AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

heck for updates

The Stata Journal (2021)
21, Number 3, pp. 659–686 DOI: 10.1177/1536867X211045558

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

Sebastian Kripfganz
University of Exeter
Exeter, UK
S.Kripfganz@exeter.ac.uk

Vasilis Sarafidis
BI Norwegian Business School
Oslo, Norway
and
Monash University
Melbourne, Australia
vasilis.sarafidis@bi.no

Abstract. In this article, we introduce the `xtivdfreg` command, which implements a general instrumental-variables (IV) approach for fitting panel-data models with many time-series observations, T, and unobserved common factors or interactive effects, as developed by Norkute et al. (2021, *Journal of Econometrics* 220: 416–446) and Cui et al. (2020a, ISER Discussion Paper 1101). The underlying idea of this approach is to project out the common factors from exogenous covariates using principal-components analysis and to run IV regression in both of two stages, using defactored covariates as instruments. The resulting two-stage IV estimator is valid for models with homogeneous or heterogeneous slope coefficients and has several advantages relative to existing popular approaches.

In addition, the `xtivdfreg` command extends the two-stage IV approach in two major ways. First, the algorithm accommodates estimation of unbalanced panels. Second, the algorithm permits a flexible specification of instruments.

We show that when one imposes zero factors, the `xtivdfreg` command can replicate the results of the popular Stata `ivregress` command. Notably, unlike `ivregress`, `xtivdfreg` permits estimation of the two-way error-components panel-data model with heterogeneous slope coefficients.

Keywords: st0650, xtivdfreg, xtivdfreg postestimation, large-T panels, two-stage instrumental-variable estimation, common factors, interactive effects, defactoring, cross-sectional dependence, two-way error-components panel-data model, heterogeneous slope coefficients

1 Introduction

The common factor approach is highly popular among panel-data practitioners because it offers a wide scope for controlling for omitted variables and rich sources of unobserved heterogeneity, including models with cross-sectional dependence; see, for example, Chudik and Pesaran (2015), Juodis and Sarafidis (2018), and Sarafidis and Wansbeek (2012, 2021).

For panels where both of the cross-sectional and time-series dimensions (N and T, respectively) tend to be large, popular estimation approaches have been developed by Pesaran (2006) and Bai (2009) known in the literature as common correlated effects

© 2021 StataCorp LLC

(CCE) and iterative principal components (IPC). Both methods involve least squares and project out the common factors using either cross-sectional averages of observables or principal-components analysis (PCA). To date, CCE and IPC have been applied to a large range of empirical areas and have been extended to several additional theoretical settings; see, for example, Su and Jin (2012), Moon and Weidner (2015, 2017), Baltagi, Ka, and Wang (2021), Harding, Lamarche, and Pesaran (2020), Kapetanios, Serlenga, and Shin (2021), and Li, Cui, and Lu (2020), among others.

Recently, Norkute et al. (2021) and Cui et al. (2020) developed a general instrumental-variables (IV) approach for estimating panel regressions with unobserved common factors when N and T are both large. The underlying idea is to project out the common factors from exogenous covariates using PCA and to construct instruments from defactored covariates. This first-stage IV (1SIV) estimator is consistent. In a second stage, the entire model is defactored based on factors extracted from the first-stage residuals, and IV regression is implemented again using the same instruments.

The resulting two-stage instrumental-variables (2SIV) approach combines features from both Pesaran (2006) and Bai (2009). In particular, following Pesaran (2006), the covariates of the model are assumed to be subject to a linear common factor structure. However, following Bai (2009), the common factors are projected out using PCA rather than cross-sectional averages. A major distinctive feature of 2SIV is that it eliminates the common factors from the error term and the regressors *separately* in two stages. In comparison, CCE eliminates the factors from the error and the regressors *jointly*, whereas IPC eliminates only the factors in the error.

2SIV is appealing for several reasons. First, CCE and IPC suffer from incidental parameters bias because an increasing number of parameters needs to be estimated as either T or N grows; see Westerlund and Urbain (2015) and Juodis, Karabiyik, and Westerlund (2021). Therefore, bias correction is required to ensure that inferences remain valid asymptotically. In contrast, 2SIV does not require bias correction in either dimension. This property is important because approximate procedures aiming to recenter the limiting distribution of particular estimators may not be able to fully eliminate all bias terms, especially those of high order; in such cases, substantial size distortions can occur in finite samples. Second, the CCE approach requires the so-called rank condition, which assumes that the number of factors does not exceed the rank of the (unknown) matrix of cross-sectional averages of the unobserved factor loadings. 2SIV does not require such a condition because the factors are estimated using PCA rather than cross-sectional averages. Third, the 2SIV objective function is linear in the parameters, and therefore the method is robust and computationally inexpensive.[1] In comparison, IPC relies on nonlinear optimization, and therefore convergence to the global optimum might not be guaranteed (Jiang et al. Forthcoming). Fourth, 2SIV shares a major attractive feature of CCE over IPC because it permits estimation of panels with heterogeneous slope coefficients. Last, 2SIV allows for endogenous regressors, so long as external instruments are available.

1. In the context of fixed-T panels, a linear generalized method of moments estimator has been proposed by Juodis and Sarafidis (Forthcoming).

