# xtseqreg: Sequential (two-stage) estimation of linear panel data models

## and some pitfalls in the estimation of dynamic panel models

Sebastian Kripfganz

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

German Stata Users Group Meeting

Berlin, June 23, 2017

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

# Time-invariant regressors in linear panel models

- In many applications, important determinants of the outcome variable can be time invariant.
  - Education, gender, nationality, ethnic and religious background, and other individual-specific characteristics play important roles in the determination of labor market or health outcomes.
  - Institutional, socio-economic, and geographic factors matter in convergence models of economic growth, and they are key variables in gravity models of international trade and investment flows.
- A researcher might be particularly interested in their effects. Yet, traditional "fixed-effects" procedures (`xtreg, fe`) wipe out all time-invariant variables from the model. In contrast, "random-effects" estimators (`xtreg, re`) rely on exogeneity assumptions that are often too strong to be acceptable.

# Time-invariant regressors in linear panel models

- Intermediate solutions include the correlated random-effects model (Mundlak, 1978; Chamberlain, 1982) and similar "hybrid" models (Allison, 2009). See Schunck (2013) for a discussion. Schunck and Perales (2017) recently provided the Stata implementation `xthybrid`.
- If some time-invariant regressors are allowed to be correlated with the unobserved effects while some time-varying regressors are not, the Hausman and Taylor (1981) estimator might be applicable, implemented in Stata as `xthtaylor`.
- Both strategies can be manually implemented with `xtreg` or `xtivreg`.

# Time-invariant regressors in linear panel models

- All procedures so far rely on the assumption of strictly exogenous regressors with respect to the idiosyncratic error component. In the presence of predetermined variables (e.g. a lagged dependent variable) or endogenous variables, additional internal or external instruments might be needed.
- In the context of dynamic panel models, generalized method of moments (GMM) estimators in the spirit of Arellano and Bover (1995) and Blundell and Bond (1998) are frequently employed, implemented in Stata as `xtdpd`, `xtdpdsys`, and the user-written command `xtabond2` (Roodman, 2009).
- If the interest is on the coefficients of time-invariant regressors, the Arellano and Bond (1991) GMM estimator (`xtabond`) is not helpful because all time-invariant variables are removed by a first-difference transformation.

# Time-invariant regressors in linear panel models

- To identify the coefficients of time-invariant regressors, the assumption that a sufficient number of regressors (or excluded instrumental variables) is uncorrelated with the unit-specific error component cannot be avoided.
- Incorrect assumptions about the exogeneity of some variables may cause inconsistency of all coefficient estimates.
- A sequential procedure can provide partial robustness to such misspecification. In a first stage, only the coefficients of time-varying regressors are estimated. In a second stage, the coefficients of time-invariant regressors are recovered.

⇒ New Stata command: `xtseqreg`

# Two-stage estimation

- Linear panel data model with time-invariant regressors and error-components structure:

$$y_{it} = \mathbf{x}_{it}'\beta + \mathbf{f}_i'\gamma + u_i + e_{it}$$

- Sequential estimation procedure:
  1. Estimation of the coefficients of time-varying regressors:

$$y_{it} = \mathbf{x}_{it}'\beta + \tilde{u}_i + e_{it}, \quad \tilde{u}_i = \mathbf{f}_i'\gamma + u_i$$

  2. Estimation of the coefficients of time-invariant regressors:

$$y_{it} - \mathbf{x}_{it}'\hat{\beta} = \mathbf{f}_i'\gamma + u_i + \tilde{e}_{it}, \quad \tilde{e}_{it} = e_{it} - \mathbf{x}_{it}'(\hat{\beta} - \beta)$$

- Conventional standard errors at the second stage are incorrect and often far too small.

⇒ `xtseqreg` computes proper standard errors with the analytical correction term derived by Kripfganz and Schwarz (2015).

