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

Sebastian Kripfganz\textsuperscript{1} Vasilis Sarafidis\textsuperscript{2}

\textsuperscript{1}University of Exeter Business School, Department of Economics, Exeter, UK

\textsuperscript{2}BI Norwegian Business School, Department of Economics, Oslo, Norway

Stata Economics Virtual Symposium
November 10, 2021

ssc install xtivdfreg
net install xtivdfreg, from(http://www.kripfganz.de/stata/)
Common factors in panel data models

Consider the following (dynamic) panel data model:

\[ y_{it} = \alpha y_{i,t-1} + \beta' x_{it} + u_{it} \]

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:

\[ u_{it} = \gamma'_{y,i} f_{y,t} + \varepsilon_{it} \]

- Factors \( f_{y,t} \) 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.
- 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.

Sebastian Kripfganz and Vasilis Sarafidis (2021)
Common factors in panel data models

- The unobserved factors are typically allowed to be correlated with the observed explanatory variables, which may themselves be driven by common factors:

\[ x_{it} = \Gamma'_{x,i} f_{x,t} + v_{it} \]

- 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.
This IV approach is implemented in our new *xtivdfreg* package. It offers a lot of flexibility and is computationally simple due to a linear objective function.

- External instruments can be incorporated.
- The covariates and the error term can be driven by different factors.
- A model with heterogeneous slopes can be estimated using a mean-group estimator.
- (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)

|                    | Coefficient | std. err. | z    | P>|z|  | [95% conf. interval] |
|--------------------|-------------|-----------|------|------|---------------------|
| CAR                |             |           |      |      |                     |
| L1.                | 0.3732316   | 0.0315035 | 11.85| 0.000| 0.3114859 - 0.4349773|
| size               | -2.025311   | 0.1770844 | -11.44| 0.000| -2.37239 -1.678232  |
| ROA                | 0.1999087   | 0.0295306 | 6.77 | 0.000| 0.1420297 - 0.2577877|
| liquidity          | 1.998128    | 0.4538704 | 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                | 0.77650224  |           |      |      | (fraction of variance due to factors) |

Hansen test of the overidentifying restrictions  
chi2(5) = 7.3151  
Prob > chi2 = 0.1982

Sebastian Kripfganz and Vasilis Sarafidis (2021)
Determinants of banks’ capital adequacy ratios

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

Number of instruments = 9  
Obs per group min = 54  
Number of factors in X = 0  
avg = 54  
Number of factors in u = 0  
max = 54

Second-stage estimator (model with homogeneous slope coefficients)

| CAR | Coefficient | std. err. | z   | P>|z| | [95% conf. interval] |
|-----|-------------|-----------|-----|-----|---------------------|
| CAR |             |           |     |     |                     |
| L1  | .291951     | .1070032  | 2.73| 0.006| .0822287            |
|     |             |           |     |     |                     |
| size| -.388992    | .0839478  | -4.63| 0.000| -.5535267           |
| ROA | .2213907    | .0687908  | 3.22| 0.001| .0865632            |
| liquidity| -.1206136| .376421 | -0.32| 0.749| -.8583851           |
| _cons| 12.55552    | 3.501715  | 3.59| 0.000| 5.692282            |

sigma_f | 0 (std. dev. of factor error component)  
sigma_e | 2.0686632 (std. dev. of idiosyncratic error component)  
rho | 0 (fraction of variance due to factors)

**Hansen test** of the overidentifying restrictions  
chi2(5) = 19.1115  
Prob > chi2 = 0.0018

```
.ivreghdfe CAR size ROA liquidity (L.CAR = L(0/2).(size ROA liquidity)), gmm2s absorb(id t) cluster(id)
(output omitted)
```
Determinants of banks’ capital adequacy ratios

```
. xtitvdfreg 1(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
Time variable: t

| Number of obs = 16200 |
| 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)

| 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 |

Sebastian Kripfganz and Vasilis Sarafidis (2021)
Determinants of banks’ capital adequacy ratios

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

Defactored instrumental variables estimation

Group variable: id
Time variable: t

Number of obs = 16200
Number of groups = 300
Number of instruments = 7
Obs per group min = 54
avg = 54
max = 54

Number of factors in X = *

Mean-group estimator (model with heterogeneous slope coefficients)

```
| Robust                       |
|-----------------------------|
|                           | CAR | Coef.     Std. Err. | z    | P>|z|   | [95% Conf. Interval] |
|-----------------------------|
|                             | CAR |           |      |      |               |
| L1. |                 .3768387 | .0215774 | 17.46 | 0.000 | .3345478     | .4191297     |
| size | -2.199214 | .1688277 | -13.03 | 0.000 | -2.530111     | -1.868318    |
| ROA | .2229961 | .0394674 | 5.65  | 0.000 | .1456415     | .3003508     |
| liquidity | 1.473673 | .2578282 | 5.72  | 0.000 | .9683387     | 1.979007     |
| _cons | 32.13583 | 2.098844 | 15.31 | 0.000 | 28.02217     | 36.24949     |
```

* Number of factors in stage 1:
  1 -> size ROA
  0 -> liquidity
  1 -> size ROA (doubledefact)
The new xtivdfreg command enables flexible IV estimation of large-$N$, large-$T$ panel data models with a multifactor error structure. It can accommodate
- 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 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.