Instrumental-variable estimation of large-T panel-data models with common factors

Sebastian Kripfganz[1] Vasilis Sarafidis[2]

[1]University of Exeter Business School, Department of Economics, Exeter, UK

[2]BI Norwegian Business School, Department of Economics, Oslo, Norway

UK Stata Conference

September 9, 2021

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

Common factors in panel data models

- Causal inference from regression estimates is often hampered by omitted variables for which no data are observed.
- With panel data, a popular approach to account for omitted variables, unobserved heterogeneity, and cross-sectional dependence is to assume a common-factor structure for the regression errors: $\gamma_{y,i}' \mathbf{f}_{y,t} + \varepsilon_{it}$
  - Factors $\mathbf{f}_{y,t}$ are a compact way of summarizing the unobserved variation over time that is common for all units (countries, firms, individuals, ...).
  - The corresponding factor loadings $\gamma_{y,i}$ allow for heterogeneous effects on the units' outcome.
- These unobserved factors are typically allowed to be correlated with the observed explanatory variables (which may themselves be driven by common factors).
  - Unit-fixed effects and time-fixed effects are special cases.

Common factors in panel data models

- A common approach to estimating common-factor models is the Pesaran (2006) common correlated effects (CCE) estimator:
  - Unobserved common factors are projected out by observed cross-sectional averages.
  - Stata implementation: `xtdcce2` (Ditzen, 2018).
- An alternative is the iterative principal components (IPC) approach of Bai (2009):
  - Principal components are factored out from the error term using nonlinear optimization techniques.
  - Stata implementation: `regife` (Gomez, 2015).
- These approaches suffer from potential shortcomings such as incidental-parameters bias (and size distortions due to ineffective bias correction), the necessity of additional assumptions, computational complexity, and limited flexibility.

Common factors in panel data models

- Norkute, Sarafidis, Yamagata, and Cui (2021) and Cui, Norkute, Sarafidis, and Yamagata (2021) developed a new two-stage instrumental variables (IV) approach.
  - In the first stage, principal components analysis (PCA) is used to project out common factors from exogenous covariates (and their lags). The defactored covariates are valid instruments.
  - In the second stage, PCA is applied to extract factors from the first-stage residuals and to defactor the entire model. The same instruments as in the first stage remain valid.
  - The approach is implemented in our new `xtivdfreg` package.
- This IV approach is very flexible (e.g. external instruments, different factors for covariates and error term, heterogeneous slopes) and computationally simple due to a linear objective function.
  - (High-dimensional) fixed effects can be partialled out prior to the estimation; `xtivdfreg` utilizes `reghdfe` (Correia, 2016).
  - Unbalanced panel data set are supported.

Determinants of banks' capital adequacy ratios

```
. xtivdfreg L(0/1).CAR size ROA liquidity, absorb(id t) iv(size ROA liquidity, lags(2)) factmax(3)

Defactored instrumental variables estimation

Group variable: id                         Number of obs        =     16200
Time variable: t                           Number of groups     =       300

Number of instruments  =      9            Obs per group    min =        54
Number of factors in X =      1                             avg =        54
Number of factors in u =      1                             max =        54

Second-stage estimator (model with homogeneous slope coefficients)
------------------------------------------------------------------------------
             |               Robust
         CAR | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         CAR |
         L1. |   .3732316   .0315035    11.85   0.000     .3114859    .4349773
             |
        size |  -2.025311   .1770844   -11.44   0.000     -2.37239   -1.678232
         ROA |   .1999087   .0295306     6.77   0.000     .1420297    .2577877
   liquidity |   1.998128   .4538704     4.40   0.000     1.108559    2.887698
       _cons |   29.99368    4.12824     7.27   0.000     21.90248    38.08488
-------------+----------------------------------------------------------------
     sigma_f |  2.0800886   (std. dev. of factor error component)
     sigma_e |   1.115956   (std. dev. of idiosyncratic error component)
         rho |  .77650224   (fraction of variance due to factors)
------------------------------------------------------------------------------
Hansen test of the overidentifying restrictions       chi2(5)      =    7.3151
H0: overidentifying restrictions are valid            Prob > chi2 =    0.1982
```

Determinants of banks' capital adequacy ratios

```
. xtivdfreg l(0/1).CAR size ROA liquidity, absorb(id t) iv(size ROA liquidity, lags(2)) factmax(3) mg

Defactored instrumental variables estimation

Group variable: id                           Number of obs        =     16200
Time variable: t                             Number of groups     =       300

Number of instruments   =       9            Obs per group     min =        54
Number of factors in X =       1                               avg =        54
                                                               max =        54

Mean-group estimator (model with heterogeneous slope coefficients)
------------------------------------------------------------------------------
             |               Robust
         CAR | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         CAR |
         L1. |   .3751735   .0172599    21.74   0.000     .3413447    .4090022
             |
        size |  -2.178075   .1683235   -12.94   0.000    -2.507983   -1.848167
         ROA |   .2142237   .0375084     5.71   0.000     .1407086    .2877388
   liquidity |   1.456521   .2479702     5.87   0.000     .9705085    1.942534
       _cons |   31.90236   2.083698    15.31   0.000     27.81838    35.98633
------------------------------------------------------------------------------
```

Summary

- The new `xtivdfreg` command enables flexible IV estimation of large-N, large-T panel data models with a multifactor error structure. It can accomodate
  - static and dynamic models,
  - homogeneous and heterogeneous slopes,
  - high-dimensional fixed effects,
  - unbalanced panel data,
  - external instruments,
  - and flexible assumptions about the factor structure of the exogenous covariates.
- For further technical details and examples, see the help file and our forthcoming article in the *Stata Journal 21* (3).

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

help xtivdfreg
```

References

- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77* (4): 1229–1279.
- Cui, G., M. Norkuté, V. Sarafidis, and T. Yamagata (2021). Two-stage instrumental variable estimation of linear panel data models with interactive effects. *Econometrics Journal*: forthcoming.
- Correia, S. (2016). Estimating multi-way fixed effect models with `reghdfe`. *Proceedings of the 2016 Stata Conference*: Chicago.
- Ditzen, J. (2018). Estimating dynamic common-correlated effects in Stata. *Stata Journal 18* (3): 585–617.
- Gomez, M. (2015). Stata module to estimate linear models with interactive fixed effects. *Statistical Software Components*: S458042.
- Norkuté, M., V. Sarafidis, T. Yamagata, and G. Cui (2021). Instrumental variable estimation of dynamic linear panel data models with defactored regressors and a multifactor error structure. *Journal of Econometrics 220* (2): 416–446.
- Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica 74* (4): 967–1010.