# Stata syntax of the xtseqreg command

**Syntax**

```
xtseqreg depvar [(indepvars1)] [indepvars2] [if] [in] [, options]
```

| *options* | Description |
|---|---|
| Model | |
| **first**(*first_spec*) | specify first-stage estimation results |
| **both** | estimate both stages |
| **nocommonesample** | do not restrict estimation samples to be the same |
| **iv**(*iv_spec*) | standard instruments; can be specified more than once |
| **gmmiv**(*gmmiv_spec*) | GMM-type instruments; can be specified more than once |
| **wmatrix**(*wmat_spec*) | specify initial weighting matrix |
| **twostep** | compute two-step instead of one-step estimator |
| **teffects** | add time effects to the model |
| **noconstant** | suppress constant term |
| SE/Robust | |
| **vce**(*vcetype*) | *vcetype* may be **conventional**, **ec**, or **robust** |
| Reporting | |
| **combine** | combine the estimation results for both equations |
| **level**(#) | set confidence level; default is **level(95)** |
| **noheader** | suppress output header |
| **notable** | suppress coefficient table |
| **noomitted** | suppress omitted variables |

# Stata syntax of xtseqreg postestimation commands

**Syntax for predict**

```
predict [type] newvar [if] [in] [, xb stdp ue xbu u e equation(eqno)]

predict [type] {stub*|newvar1 ... newvarq} [if] [in] , scores
```

**Syntax for estat**

Arellano-Bond test for autocorrelated residuals

```
estat serial [, ar(numlist)]
```

Hansen's J-test of overidentifying restrictions

```
estat overid
```

Difference-in-Hansen test of overidentifying restrictions

```
estat overid name
```

Generalized Hausman test for model misspecification

```
estat hausman name [(varlist)] [, df(#) nonested]
```

where *name* is a name under which estimation results were stored via **estimates store**.

# Empirical example: distance and FDI

- Estimation of a gravity model for U.S. outward FDI.
- Annual data, 1989–1999, for 341 bilateral industry-level relationships, compiled by Egger and Pfaffermayr (2004).

```
. describe

Contains data from C:\data_us.dta
  obs:         2,767                          Egger and Pfaffermayr (2004, JAE)
 vars:            13                          8 Aug 2003 03:39
 size:       118,981
--------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
--------------------------------------------------------------------------------
ind             byte   %9.0g                  industry identifier
codeim          int    %8.0g                  country identifier
year            int    %9.0g                  year
lrfdi           float  %9.0g                  log real outward foreign direct investment
lgdt            float  %9.0g                  log bilateral gross domestic product
lsimi           float  %9.0g                  log similarity in country size
lrk             float  %9.0g                  log relative physical capital endowment
lrh             float  %9.0g                  log relative human capital endowment
lrl             float  %9.0g                  log relative labor endowment
ldist           float  %9.0g                  log geographical distance
lkgdt           float  %9.0g                  = lgdt * abs(lrk)
lkldist         float  %9.0g                  = ldist * (lrk - lrl)
id              int    %9.0g                  group(codeim ind)
--------------------------------------------------------------------------------
Sorted by: id  year
```

# First-stage system GMM estimation

```
. xtseqreg L(0/1).lrfdi lkldist lgdt lkgdt lsimi lrk lrh lrl, twostep vce(robust) ///
> gmmiv(L.lrfdi, lag(1 5) collapse model(difference)) ///
> gmmiv(lkldist lgdt lkgdt lsimi lrk lrh lrl, lag(0 4) collapse model(difference)) ///
> iv(L.lrfdi, difference model(level)) ///
> iv(lkldist lgdt lkgdt lsimi lrk lrh lrl, difference model(level))

Group variable: id                          Number of obs         =       2198
Time variable: year                         Number of groups      =        337

                                            Obs per group:    min =          1
                                                              avg =   6.522255
                                                              max =         10

                                            Number of instruments =         49

                                       (Std. Err. adjusted for clustering on id)
```

| lrfdi | Coef. | WC-Robust Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| lrfdi | | | | | | |
| L1. | .8956164 | .063313 | 14.15 | 0.000 | .7715252 | 1.019708 |
| lkldist | -.0978499 | .1490779 | -0.66 | 0.512 | -.3900371 | .1943374 |
| lgdt | -.1502013 | .2320426 | -0.65 | 0.517 | -.6049964 | .3045939 |
| lkgdt | .0072154 | .0053281 | 1.35 | 0.176 | -.0032276 | .0176584 |
| lsimi | .3100215 | .2370884 | 1.31 | 0.191 | -.1546632 | .7747062 |
| lrk | .7471581 | 1.291878 | 0.58 | 0.563 | -1.784877 | 3.279193 |
| lrh | -.0897363 | .1311771 | -0.68 | 0.494 | -.3468386 | .1673661 |
| lrl | -.8973519 | 1.30242 | -0.69 | 0.491 | -3.450048 | 1.655344 |
| _cons | 4.926161 | 5.971464 | 0.82 | 0.409 | -6.777694 | 16.63002 |

```
. estimates store gmm1
```