In this article, we introduce a new command, `xtivdfreg`, that implements the 2SIV approach and extends it in two major ways. First, the algorithm accommodates estimation of unbalanced panels. To achieve this, we use a variant of the expectation-maximization approach proposed by Stock and Watson (1998) and Bai, Liao, and Yang (2015). Second, the algorithm permits a flexible specification of instruments. In particular, it accommodates cases where 1) the covariates are driven by entirely different factors; 2) the covariates have a different number of factors, including no factors at all; and 3) different lags of defactored covariates are used as instruments.

We show that when one imposes zero factors and requests the 1SIV estimator, the `xtivdfreg` command can replicate the results of the popular `ivregress` command. Essentially, the two-stage least-squares (2SLS) estimator of the two-way error-components panel-data model can be viewed as a special case of the proposed 2SIV approach in that the former does not defactor the instruments. Notably, unlike `ivregress`, `xtivdfreg` permits estimation of the two-way error-components panel-data model with *heterogeneous* slope coefficients.

We illustrate the method with two examples. First, we use a panel dataset consisting of 300 U.S. financial institutions, each one observed over 56 time periods. We attempt to shed some light on the determinants of banks' capital adequacy ratios. The results are compared with those obtained by using popular panel methods, such as the fixed-effects and 2SLS estimators, as well as the CCE estimator of Pesaran (2006). In the second example, we use macrodata used by Eberhardt and Teal (2010) for the estimation of cross-country production functions in the manufacturing sector. The dataset is unbalanced, containing observations on 48 developing and developed countries during the period 1970 to 2002.

The remainder of the article is organized as follows. Section 2, outlines the 2SIV approach developed by Norkute et al. (2021) and Cui et al. (2020) and discusses implementation with unbalanced panel data. Section 3 describes the syntax of the `xtivdfreg` command. Section 4 illustrates the command using real datasets. Section 5 concludes.

2 IV estimation of large panels with common factors

2.1 Models with homogeneous coefficients

We consider the following autoregressive distributed lag panel-data model with homogeneous slopes and a multifactor error structure:[2]

$$y_{it} = \alpha y_{i,t-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + u_{it}; \quad i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T$$

and

$$u_{it} = \boldsymbol{\gamma}'_{y,i}\mathbf{f}_{y,t} + \varepsilon_{it}$$

2. The estimation procedures described in this article apply also to static panels that arise by imposing $\alpha = 0$ or models with higher-order lags of $y_{i,t}$ and $\mathbf{x}_{it}$. Models with heterogeneous slopes are considered in section 2.2.

$|\alpha| < 1$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_K)'$ such that at least one of $\{\beta_k\}_{k=1}^K$ is nonzero, and $\mathbf{x}_{it} = (x_{it}^{(1)}, x_{it}^{(2)}, \ldots, x_{it}^{(K)})'$ is a $K \times 1$ vector of regressors. The error term of the model is composite, where $\mathbf{f}_{y,t}$ and $\boldsymbol{\gamma}_{y,i}$ denote $m_y \times 1$ vectors of true unobserved factors and factor loadings, respectively, and ε_{it} is an idiosyncratic error.

The vector of regressors $\mathbf{x}_{it}$ is assumed to be subject to the following data-generating process:[3]

$$\mathbf{x}_{it} = \boldsymbol{\Gamma}_{x,i}' \mathbf{f}_{x,t} + \mathbf{v}_{it} \tag{1}$$

$\mathbf{f}_{x,t}$ denotes an $m_x \times 1$ vector of true factors, $\boldsymbol{\Gamma}_{x,i} = (\boldsymbol{\gamma}_{1i}, \boldsymbol{\gamma}_{2i}, \ldots, \boldsymbol{\gamma}_{Ki})$ denotes the corresponding $m_x \times K$ factor loading matrix, and $\mathbf{v}_{it} = (\mathrm{v}_{1it}, \mathrm{v}_{2it}, \ldots, \mathrm{v}_{Kit})'$ is an idiosyncratic error term that is assumed to be independent from ε_{it}.[4] Thus, $\mathbf{x}_{it}$ satisfies strict exogeneity with respect to ε_{it}, although it can be endogenous with respect to the total error term, u_{it}, via the factor component. This assumption ensures that one does not need to seek for external instruments. However, as discussed in remark 4, endogeneity with respect to ε_{it} can be allowed straightforwardly, provided there are valid external instruments available for estimation.

Stacking the T observations for each i yields

$$\mathbf{y}_i = \alpha \mathbf{y}_{i,-1} + \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i; \quad \mathbf{u}_i = \mathbf{F}_y \boldsymbol{\gamma}_{y,i} + \boldsymbol{\varepsilon}_i$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, $\mathbf{y}_{i,-1} = L^1 \mathbf{y}_i = (y_{i0}, y_{i1}, \ldots, y_{iT-1})'$ with L^j defined as the jth lag operator, $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT})'$ is $T \times K$, $\mathbf{u}_i = (u_{i1}, u_{i2}, \ldots, u_{iT})'$, $\mathbf{F}_y = (\mathbf{f}_{y,1}, \mathbf{f}_{y,2}, \ldots, \mathbf{f}_{y,T})'$ is $T \times m_y$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})'$. Similarly,

$$\mathbf{X}_i = \mathbf{F}_x \boldsymbol{\Gamma}_{x,i} + \mathbf{V}_i \tag{2}$$

where $\mathbf{F}_x = (\mathbf{f}_{x,1}, \mathbf{f}_{x,2}, \ldots, \mathbf{f}_{x,T})'$ is a $T \times m_x$ matrix and $\mathbf{V}_i = (\mathbf{v}_{i1}, \mathbf{v}_{i2}, \ldots, \mathbf{v}_{iT})'$ is $T \times K$.[5]

