9656      7       0.9954
  5, model(fodev) |    22.9058     27       0.6901 |      4.2674      8       0.8322
  6, model(fodev) |    22.4188     27       0.7158 |      4.7544      8       0.7835
   7, model(mdev) |    26.4764     34       0.8179 |      0.6968      1       0.4039
  8, model(level) |    24.8303     30       0.7332 |      2.3430      5       0.7999
     model(fodev) |          .    -11            . |           .      .            .
```

Sequential model selection process in Stata

```
. estat mmsc model7pre model7

Andrews-Lu model and moment selection criteria

       Model | ngroups          J   nmom   npar    MMSC-AIC    MMSC-BIC   MMSC-HQIC
-------------+--------------------------------------------------------------------
           . |     140    27.1733     53     18    -42.8267   -145.7842    -85.7840
   model7pre |     140    24.6025     52     18    -43.3975   -143.4134    -85.1274
      model7 |     140    22.2653     48     18    -37.7347   -125.9840    -74.5552

. underid, underid kp noreport

Underidentification test: Kleibergen-Paap robust LIML-based (LM version)
  Test statistic robust to heteroskedasticity and clustering on id
j=    59.95  Chi-sq( 36) p-value=0.0074
```

- Treating `k` as strictly exogenous does not improve the MMSC much but it apparently helps a lot to pass the underidentification test.

Sequential model selection process

6. Possibly, repeat step 3 based on the new MSM from step 5.
   - Note that L will be generally larger after steps 4 and 5. A further reduction of the instrument count by collapsing and/or curtailing might become necessary.
   - If predicted by theory, it might be worth exploring other coefficient restrictions besides those of equality to zero.
   - Keep in mind that statistical insignificance per se is not a sufficient reason to exclude a variable, in particular if the point estimate is (economically) large or if the effect of this variable is of particular interest in the analysis.

Sequential model selection process

7. If there are any time-invariant regressors of particular interest (beyond the mere desire to control for them), add them and sufficiently many instruments for the level model. Estimate the model by two-step or iterated sys-GMM with Windmeijer-corrected standard errors
   - Keep in mind that the inclusion of time-invariant regressors generally requires potentially strong identifying assumption.
   - If the coefficients of the time-invariant regressors are overidentified, check the incremental overidentification tests (and possibly underidentification tests as well).

Sequential model selection process

8. Unless there is opposing theory or evidence, add the additional instruments that are valid under the Blundell and Bond (1998) initial-conditions assumption. Estimate the model by two-step or iterated sys-GMM with Windmeijer-corrected standard errors, and check the incremental overidentification tests.
   - Separately investigate the additional instruments $\Delta\mathbf{x}_{it}$ (or $\Delta\mathbf{x}_{i,t-1}$) one by one for the level model first. Only if there is sufficiently strong evidence that all of those instruments are valid, add the extra instruments $\Delta y_{i,t-1}$.
   - Keep an eye on the other specification tests as well.

Sequential model selection process in Stata

- Skipping some intermediate steps, using differences of `w` and `k` as instruments for the level model might be acceptable:

```
. xtdpdgmm L(0/2).n L(0/2).w k L(0/3).ys c.w#c.w c.w#c.k, model(fod) collapse gmm(n, lag(1 .)) ///
> gmm(w, lag(0 .)) gmm(k, lag(0 .)) gmm(ys, lag(1 .)) gmm(c.w#c.w, lag(0 .)) gmm(c.w#c.k, lag(0 .)) ///
> gmm(k, lag(0 0) model(md)) gmm(w k, lag(0 0) diff model(level)) teffects two vce(r) overid
(Some output omitted)
Instruments corresponding to the linear moment conditions:
 1, model(fodev):
   L1.n L2.n L3.n L4.n L5.n L6.n L7.n
 2, model(fodev):
   w L1.w L2.w L3.w L4.w L5.w L6.w L7.w
 3, model(fodev):
   k L1.k L2.k L3.k L4.k L5.k L6.k L7.k
 4, model(fodev):
   L1.ys L2.ys L3.ys L4.ys L5.ys L6.ys L7.ys
 5, model(fodev):
   c.w#c.w L1.c.w#c.w L2.c.w#c.w L3.c.w#c.w L4.c.w#c.w L5.c.w#c.w L6.c.w#c.w
   L7.c.w#c.w
 6, model(fodev):
   c.w#c.k L1.c.w#c.k L2.c.w#c.k L3.c.w#c.k L4.c.w#c.k L5.c.w#c.k L6.c.w#c.k
   L7.c.w#c.k
 7, model(mdev):
   k
 8, model(level):
   D.w D.k
 9, model(level):
   1980bn.year 1981.year 1982.year 1983.year 1984.year
 10, model(level):
   _cons
```

Sequential model selection process in Stata