# First-stage system GMM estimation

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

Arellano-Bond test for autocorrelation of the first-differenced residuals
H0: no autocorrelation of order 1:     z =   -7.3012   Prob > |z|  =    0.0000
H0: no autocorrelation of order 2:     z =   -0.0535   Prob > |z|  =    0.9573
H0: no autocorrelation of order 3:     z =   -0.3725   Prob > |z|  =    0.7095

. estat overid

Hansen's J-test                                        chi2(40)    =   45.7042
H0: overidentifying restrictions are valid             Prob > chi2 =    0.2471
```

- Replication with `xtabond2`:

```
. xtabond2 L(0/1).lrfdi lkldist lgdt lkgdt lsimi lrk lrh lrl, twostep robust ar(3) ///
>   gmm(lrfdi, lag(2 6) collapse equation(diff)) ///
>   gmm(lkldist lgdt lkgdt lsimi lrk lrh lrl, lag(0 4) collapse equation(diff)) ///
>   iv(LD.lrfdi, equation(level) mz) ///
>   iv(D.lkldist D.lgdt D.lkgdt D.lsimi D.lrk D.lrh D.lrl, equation(level) mz)


------------------------------------------------------------------------------
Arellano-Bond test for AR(1) in first differences: z =  -6.69  Pr > z =  0.000
Arellano-Bond test for AR(2) in first differences: z =  -0.05  Pr > z =  0.957
Arellano-Bond test for AR(3) in first differences: z =  -0.37  Pr > z =  0.709
------------------------------------------------------------------------------
Sargan test of overid. restrictions: chi2(40)    =  80.12  Prob > chi2 =  0.000
  (Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(40)    =  45.70  Prob > chi2 =  0.247
  (Robust, but weakened by many instruments.)
```

# How (not) to do xtabond2: Always double check!

- The first two specifications yield identical estimation results. The results from the last specification differ (but should not):

```
. xtseqreg L(0/1).lrfdi lkldist lgdt lkgdt lsimi lrk lrh lrl, twostep vce(robust) ///
>  gmmiv(L.lrfdi, lag(1 5) collapse model(difference)) ///
>  gmmiv(lkldist lgdt lkgdt lsimi lrk lrh lrl, lag(0 4) collapse model(difference)) ///
>  iv(L.lrfdi, difference model(level)) ///
>  iv(lkldist lgdt lkgdt lsimi lrk lrh lrl, difference model(level))
```

```
. xtabond2 L(0/1).lrfdi lkldist lgdt lkgdt lsimi lrk lrh lrl, twostep robust ar(3) ///
>  gmm(lrfdi, lag(2 6) collapse equation(diff)) ///
>  gmm(lkldist lgdt lkgdt lsimi lrk lrh lrl, lag(0 4) collapse equation(diff)) ///
>  iv(LD.lrfdi, equation(level) mz) ///
>  iv(D.lkldist D.lgdt D.lkgdt D.lsimi D.lrk D.lrh D.lrl, equation(level) mz)
```

```
. xtabond2 L(0/1).lrfdi lkldist lgdt lkgdt lsimi lrk lrh lrl, twostep robust ar(3) ///
>  gmm(L.lrfdi, lag(1 5) collapse equation(diff)) ///
>  gmm(lkldist lgdt lkgdt lsimi lrk lrh lrl, lag(0 4) collapse equation(diff)) ///
>  iv(LD.lrfdi, equation(level) mz) ///
>  iv(D.lkldist D.lgdt D.lkgdt D.lsimi D.lrk D.lrh D.lrl, equation(level) mz)
```

# Second-stage 2SLS estimation

```
. xtseqreg lrfdi (L.lrfdi lkldist lgdt lkgdt lsimi lrk lrh lrl) ldist, vce(robust) ///
> first(gmm1, nocons) iv(lsimi lrh)

Group variable: id                              Number of obs      =      2198
Time variable: year                             Number of groups   =       337

------------------------------------------------------------------------------
Equation _first                                 Equation _second
Number of obs         =      2198               Number of obs         =      2198
Number of groups      =       337               Number of groups      =       337