Let $\mathbf{W}_i = (\mathbf{y}_{i,-1}, \mathbf{X}_i)$ and $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}')'$. The model can be written more succinctly as

$$\mathbf{y}_i = \mathbf{W}_i \boldsymbol{\theta} + \mathbf{u}_i$$

The 2SIV approach involves two stages. In the first stage, the common factors in $\mathbf{X}_i$ are asymptotically eliminated using PCA, and the defactored regressors are used as instruments to obtain consistent estimates of the structural parameters of the model, $\boldsymbol{\theta}$.

3. As in Pesaran (2006), (1) implies a genuine restriction on the data-generating process, which is not actually required for the IPC estimator of Bai (2009). However, while this assumption is typically taken for granted by practitioners when using CCE, it is testable within the 2SIV framework, based on the overidentifying restrictions test statistic that is readily available in overidentified models. This issue is discussed in more detail at the end of section 2.1.
4. Individual-specific and time-specific effects can be easily accommodated by replacing $\{y_{it}, \mathbf{x}_{it}'\}$ with the transformed variables $\{\dot{y}_{it}, \dot{\mathbf{x}}_{it}'\}$, where $\dot{y}_{it} = y_{it} - \overline{y}_i - \overline{y}_t + \overline{y}$, $\overline{y}_i = T^{-1} \sum_{t=1}^T y_{it}$, $\overline{y}_t = N^{-1} \sum_{i=1}^N y_{it}$, $\overline{y} = N^{-1} \sum_{i=1}^N \overline{y}_i$, and $\dot{\mathbf{x}}_{it}$ is defined analogously.
5. In practice, it is not necessary that all regressors be subject to a common factor structure and thus correlated with the factor component of the error term, u_{it}. We discuss one such situation in section 4.1, remark 7.

In the second stage, the entire model is defactored based on estimated factors extracted from the first-stage residuals, and another IV regression is implemented using the same instruments as in stage one.

2.1.1 First-stage IV estimator

Define $\widehat{\mathbf{F}}_x$ as $\sqrt{T}$ times the eigenvectors corresponding to the m_x largest eigenvalues of the $T \times T$ matrix $\sum_{i=1}^{N} \mathbf{X}_i \mathbf{X}_i' / NT$. Also, let $\widehat{\mathbf{F}}_{x,-1}$ denote a matrix defined similarly, except that it is based on $\sum_{i=1}^{N} \mathbf{X}_{i,-1} \mathbf{X}_{i,-1}' / NT$, where $\mathbf{X}_{i,-1} = L^1 \mathbf{X}_i$.[6]

Consider the following empirical projection matrices:

$$\mathbf{M}_{\widehat{F}_x} = \mathbf{I}_T - \widehat{\mathbf{F}}_x \left(\widehat{\mathbf{F}}_x' \widehat{\mathbf{F}}_x \right)^{-1} \widehat{\mathbf{F}}_x'; \quad \mathbf{M}_{\widehat{F}_{x,-1}} = \mathbf{I}_T - \widehat{\mathbf{F}}_{x,-1} \left(\widehat{\mathbf{F}}_{x,-1}' \widehat{\mathbf{F}}_{x,-1} \right)^{-1} \widehat{\mathbf{F}}_{x,-1}' \tag{3}$$

In this case, the matrix of instruments can be formulated as

$$\widehat{\mathbf{Z}}_i = \left(\mathbf{M}_{\widehat{F}_x} \mathbf{X}_i, \mathbf{M}_{\widehat{F}_{x,-1}} \mathbf{X}_{i,-1} \right) \tag{4}$$

which is of dimension $T \times 2K$. Thus, the degree of overidentification of the model is $2K - (K+1)$.

Remark 1. Further lags of $\mathbf{X}_i$ can be used as instruments straightforwardly. To illustrate, let q_z denote the total number of lags of $\mathbf{X}_i$ used as instruments, and define $\widehat{\mathbf{F}}_{x,-\tau}$ as $\sqrt{T}$ times the eigenvectors corresponding to the m_x largest eigenvalues of the $T \times T$ matrix $\sum_{i=1}^{N} \mathbf{X}_{i,-\tau} \mathbf{X}_{i,-\tau}' / NT$, where $\mathbf{X}_{i,-\tau} = L^\tau \mathbf{X}_i$ for $\tau = 1, \ldots, q_z$. The corresponding empirical projection matrices are of the same form as in (3) with $\widehat{\mathbf{F}}_{x,-1}$ replaced by $\widehat{\mathbf{F}}_{x,-\tau}$. Moreover, in the case where the covariates are strictly exogenous, leads of $\mathbf{X}_i$ can also be used as instruments; see remark 8 in section 4.1 for more details. In the absence of any lags of $\mathbf{X}_i$ (and further lags of $\mathbf{y}_i$) included in the model as regressors, the degree of overidentification is equal to $q_z K - (K+1)$.