```
. estat serial, ar(1/3)
(Output omitted)

. estat overid
(Some output omitted)
2-step moment functions, 2-step weighting matrix       chi2(37)     =   31.5940
                                                       Prob > chi2 =    0.7202
(Some output omitted)

. estat overid, difference

Sargan-Hansen (difference) test of the overidentifying restrictions
H0: (additional) overidentifying restrictions are valid

2-step weighting matrix from full model

                 | Excluding                       | Difference
Moment conditions |       chi2     df          p |       chi2     df          p
-----------------+-------------------------------+------------------------------
  1, model(fodev) |    30.5644     30     0.4370 |     1.0296      7     0.9943
  2, model(fodev) |    25.8607     29     0.6329 |     5.7333      8     0.6771
  3, model(fodev) |    26.6376     29     0.5913 |     4.9564      8     0.7622
  4, model(fodev) |    27.3258     30     0.6061 |     4.2682      7     0.7484
  5, model(fodev) |    25.8421     29     0.6339 |     5.7518      8     0.6750
  6, model(fodev) |    27.0201     29     0.5706 |     4.5739      8     0.8020
   7, model(mdev) |    31.5847     36     0.6786 |     0.0093      1     0.9233
  8, model(level) |    31.3841     35     0.6434 |     0.2099      2     0.9004
  9, model(level) |    28.2006     32     0.6594 |     3.3934      5     0.6396
     model(fodev) |          .     -9          . |          .      .          .
     model(level) |    28.1268     30     0.5637 |     3.4672      7     0.8387
```

Sequential model selection process

9. If (many of) the additional level moment conditions in step 8 are rejected, add instead the nonlinear Ahn and Schmidt (1995) moment conditions valid under no serial correlation of u_{it}. Estimate the model by two-step or iterated GMM with Windmeijer-corrected standard errors.
   - A rejection of this model by the specification tests causes doubt on the MSM and might require to revoke some of the decisions made in earlier steps.
   - To improve the efficiency, it might be worth utilizing the nonlinear Ahn and Schmidt (1995) moment conditions valid under homoskedasticity. A generalized Hausman test can be used as a specification test but be aware that it tends to perform poorly in small samples.
   - It might be reasonable to add the nonlinear moment conditions already at a previous step to circumvent identification problems.

Sequential model selection process in Stata: final model

- Given that not all instruments under the initial-conditions assumption appear valid, the GMM estimator with the (collapsed) Ahn and Schmidt (1995) nonlinear moment conditions might be preferable:

```
. xtdpdgmm L(0/2).n L(0/2).w k L(0/3).ys c.w#c.w c.w#c.k, model(fod) collapse gmm(n, lag(1 .)) ///
> gmm(w, lag(0 .)) gmm(k, lag(0 .)) gmm(ys, lag(1 .)) gmm(c.w#c.w, lag(0 .)) gmm(c.w#c.k, lag(0 .)) ///
> gmm(k, lag(0 0) model(md)) teffects nl(noserial) two vce(r) overid
(Some output omitted)
Group variable: id                          Number of obs         =        611
Time variable: year                         Number of groups      =        140

Moment conditions:     linear =        53   Obs per group:    min =          4
                    nonlinear =         1                     avg =   4.364286
                        total =        54                     max =          6

                                  (Std. Err. adjusted for 140 clusters in id)
------------------------------------------------------------------------------
             |              WC-Robust
           n |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |
         L1. |   .9012081   .0928039     9.71   0.000     .7193157      1.0831
         L2. |  -.1379692   .0741801    -1.86   0.063    -.2833595    .0074212
             |
(Continued on next page)
```

Sequential model selection process in Stata: final model