Obs per group:    min =         1               Obs per group:    min =         1
                  avg =  6.522255                                 avg =  6.522255
                  max =        10                                 max =        10

Number of instruments =        49               Number of instruments =         3

                                   (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
       lrfdi |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_first       |
       lrfdi |
         L1. |   .8956164    .063313    14.15   0.000     .7715252    1.019708
             |
     lkldist |  -.0978499   .1490779    -0.66   0.512    -.3900371    .1943374
        lgdt |  -.1502013   .2320426    -0.65   0.517    -.6049964    .3045939
       lkgdt |   .0072154   .0053281     1.35   0.176    -.0032276    .0176584
       lsimi |   .3100215   .2370884     1.31   0.191    -.1546632    .7747062
         lrk |   .7471581   1.291878     0.58   0.563    -1.784877    3.279193
         lrh |  -.0897363   .1311771    -0.68   0.494    -.3468386    .1673661
         lrl |  -.8973519    1.30242    -0.69   0.491    -3.450048    1.655344
-------------+----------------------------------------------------------------
_second      |
       ldist |  -.1213967   .5854263    -0.21   0.836    -1.268811    1.026018
       _cons |   5.966496     8.5777     0.70   0.487    -10.84549    22.77848
------------------------------------------------------------------------------
```

# Second-stage 2SLS estimation

```
. estat overid

Hansen's J-test for equation _first                   chi2(40)     =   45.7042
H0: overidentifying restrictions are valid            Prob > chi2  =    0.2471

Hansen's J-test for equation _second                  chi2(1)      =    1.1989
H0: overidentifying restrictions are valid            Prob > chi2  =    0.2735
```

- Replication with `ivregress` (incorrect standard errors):

```
. quietly estimates restore gmm1

. quietly predict residuals, ue

. ivregress 2sls residuals (ldist = lsimi lrh), vce(cluster id)

Instrumental variables (2SLS) regression          Number of obs     =      2,198
                                                  Wald chi2(1)      =       2.15
                                                  Prob > chi2       =     0.1422
                                                  R-squared         =     0.0107
                                                  Root MSE          =     .46723

                                   (Std. Err. adjusted for 337 clusters in id)
------------------------------------------------------------------------------
             |               Robust
   residuals |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       ldist |  -.1213967   .0827107    -1.47   0.142    -.2835066    .0407132
       _cons |   1.040335   .7110881     1.46   0.143    -.3533725    2.434042
------------------------------------------------------------------------------
Instrumented:  ldist
Instruments:   lsimi lrh
```

# One-stage GMM estimation

```
. xtseqreg L(0/1).lrfdi lkldist lgdt lkgdt lsimi lrk lrh lrl ldist, twostep vce(robust) ///
> gmmiv(L.lrfdi, lag(1 5) collapse model(difference)) ///
> gmmiv(lkldist lgdt lkgdt lsimi lrk lrh lrl, lag(0 4) collapse model(difference)) ///
> iv(L.lrfdi, difference model(level)) ///
> iv(lkldist lgdt lkgdt lsimi lrk lrh lrl, difference model(level)) ///
> iv(lsimi lrh)

Group variable: id                          Number of obs         =      2198
Time variable: year                         Number of groups      =       337

                                            Obs per group:    min =         1
                                                              avg =  6.522255
                                                              max =        10

                                            Number of instruments =        51

                                         (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |              WC-Robust
       lrfdi |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lrfdi |
         L1. |    .874835   .0658537    13.28   0.000     .7457641    1.003906
             |
     lkldist |  -.0894573   .1552895    -0.58   0.565    -.3938191    .2149044
        lgdt |   -.100095   .2389068    -0.42   0.675    -.5683437    .3681537
       lkgdt |   .0103749   .0053781     1.93   0.054     -.000166    .0209159
       lsimi |   .3735686   .2467129     1.51   0.130    -.1099798    .8571171
         lrk |   .6246915   1.349609     0.46   0.643    -2.020494    3.269877
         lrh |  -.0007819   .1125051    -0.01   0.994    -.2212878    .2197241
         lrl |  -.7648876    1.37943    -0.55   0.579    -3.468521    1.938746
       ldist |  -.0825973   .1385583    -0.60   0.551    -.3541665    .1889719
       _cons |   4.320648    6.06585     0.71   0.476      -7.5682     16.2095
------------------------------------------------------------------------------

. estat hausman gmm1 (L.lrfdi lkldist lgdt lkgdt lsimi lrk lrh lrl)

Generalized Hausman test                              chi2(1)      =    4.4792
H0: coefficients do not systematically differ         Prob > chi2  =    0.0343
```

# How (not) to do xtabond2: Remember the assumptions!