The 1SIV estimator of $\boldsymbol{\theta}$ is defined as

$$\widehat{\boldsymbol{\theta}}_{\text{1SIV}} = \left(\widehat{\mathbf{A}}_{NT}' \widehat{\mathbf{B}}_{NT}^{-1} \widehat{\mathbf{A}}_{NT} \right)^{-1} \widehat{\mathbf{A}}_{NT}' \widehat{\mathbf{B}}_{NT}^{-1} \widehat{\mathbf{g}}_{NT} \tag{5}$$

where

$$\widehat{\mathbf{A}}_{NT} = \frac{1}{NT} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i' \mathbf{W}_i; \quad \widehat{\mathbf{B}}_{NT} = \frac{1}{NT} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i' \widehat{\mathbf{Z}}_i; \quad \widehat{\mathbf{g}}_{NT} = \frac{1}{NT} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i' \mathbf{y}_i$$

The 1SIV estimator is $\sqrt{NT}$ consistent; that is,

$$\sqrt{NT} \left(\widehat{\boldsymbol{\theta}}_{\text{1SIV}} - \boldsymbol{\theta} \right) = O_p(1)$$

6. In this section, both m_y and m_x are treated as known. In practice, these quantities can be estimated consistently using standard methods proposed in the literature, such as the information criteria proposed by Bai and Ng (2002) or the eigenvalue ratio test of Ahn and Horenstein (2013). The `xtivdfreg` command uses the latter.

as N and T grow jointly to infinity, that is, $(N,T) \xrightarrow{j} \infty$, such that $N/T \to c$, $0 < c < \infty$. However, $\widehat{\boldsymbol{\theta}}_{1\text{SIV}}$ is asymptotically biased. Rather than bias correcting this estimator, Norkute et al. (2021) and Cui et al. (2020) put forward a second-stage estimator, which is free from asymptotic bias and is potentially more efficient. For this purpose, the first-stage estimator is useful because it provides a consistent estimate of the error term of the model, which is required to implement the second-stage IV estimator.

Remark 2. In the static panel case, where no lags of $\mathbf{y}_i$ are included on the right-hand side *and* the model is exactly identified (that is, no lags of the regressors are used as instruments), the 1SIV estimator reduces to

$$\widehat{\boldsymbol{\theta}}_{1\text{SIV}} = \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M}_{\widehat{F}_x} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M}_{\widehat{F}_x} \mathbf{y}_i$$

2.1.2 Second-stage IV estimator

To implement the second stage, extract estimates of the space spanned by $\mathbf{F}_y$ using residuals from the first stage; that is,

$$\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{W}_i \widehat{\boldsymbol{\theta}}_{1\text{SIV}}$$

Subsequently, the entire model is defactored, and a second IV regression is run using the same instruments as in stage one.

In particular, let

$$\mathbf{M}_{\widehat{F}_y} = \mathbf{I}_T - \widehat{\mathbf{F}}_y \left(\widehat{\mathbf{F}}_y' \widehat{\mathbf{F}}_y\right)^{-1} \widehat{\mathbf{F}}_y'$$

where $\widehat{\mathbf{F}}_y$ is defined as $\sqrt{T}$ times the eigenvectors corresponding to the m_y largest eigenvalues of the $T \times T$ matrix $\sum_{i=1}^{N} \widehat{\mathbf{u}}_i \widehat{\mathbf{u}}_i' / NT$.

The (optimal) second-stage IV estimator is defined as

$$\widehat{\theta}_{2\text{SIV}} = \left(\widehat{\widehat{\mathbf{A}}}_{NT}' \widehat{\widehat{\boldsymbol{\Omega}}}_{NT}^{-1} \widehat{\widehat{\mathbf{A}}}_{NT}\right)^{-1} \widehat{\widehat{\mathbf{A}}}_{NT}' \widehat{\widehat{\boldsymbol{\Omega}}}_{NT}^{-1} \widehat{\widehat{\mathbf{g}}}_{NT} \tag{6}$$

where

$$\widehat{\widehat{\mathbf{A}}}_{NT} = \frac{1}{NT} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i' \mathbf{M}_{\widehat{F}_y} \mathbf{W}_i; \quad \widehat{\widehat{\mathbf{g}}}_{NT} = \frac{1}{NT} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i' \mathbf{M}_{\widehat{F}_y} \mathbf{y}_i$$

and

$$\widehat{\widehat{\boldsymbol{\Omega}}}_{NT} = \frac{1}{NT} \sum_{i=1}^{N} \widehat{\mathbf{Z}}_i' \mathbf{M}_{\widehat{F}_y} \widehat{\mathbf{u}}_i \widehat{\mathbf{u}}_i' \mathbf{M}_{\widehat{F}_y} \widehat{\mathbf{Z}}_i \tag{7}$$

As shown by Norkute et al. (2021), $\widehat{\boldsymbol{\theta}}_{\text{2SIV}}$ is $\sqrt{NT}$ consistent and asymptotically normally distributed, such that

$$\sqrt{NT}\left(\widehat{\boldsymbol{\theta}}_{\text{2SIV}} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left\{\mathbf{0}, \left(\mathbf{A}'\boldsymbol{\Omega}^{-1}\mathbf{A}\right)^{-1}\right\}$$

as $(N,T) \xrightarrow{j} \infty$ with $N/T \to c$, $0 < c < \infty$.[7]

Notice that the limiting distribution of $\widehat{\boldsymbol{\theta}}_{\text{2SIV}}$ is correctly centered, and thus no bias correction is required. As demonstrated by Cui et al. (2020), the main intuition of this result lies in that $\mathbf{F}_x\boldsymbol{\Gamma}_{x,i}$ is estimated from $\mathbf{X}_i$, whereas $\mathbf{F}_y\boldsymbol{\gamma}_{y,i}$ is estimated from $\mathbf{u}_i$. Because $\mathbf{V}_i$, $\mathbf{F}_y\boldsymbol{\gamma}_{y,i}$, and $\boldsymbol{\varepsilon}_i$ are independent from one another, any correlations that arise because of the estimation error of $\widehat{\mathbf{F}}_y$ and $\widehat{\mathbf{F}}_x$ are asymptotically negligible.