```
           w |
         --. |   2.512475    1.87464     1.34   0.180    -1.161752    6.186703
         L1. |   .3823612   .1126858     3.39   0.001      .161501    .6032213
         L2. |     -.1416   .0962713    -1.47   0.141    -.3302883    .0470883
             |
           k |   .3949068   .2491891     1.58   0.113    -.0934948    .8833084
             |
          ys |
         --. |   .7105045   .2429949     2.92   0.003     .2342433    1.186766
         L1. |  -.9600985    .263219    -3.65   0.000    -1.475998   -.4441988
         L2. |   .1624694   .1969018     0.83   0.409     -.223451    .5483898
         L3. |  -.2515405   .2289312    -1.10   0.272    -.7002374    .1971564
             |
     c.w#c.w |  -.5461882   .3219414    -1.70   0.090    -1.177182    .0848054
             |
     c.w#c.k |     -.0272   .0694873    -0.39   0.695    -.1633926    .1089927
             |
        year |
       1980  |  -.0070694   .0253212    -0.28   0.780    -.0566981    .0425593
       1981  |  -.0350353   .0411753    -0.85   0.395    -.1157374    .0456669
       1982  |  -.0277518   .0501244    -0.55   0.580    -.1259939    .0704904
       1983  |   .0106885   .0554848     0.19   0.847    -.0980598    .1194368
       1984  |  -.0116853    .044806    -0.26   0.794    -.0995034    .0761328
             |
       _cons |  -1.249312     2.6188    -0.48   0.633    -6.382065    3.883442
------------------------------------------------------------------------------
(Some output omitted)

. estat serial, ar(1/3)
(Output omitted)
```

Sequential model selection process in Stata: final model

```
. estat overid

Sargan-Hansen test of the overidentifying restrictions
H0: overidentifying restrictions are valid

2-step moment functions, 2-step weighting matrix       chi2(36)    =   27.7499
                                                       Prob > chi2 =    0.8360

2-step moment functions, 3-step weighting matrix       chi2(36)    =   39.1583
                                                       Prob > chi2 =    0.3300

. estat overid, difference

Sargan-Hansen (difference) test of the overidentifying restrictions
H0: (additional) overidentifying restrictions are valid

2-step weighting matrix from full model

                  | Excluding                     | Difference
Moment conditions |       chi2     df           p |       chi2     df           p
------------------+-------------------------------+------------------------------
  1, model(fodev) |    25.4072     29      0.6570 |     2.3428      7      0.9385
  2, model(fodev) |    23.1059     28      0.7277 |     4.6440      8      0.7949
  3, model(fodev) |    22.3165     28      0.7664 |     5.4334      8      0.7104
  4, model(fodev) |    26.3066     29      0.6091 |     1.4433      7      0.9842
  5, model(fodev) |    23.2937     28      0.7182 |     4.4563      8      0.8138
  6, model(fodev) |    22.9352     28      0.7363 |     4.8147      8      0.7772
   7, model(mdev) |    27.4318     35      0.8154 |     0.3181      1      0.5727
  8, model(level) |    25.3010     31      0.7541 |     2.4489      5      0.7842
    nl(noserial)  |    27.1247     35      0.8268 |     0.6253      1      0.4291
     model(fodev) |          .    -10           . |          .      .           .
```

Disclaimer

- The above procedure serves as a guideline and should not be followed too mechanically.[7] Specification tests cannot provide a definite answer. Each application has its own peculiarities.
- The (finite-sample) properties of the estimators and specification tests depend on characteristics of the (unknown) data-generating process. (For some extensive Monte Carlo evidence, see Kiviet, Pleus, and Poldermans, 2017).
  - There is no unequivocal ranking of curtailing versus collapsing or a combination of both.
  - Even if it is asymptotically inefficient, in some cases the one-step estimator might have better finite-sample properties than the two-step or the iterated GMM estimator.
- Do not use the default settings of statistical software packages unhesitatingly. In case of doubt, make all desired specifications explicit in the command line.

[7] All examples are simplified for expositional purposes.

Summary: the xtdpdgmm package for Stata

- The xtdpdgmm package enables generalized method of moments estimation of linear (dynamic) panel data models.
  - Besides the conventional *difference GMM*, *system GMM*, and GMM with forward-orthogonal deviations, additional nonlinear moment conditions can be incorporated.
  - Besides one-step and feasible efficient two-step estimation, iterated GMM estimation is possible as well.
  - Combining the command with other packages in the Stata universe opens up further possibilities.