```
. estat overid gmm1

Difference-in-Hansen test                               chi2(1)      =     2.6932
H0: overidentifying restrictions are valid              Prob > chi2 =     0.1008
```

- Instruments for the first-differenced equation are uncorrelated with time-invariant variables by construction, first-differenced instruments for the level equation by assumption.

⇒ Difference-in-Hansen tests might be based on asymptotically incorrect (or at least debatable) degrees of freedom:

```
. xtabond2 L(0/1).lrfdi lkldist lgdt lkgdt lsimi lrk lrh lrl ldist, twostep robust ///
>  gmm(lrfdi, lag(2 6) collapse equation(diff)) ///
>  gmm(lkldist lgdt lkgdt lsimi lrk lrh lrl, lag(0 4) collapse equation(diff)) ///
>  iv(LD.lrfdi, equation(level) mz) ///
>  iv(D.lkldist D.lgdt D.lkgdt D.lsimi D.lrk D.lrh D.lrl, equation(level) mz) ///
>  iv(lsimi lrh, equation(level) mz)


Difference-in-Hansen tests of exogeneity of instrument subsets:

  iv(lsimi lrh, mz eq(level))
    Hansen test excluding group:     chi2(39)    =  45.95  Prob > chi2 =  0.206
    Difference (null H = exogenous): chi2(2)     =   2.44  Prob > chi2 =  0.295
```

# Alternative first-stage QML estimator

- First-stage QML estimator of Hsiao et al. (2002):

```
. quietly xtdpdqml lrfdi lkldist lgdt lkgdt lsimi lrk lrh lrl, fe mlparam vce(robust)

. xtseqreg lrfdi (L.lrfdi lkldist lgdt lkgdt lsimi lrk lrh lrl) ldist, vce(robust) ///
>  first(, nocons) iv(lsimi lrh) noheader
note: first-stage variable names do not match with coefficient list from xtdpdqml
note: dependent variable D.lrfdi from xtdpdqml does not match with lrfdi

                                  (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
       lrfdi |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_first       |
       lrfdi |
         L1. |   .8000757   .0539962    14.82   0.000     .6942451    .9059062
             |
     lkldist |  -.7160072   .5053811    -1.42   0.157    -1.706536    .2745216
        lgdt |   .4346637   .1907476     2.28   0.023     .0608052    .8085221
       lkgdt |   .0028906   .0068807     0.42   0.674    -.0105954    .0163766
       lsimi |   .3172032   .3605734     0.88   0.379    -.3895076    1.023914
         lrk |   6.152142   4.400668     1.40   0.162    -2.473009    14.77729
         lrh |   .0758457   .0869135     0.87   0.383    -.0945017    .2461931
         lrl |   -5.60704   4.175718    -1.34   0.179     -13.7913    2.577216
-------------+----------------------------------------------------------------
_second      |
       ldist |    2.41061   2.285819     1.05   0.292    -2.069514    6.890734
       _cons |  -31.43894   21.15977    -1.49   0.137    -72.91133    10.03345
------------------------------------------------------------------------------

. estat overid

Hansen's J-test for equation _second                  chi2(1)     =    0.8358
H0: overidentifying restrictions are valid            Prob > chi2 =    0.3606
```

# Alternative first-stage GMM estimator

- First-stage GMM estimator of Ahn and Schmidt (1995):

```
. quietly xtdpdgmm L(0/1).lrfdi lkldist lgdt lkgdt lsimi lrk lrh lrl, twostep noserial ///
> vce(robust) aux gmmiv(L.lrfdi, lag(1 5) collapse model(difference)) ///
> gmmiv(lkldist lgdt lkgdt lsimi lrk lrh lrl, lag(0 4) collapse model(difference))

. xtseqreg lrfdi (L.lrfdi lkldist lgdt lkgdt lsimi lrk lrh lrl) ldist, vce(robust) ///
> first(, copy) iv(lsimi lrh) noheader
note: first-stage standard errors may not be robust