Remark 3. In the static panel case, where no lags of $\mathbf{y}_i$ are included on the right-hand side *and* the model is exactly identified, the second-stage IV estimator can be expressed as

$$\widehat{\boldsymbol{\theta}}_{\text{2SIV}} = \left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{F}_x}\mathbf{M}_{\widehat{F}_y}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\widehat{F}_x}\mathbf{M}_{\widehat{F}_y}\mathbf{y}_i$$

In this case, proposition 3.2 in Cui et al. (2020) reveals that the second-stage estimator is asymptotically equivalent to a least-squares estimator obtained by regressing $\mathbf{y}_i - \mathbf{F}_y\boldsymbol{\gamma}_{y,i}$ on $\mathbf{X}_i - \mathbf{F}_x\boldsymbol{\Gamma}_{x,i}$. Moreover, the authors show that $\widehat{\boldsymbol{\theta}}_{\text{2SIV}}$ is asymptotically as efficient as the bias-corrected CCE and IPC estimators.

Remark 4. The assumptions imposed thus far imply that $\mathbf{X}_i$ satisfies strict exogeneity with respect to $\boldsymbol{\varepsilon}_i$ because otherwise extracting principal components from $\mathbf{X}_i$ may be invalid. When some of the regressors are endogenous (or weakly exogenous) with respect to ε_{it}, 2SIV requires using external exogenous instruments.[8] To illustrate, let $\mathbf{X}_i = (\mathbf{X}_i^{(\text{exog})}, \mathbf{X}_i^{(\text{endog})})$, where $\mathbf{X}_i^{(\text{exog})}$ and $\mathbf{X}_i^{(\text{endog})}$ refer to the strictly exogenous and endogenous regressors, respectively, which are of dimension $T\times K^{(\text{exog})}$ and $T\times K^{(\text{endog})}$. Furthermore, let $\mathbf{X}_i^{+} = (\mathbf{X}_i^{(\text{exog})}, \mathbf{X}_i^{(\text{ext})})$, a $T \times K^*$ matrix with $K^* = K^{(\text{exog})} + K^{(\text{ext})}$, where $\mathbf{X}_i^{(\text{ext})}$ denotes the matrix of external exogenous covariates. $\mathbf{X}_i^{(\text{ext})}$ can still be correlated with the factor component; that is, it may be subject to a similar data-generating process as in (2). Define $\widehat{\mathbf{F}}_x^{+}$ as $\sqrt{T}$ times the eigenvectors corresponding to the m_x^{+} largest eigenvalues of the $T\times T$ matrix $\sum_{i=1}^{N}\mathbf{X}_i^{+}\left(\mathbf{X}_i^{+}\right)'/NT$. The corresponding projection matrices are defined in the same way as in (3) with $\widehat{\mathbf{F}}_x$ ($\widehat{\mathbf{F}}_{x,-1}$) replaced by $\widehat{\mathbf{F}}_x^{+}$ ($\widehat{\mathbf{F}}_{x,-1}^{+}$). In this case, the matrix of instruments becomes

$$\widehat{\mathbf{Z}}_i = \left(\mathbf{M}_{\widehat{F}_x^{+}}\mathbf{X}_i^{+}, \mathbf{M}_{\widehat{F}_{x,-1}^{+}}\mathbf{X}_{i,-1}^{+}\right) \tag{8}$$

7. One could extend the 1SIV estimator of $\boldsymbol{\theta}$ defined in (5) by defactoring the entire model based on $\mathbf{M}_{\widehat{F}_x}$, that is, by using $\mathbf{M}_{\widehat{F}_x}\widehat{\mathbf{Z}}_i$ instead of $\widehat{\mathbf{Z}}_i$. However, in this case, when the space of $\mathbf{F}_y$ spans the space of $\mathbf{F}_x$, the resulting estimator would be asymptotically equivalent to the existing one defined in (6).
8. If external instruments cannot be found, identification requires that 1) the number of strictly exogenous regressors within $\mathbf{X}_i$ be sufficiently large; and 2) these exogenous regressors be correlated with the endogenous ones so that they (and their lags) serve as informative instruments.

The overidentifying restrictions J-test statistic associated with the second-stage IV estimator is given by

$$J_{NT} = \frac{1}{NT}\left(\sum_{i=1}^{N}\widehat{\widehat{\mathbf{u}}}_i'\mathbf{M}_{\widehat{F}_y}\widehat{\mathbf{Z}}_i\right)\widehat{\widehat{\boldsymbol{\Omega}}}_{NT}^{-1}\left(\sum_{i=1}^{N}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{F}_y}\widehat{\widehat{\mathbf{u}}}_i\right)$$

where $\widehat{\widehat{\mathbf{u}}}_i = \mathbf{y}_i - \mathbf{W}_i\widehat{\boldsymbol{\theta}}_{2\text{SIV}}$ and $\widehat{\widehat{\boldsymbol{\Omega}}}_{NT}$ is defined in (7).