```
ssc install xtdpdgmm
net install xtdpdgmm, from(http://www.kripfganz.de/stata/)

help xtdpdgmm
help xtdpdgmm postestimation
```

Acknowledgment: This presentation and the current version of the xtdpdgmm package benefited significantly from discussions with the Stata community, in particular Mark Schaffer and Jan Kiviet.

References

- Ahn, S. C., and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics 68* (1): 5–27.
- Anderson, T. W., and C. Hsiao (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association 76* (375): 598–606.
- Andrews, D. W. K, and B. Lu (2001). Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics 101* (1): 123–164.
- Arellano, M., and S. R. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58* (2): 277–297.
- Arellano, M., and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68* (1): 29–51.
- Baum, C. F., M. E. Schaffer, and S. Stillman (2003). Instrumental variables and GMM: Estimation and testing. *Stata Journal 3* (1): 1–31.
- Baum, C. F., M. E. Schaffer, and S. Stillman (2007). Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal 7* (4): 465–506.
- Blundell, R., and S. R. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87* (1): 115–143.
- Blundell, R., S. R. Bond, and F. Windmeijer (2001). Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator. *Advances in Econometrics 15* (1): 53–91.
- Cragg, J. G., and S. G. Donald (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory 9* (2): 222-240.
- Eichenbaum, M. S., L. P. Hansen, and K. J. Singleton (1988). A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *Quarterly Journal of Economics 103* (1): 51–78.

References

- Gørgens, T., C. Han, and S. Xue (2019). Moment restrictions and identification in linear dynamic panel data models. *Annals of Economics and Statistics 134*: 149–176.
- Hansen, B. E., and S. Lee (2019). Inference for iterated GMM under misspecification. *Manuscript*, University of Wisconsin and University of New South Wales.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50* (4): 1029–1054.
- Hansen, L. P., J. Heaton, and A. Yaron (1996). Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics 14* (3): 262–280.
- Hausman, J. A. (1978). Specification tests in Econometrics. *Econometrica 46* (6): 1251–1271.
- Hausman, J. A., and W. E. Taylor (1981). Panel data and unobservable individual effects. *Econometrica 49* (6): 1377–1398.
- Hayakawa, K., M. Qi, and J. Breitung (2019). Double filter instrumental variable estimation of panel data models with weakly exogenous variables. *Econometric Reviews 38* (9): 1055–1088.
- Kiviet, J. F. (2019). Microeconometric dynamic panel data methods: Model specification and selection issues. *MPRA Paper* 95159, Munich Personal RePEc Archive.
- Kiviet, J. F., M. Pleus, and R. W. Poldermans (2017). Accuracy and efficiency of various GMM inference techniques in dynamic micro panel data models. *Econometrics 5* (1): 14.
- Kleibergen, F., and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics 133* (1): 97–126.
- Kripfganz, S., and C. Schwarz (2019). Estimation of linear dynamic panel data models with time-invariant regressors. *Journal of Applied Econometrics 34* (4): 526–546.

References

- Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica 46* (1): 69–85.
- Newey, W. K. (1985). Generalized method of moments specification testing. *Journal of Econometrics 29* (3): 229–256.
- Roodman, D. (2009a). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics 71* (1): 135–158.
- Roodman, D. (2009b). How to do xtabond2? An introduction to difference and system GMM in Stata. *Stata Journal 9* (1): 86–136.
- Sanderson, E., and F. Windmeijer (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics 190* (2): 212–221.
- Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica 26* (3): 393–415.
- Windmeijer, F. (2000). Efficiency comparisons for a system GMM estimator in dynamic panel data models. In R. D. H. Heijmans, D. S. G. Pollock, and A. Sattora (Eds.), *Innovations in multivariate statistical analysis: A festschrift for Heinz Neudecker*, Ch. 11: 175–184.
- Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics 126* (1): 25–51.
- Windmeijer, F. (2018). Testing over- and underidentification in linear models, with applications to dynamic panel data and asset-pricing models. *Economics Discussion Paper* 18/696, University of Bristol.
- White, H. L. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50* (1): 1–25.