                              (Std. Err. adjusted for clustering on id)
```

| lrfdi | Coef. | Robust Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| _first | | | | | | |
| lrfdi L1. | .8017069 | .1204806 | 6.65 | 0.000 | .5655692 | 1.037845 |
| lkldist | -.2290635 | .7040092 | -0.33 | 0.745 | -1.608896 | 1.150769 |
| lgdt | -.0748559 | .2905325 | -0.26 | 0.797 | -.6442891 | .4945773 |
| lkgdt | -.0186638 | .0112666 | -1.66 | 0.098 | -.0407459 | .0034183 |
| lsimi | .0212282 | .3722118 | 0.06 | 0.955 | -.7082936 | .75075 |
| lrk | 1.784527 | 6.101738 | 0.29 | 0.770 | -10.17466 | 13.74371 |
| lrh | .0299533 | .1551918 | 0.19 | 0.847 | -.2742171 | .3341238 |
| lrl | -1.580551 | 6.123368 | -0.26 | 0.796 | -13.58213 | 10.42103 |
| _cons | 3.642671 | 7.335562 | 0.50 | 0.619 | -10.73477 | 18.02011 |
| _second | | | | | | |
| ldist | .3209373 | 1.580573 | 0.20 | 0.839 | -2.776928 | 3.418803 |
| _cons | -2.761592 | 13.56865 | -0.20 | 0.839 | -29.35565 | 23.83247 |

```
. estat overid

Hansen's J-test for equation _second                chi2(1)      =    2.7079
H0: overidentifying restrictions are valid          Prob > chi2 =    0.0999
```

# Time effects

```
. xtseqreg L(0/1).lrfdi, teffects twostep vce(robust) ///
> gmmiv(L.lrfdi, lag(1 5) collapse model(difference)) iv(L.lrfdi, difference model(level))

Group variable: id                      Number of obs        =      2198
Time variable: year                     Number of groups     =       337

                                        Obs per group:   min =         1
                                                         avg =  6.522255
                                                         max =        10

                                        Number of instruments =       16

                                   (Std. Err. adjusted for clustering on id)

------------------------------------------------------------------------------
             |             WC-Robust
       lrfdi |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lrfdi |
         L1. |   1.015676   .0727146    13.97   0.000     .8731579    1.158194
             |
        year |
        1991 |  -.0975429   .0419594    -2.32   0.020    -.1797819   -.0153039
        1992 |  -.0670002   .0476785    -1.41   0.160    -.1604484    .0264479
        1993 |  -.0945048   .0457007    -2.07   0.039    -.1840766   -.0049331
        1994 |  -.0644637   .0701426    -0.92   0.358    -.2019406    .0730132
        1995 |  -.0513381   .0426408    -1.20   0.229    -.1349125    .0322363
        1996 |  -.0605227   .0481965    -1.26   0.209    -.1549861    .0339408
        1997 |  -.1211606   .0594696    -2.04   0.042    -.2377189   -.0046024
        1998 |  -.1699316   .0552347    -3.08   0.002    -.2781895   -.0616736
        1999 |  -.1261552   .0830178    -1.52   0.129    -.2888672    .0365568
             |
       _cons |   .0937689   .3189754     0.29   0.769    -.5314114    .7189492
------------------------------------------------------------------------------

. estat overid

Hansen's J-test                                    chi2(5)      =    13.2885
H0: overidentifying restrictions are valid         Prob > chi2  =     0.0208
```

# How (not) to do xtabond2: Beware of the dummy trap!