The overidentifying restrictions test is particularly useful in this approach. First, it is expected to pick up a violation of the exogeneity of the defactored covariates with respect to the idiosyncratic error in the equation for $\mathbf{y}_i$. Second, the orthogonality condition of the instruments is violated if the slope vector, $\boldsymbol{\theta}$, is cross-sectionally heterogeneous. In this case, the estimators proposed in this section may become inconsistent, and the J test is expected to reject the null hypothesis asymptotically.

2.2 Models with heterogeneous coefficients

We now turn our focus on models with heterogeneous coefficients. Let

$$\mathbf{y}_i = \mathbf{W}_i\boldsymbol{\theta}_i + \mathbf{u}_i$$

where $\boldsymbol{\theta}_i = (\alpha_i, \boldsymbol{\beta}_i')'$ with $\sup_{1\le i\le N}|\alpha_i| < 1$.

The IV estimator of $\boldsymbol{\theta}_i$ is defined as

$$\widehat{\boldsymbol{\theta}}_{\text{IV},i} = \left(\widetilde{\mathbf{A}}_{i,T}'\widetilde{\mathbf{B}}_{i,T}^{-1}\widetilde{\mathbf{A}}_{i,T}\right)^{-1}\widetilde{\mathbf{A}}_{i,T}'\widetilde{\mathbf{B}}_{i,T}^{-1}\widetilde{\mathbf{g}}_{i,T} \tag{9}$$

where

$$\widetilde{\mathbf{A}}_{i,T} = \frac{1}{T}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{F}_x}\mathbf{W}_i; \quad \widetilde{\mathbf{B}}_{i,T} = \frac{1}{T}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{F}_x}\widehat{\mathbf{Z}}_i; \quad \widetilde{\mathbf{g}}_{i,T} = \frac{1}{T}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{F}_x}\mathbf{y}_i \tag{10}$$

$\widehat{\mathbf{Z}}_i$ is defined in (4), and $\mathbf{M}_{\widehat{F}_x}$ is defined in (3) with $\widehat{\mathbf{F}}_x$ obtained as $\sqrt{T}$ times the eigenvectors corresponding to the m_x largest eigenvalues of the $T\times T$ matrix $\sum_{i=1}^{N}\mathbf{X}_i\mathbf{X}_i'/NT$.

The mean-group instrumental-variables (MGIV) estimator of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}}_{\text{MGIV}} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\theta}}_{\text{IV},i} \tag{11}$$

As shown by Norkute et al. (2021), as $(N,T) \xrightarrow{j} \infty$ such that $N/T \to c$ with $0 < c < \infty$,

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_{\text{MGIV}} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_\eta\right)$$

and

$$\widehat{\boldsymbol{\Sigma}}_\eta - \boldsymbol{\Sigma}_\eta \xrightarrow{p} \mathbf{0}$$

where

$$\widehat{\boldsymbol{\Sigma}}_{\eta} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{IV},i} - \widehat{\boldsymbol{\theta}}_{\mathrm{MGIV}}\right)\left(\widehat{\boldsymbol{\theta}}_{\mathrm{IV},i} - \widehat{\boldsymbol{\theta}}_{\mathrm{MGIV}}\right)'$$

Note that the overidentifying restrictions test statistic is not valid for the model with heterogeneous coefficients.[9]

Remark 5. In the static panel case, where no lags of $\mathbf{y}_i$ are included on the right-hand side *and* the model is exactly identified, the individual-specific IV estimator reduces to

$$\widehat{\boldsymbol{\theta}}_{\mathrm{IV},i} = \left(\mathbf{X}_i'\mathbf{M}_{\widehat{F}_x}\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{M}_{\widehat{F}_x}\mathbf{y}_i$$

Remark 6. When the model contains endogenous regressors, the matrices listed in (10) are given by

$$\widetilde{\mathbf{A}}_{i,T} = \frac{1}{T}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{F}_x^+}\mathbf{W}_i; \quad \widetilde{\mathbf{B}}_{i,T} = \frac{1}{T}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{F}_x^+}\widehat{\mathbf{Z}}_i; \quad \widetilde{\mathbf{g}}_{i,T} = \frac{1}{T}\widehat{\mathbf{Z}}_i'\mathbf{M}_{\widehat{F}_x^+}\mathbf{y}_i$$

where $\widehat{\mathbf{Z}}_i$ is defined in (8).

2.3 Unbalanced panels

When the panel-data model is unbalanced, that is, some observations are missing at random, our procedure needs to be modified to control for the unobserved common factors. Following Stock and Watson (1998) and Bai, Liao, and Yang (2015), we may distinguish between $\mathbf{X}_i$ and $\mathbf{X}_i^*$. $\mathbf{X}_i^*$ is a $T \times K$ matrix containing the true values of the regressors, and it is defined as in (2). Let $x_{i,t}^{*(k)}$ denote the (t,k)th entry of $\mathbf{X}_i^*$, and $\iota_{i,t}^{(k)}$ denote a binary indicator that takes the value unity if the kth variable for individual i at time t is observed and zero otherwise. Thus, we set $x_{i,t}^{(k)} = x_{i,t}^{*(k)}$ if $\iota_{i,t}^{(k)} = 1$ and $x_{i,t}^{(k)}$ is unobserved otherwise, $k = 1, \ldots, K$.[10]

Let $\widehat{\mathbf{f}}_{x,t}^{(0)}$ and $\widehat{\boldsymbol{\gamma}}_{ki}^{(0)}$ denote some initial values for the factors and factor loadings, respectively. Also, let $\overline{T} = \max\{T_1, T_2, \ldots, T_N\}$, where T_i denotes the maximum number of observations for individual i.