```
. xtabond2 L(0/1).lrfdi i.year, twostep robust ///
> gmm(lrfdi, lag(2 6) collapse equation(diff)) iv(LD.lrfdi, equation(level) mz) ///
> iv(i.year, equation(level))
```

| lrfdi | Coef. | Corrected Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| lrfdi | | | | | | |
| L1. | 1.015676 | .0727146 | 13.97 | 0.000 | .8731579 | 1.158194 |
| year | | | | | | |
| 1989 | 0 | (empty) | | | | |
| 1990 | .0644637 | .0701426 | 0.92 | 0.358 | -.0730132 | .2019406 |
| 1991 | -.0330792 | .0597255 | -0.55 | 0.580 | -.150139 | .0839805 |
| 1992 | -.0025366 | .0513121 | -0.05 | 0.961 | -.1031064 | .0980333 |
| 1993 | -.0300412 | .0579887 | -0.52 | 0.604 | -.1436969 | .0836146 |
| 1994 | 0 | (omitted) | | | | |
| 1995 | .0131256 | .0551362 | 0.24 | 0.812 | -.0949394 | .1211905 |
| 1996 | .003941 | .055217 | 0.07 | 0.943 | -.1042823 | .1121643 |
| 1997 | -.056697 | .0504278 | -1.12 | 0.261 | -.1555337 | .0421398 |
| 1998 | -.1054679 | .04837 | -2.18 | 0.029 | -.2002714 | -.0106643 |
| 1999 | -.0616915 | .0540627 | -1.14 | 0.254 | -.1676525 | .0442694 |
| _cons | .0293052 | .3703467 | 0.08 | 0.937 | -.696561 | .7551714 |

```
Hansen test of overid. restrictions: chi2(3)    =  13.29  Prob > chi2 =  0.004
  (Robust, but weakened by many instruments.)
```

# How (not) to do xtabond2: Always specify equation()!

- Instruments for the time dummies should only be included for the level equation. Asymptotically, the additional instruments for the first-differenced equation are redundant.

⇒ Hansen's J-test is based on incorrect degrees of freedom:

```
. xtabond2 L(0/1).lrfdi i.year, twostep robust ///
> gmm(lrfdi, lag(2 6) collapse equation(diff)) iv(LD.lrfdi, equation(level) mz) ///
> iv(i.year, equation(diff)) iv(i.year, equation(level))

Hansen test of overid. restrictions: chi2(12)   =  14.82  Prob > chi2 =  0.252
  (Robust, but weakened by many instruments.)
```

- Never use the `iv()` option without suboption `equation()`! It is not equivalent to the joint specification of `iv(, equation(diff))` and `iv(, equation(level))`:

```
. xtabond2 L(0/1).lrfdi i.year, twostep robust ///
> gmm(lrfdi, lag(2 6) collapse equation(diff)) iv(LD.lrfdi, equation(level) mz) ///
> iv(i.year)

Hansen test of overid. restrictions: chi2(3)    =  10.79  Prob > chi2 =  0.013
  (Robust, but weakened by many instruments.)
```

# Summary: the new `xtseqreg` package for Stata

- Sequential estimation can provide partial robustness to model misspecification.
- Is is important to compute corrected standard errors at the second stage that account for the first-stage estimation error.
- The new `xtseqreg` Stata command implements this standard error correction for two-stage linear panel data models.
- The two-stage procedure is particularly relevant in the presence of time-invariant regressors, but it can be easily applied to more general settings.

Kripfganz, S., and C. Schwarz (2015). Estimation of linear dynamic panel data models with time-invariant regressors. *ECB Working Paper 1838*. European Central Bank.

```
net install xtseqreg, from(http://www.kripfganz.de/stata/)
help xtseqreg
help xtseqreg postestimation
```

# References


- Ahn, S. C., and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics 68(1)*: 5–27.
- 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.
- Blundell, R., and S. R. Bond (1991). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87(1)*: 115–143.
- Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics 18(1)*, 5–46.
- Egger, P., and M. Pfaffermayr (2004). Distance, trade and FDI: A Hausman-Taylor SUR approach. *Journal of Applied Econometrics 19(2)*, 227–246; data set available from the JAE data archive: `http://qed.econ.queensu.ca/jae/2004-v19.2/egger-pfaffermayr/`.
- Hausman, J. A., and W. E. Taylor (1981). Panel data and unobservable individual effects. *Econometrica 49(6)*, 1377–1398.
- Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics 109(1)*: 107–150.
- Kripfganz, S., and C. Schwarz (2015). Estimation of linear dynamic panel data models with time-invariant regressors. *ECB Working Paper 1838*. European Central Bank.
- Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica 46(1)*, 69–85.
- Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal 9(1)*: 86–136.
- Schunck, D. (2013). Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *Stata Journal 13(1)*: 65–76.
- Schunck, D., and F. Perales (2017). Within- and between-cluster effects in generalized linear mixed models: A discussion of approaches and the xthybrid command. *Stata Journal 17(1)*: 89–115.