9. Using a similar line of argument as that in section 2.1, one could also consider a second-stage estimator by projecting $\mathbf{F}_y$ out from the model asymptotically, that is, formulating $\mathbf{M}_{F_y}\mathbf{y}_i = \mathbf{M}_{F_y}\mathbf{W}_i\boldsymbol{\theta}_i + \mathbf{M}_{F_y}\mathbf{u}_i$ and then estimating $\boldsymbol{\theta}_i$. However, the need to deal with heterogeneous slopes here implies that (the space spanned by) $\mathbf{F}_y$ should be estimated using the residuals from the time-series IV regression, $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{W}_i\widehat{\boldsymbol{\theta}}_{\mathrm{IV},i}$. Because $\widehat{\boldsymbol{\theta}}_{\mathrm{IV},i}$ is $\sqrt{T}$ consistent rather than $\sqrt{NT}$ consistent, the estimation of $\mathbf{F}_y$ may become very inefficient. Note that the estimation of $\mathbf{F}_x$ required for the IV estimator defined in (9) does not suffer from a similar problem, because it can be estimated using the raw data $\{\mathbf{X}_i\}_{i=1}^N$.

10. When individual-specific and time-specific effects are included, $x_{i,t}^{*(k)}$ is replaced by $\dot{x}_{i,t}^{*(k)}$, which is defined similarly to footnote 4. Thus, we set $\dot{x}_{i,t}^{(k)} = \dot{x}_{i,t}^{*(k)}$ if $\iota_{i,t}^{(k)} = 1$ and $\dot{x}_{i,t}^{(k)}$ is unobserved otherwise.

In the first iteration, the values of the regressors are set such that

$$\widehat{x}_{i,t}^{(k,1)} = \begin{cases} x_{i,t}^{*(k)} & \text{if} \quad \iota_{i,t}^{(k)} = 1 \\ \left(\widehat{\mathbf{f}}_{x,t}^{(0)}\right)' \widehat{\boldsymbol{\gamma}}_{ki}^{(0)} & \text{if} \quad \iota_{i,t}^{(k)} = 0 \end{cases}$$

The factors in the first iteration, $\widehat{\mathbf{f}}_{x,t}^{(1)}$, are extracted as $\sqrt{\overline{T}}$ times the eigenvectors corresponding to the m_x largest eigenvalues of the matrix

$$\mathbf{V}_x^{(1)} = \sum_{k=1}^{K} \sum_{i=1}^{N} \widehat{\mathbf{x}}_i^{(k,1)} \left(\widehat{\mathbf{x}}_i^{(k,1)}\right)' / \left(N\overline{T}\right)$$

where $\widehat{\mathbf{x}}_i^{(k,1)} = \left(\widehat{x}_{i,1}^{(k,1)}, \widehat{x}_{i,2}^{(k,1)}, \ldots, \widehat{x}_{i,T}^{(k,1)}\right)'$. The corresponding factor loadings, $\widehat{\boldsymbol{\gamma}}_{ki}^{(1)}$, are the estimated individual-specific coefficients obtained by regressing $\widehat{x}_{i,t}^{(k,1)}$ on $\widehat{\mathbf{f}}_{x,t}^{(1)}$, $k = 1, \ldots, K$.

Subsequent iterations are based on

$$\mathbf{V}_x^{(\ell)} = \sum_{k=1}^{K} \sum_{i=1}^{N} \widehat{\mathbf{x}}_i^{(k,\ell)} \left(\widehat{\mathbf{x}}_i^{(k,\ell)}\right)' / \left(N\overline{T}\right)$$

for $\ell > 1$, until convergence. The convergence criterion is defined with respect to the objective function

$$\widetilde{V}_x^{(\ell)} = \left(N\overline{T}\right)^{-1} \sum_{k=1}^{K} \sum_{t=1}^{T} \sum_{i=1}^{N} \left\{ \widehat{x}_{it}^{(k,\ell)} - \left(\widehat{\mathbf{f}}_{x,t}^{(\ell)}\right)' \widehat{\boldsymbol{\gamma}}_{ki}^{(\ell)} \right\}^2$$

where $\widehat{x}_{it}^{(k,\ell)}$ denotes the estimated value of the kth regressor corresponding to the ℓth iteration for individual i at time t, while $\widehat{\mathbf{f}}_{x,t}^{(\ell)}$ and $\widehat{\boldsymbol{\gamma}}_{ki}^{(\ell)}$ are defined similarly as before.

The initial factor values are determined using a similar eigenvalue problem as outlined previously, this time based on $\mathbf{x}_i^{(k)}$, a column vector of length $\overline{T}$ with missing values replaced by zeros. That is, $\widehat{\mathbf{f}}_{x,t}^{(0)}$ is computed as $\sqrt{\overline{T}}$ times the eigenvectors corresponding to the m_x largest eigenvalues of the matrix

$$\mathbf{V}_x^{(0)} = \sum_{k=1}^{K} \sum_{i=1}^{N} \mathbf{x}_i^{(k)} \left(\mathbf{x}_i^{(k)}\right)'$$

with the (j_1, j_2) entry being divided by the number of summands used when this number is larger than zero.

The same procedure is followed when extracting factors from lagged values of $\mathbf{X}_i$ or from the residuals obtained from the first-stage estimation.[11]

11. In practice, it is possible that the estimated number of factors in the regressors varies across different lags. In this case, we set $\widehat{m}_x$ equal to the maximum estimated value obtained across different lags.

3 The xtivdfreg command

3.1 Syntax

xtivdfreg *depvar* [*indepvars*] [*if*] [*in*] [, absorb(*absvars*)
iv(*varlist* [, fvar(*fvars*) lags(#) factmax(#) [no]eigratio
[no]doubledefact]) factmax(#) [no]eigratio [no]doubledefact fstage
mg iterate(#) ltolerance(#) nodots noconstant level(#) coeflegend
noheader notable *display_options*]

3.2 Options

absorb(*absvars*) specifies categorical variables that identify the fixed effects to be absorbed. Typical use is absorb(*panelvar*) or absorb(*panelvar timevar*) for one-way or two-way fixed effects, respectively.[12]

iv(*varlist* [, fvar(*fvars*) lags(#) factmax(#) [no]eigratio [no]doubledefact]) specifies IV. One can specify as many sets of instruments as required. Variables in the same set are defactored jointly. External variables that are not part of the regression model can also be used as instruments in *varlist*.

fvar(*fvars*) specifies that factors be extracted from the variables in *fvars*. The default is to extract factors from all variables in *varlist*.

lags(#) specifies the # of lags of *varlist* to be added to the set of instruments. The variables at each lag order are defactored separately with factors extracted from the corresponding lag of *fvars*. The default is lags(0).

factmax(#) specifies the maximum number of factors to be extracted from *fvars*. The default is set by the global option factmax(#).

noeigratio and eigratio request either to use a fixed number of factors as specified with the suboption factmax(#) or to use the Ahn and Horenstein (2013) eigenvalue ratio test to compute the number of factors. eigratio is the default unless otherwise specified with the global option noeigratio.

doubledefact requests to include *fvars* in a further defactorization stage of the entire model for the first-stage estimator. All sets of instruments that are included in this defactorization stage are jointly defactored, excluding lags of *fvars* specified with the suboption lags(#). nodoubledefact requests to avoid implementing a further defactorization stage of the entire model for the first-stage estimator. The default is set by the global option [no]doubledefact.

12. This option requires the community-contributed commands reghdfe (Correia 2016) and ftools (Correia 2017) to be installed.

`factmax(#)` specifies the maximum number of factors for each estimation stage and each set of instruments. The default is `factmax(4)`.

`noeigratio` requests to use a fixed number of factors as specified with the option `factmax(#)`. By default, the eigenvalue ratio test of Ahn and Horenstein (2013) is used to compute the number of factors for each estimation stage and each set of instruments.

`doubledefact` requests to use a further defactorization stage of the entire model for the first-stage estimator, as, for example, described in footnote 7. `nodoubledefact` requests to avoid implementing this further defactorization stage. `doubledefact` is the default when the option `mg` is specified, and `nodoubledefact` is the default when the option `mg` is omitted.

`fstage` requests the 1SIV estimator to be computed instead of the second-stage IV estimator.

`mg` requests the mean-group estimator to be computed, which allows for heterogeneous slopes.

`iterate(#)` specifies the maximum number of iterations for the extraction of factors. If convergence is declared before this threshold is reached, it will stop when convergence is declared. The default is the number set using `set maxiter`. This option has no effect with strongly balanced panel data, in which case any iterations are redundant.

`ltolerance(#)` specifies the convergence tolerance for the objective function; see [R] **Maximize**. The default is `ltolerance(1e-4)`. This option has no effect with strongly balanced panel data.

`nodots` requests not to display dots for the iteration steps. By default, one dot character is displayed for each iteration step. This option has no effect with strongly balanced panel data.

`noconstant` suppresses the constant term.

`level(#)`, `coeflegend`; see [R] **Estimation options**.

`noheader` suppresses display of the header above the coefficient table that displays the number of observations and moment conditions.

`notable` suppresses display of the coefficient table.

display_options: `noci`, `nopvalues`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(`*style*`)`, `cformat(%`*fmt*`)`, `pformat(%`*fmt*`)`, `sformat(%`*fmt*`)`, and `nolstretch`; see [R] **Estimation options**.

3.3 Stored results

`xtivdfreg` stores the following in `e()`:

Scalars

`e(N)`	number of observations	`e(rho)`	variance fraction due to factors
`e(df_m)`	model degrees of freedom (DF)	`e(chi2_J)`	Hansen's J statistic
`e(N_g)`	number of groups	`e(df_J)`	DF of the J test
`e(g_min)`	smallest group size	`e(p_J)`	p-value of the J test
`e(g_avg)`	average group size	`e(rank)`	rank of `e(V)`
`e(g_max)`	largest group size	`e(zrank)`	number of instruments
`e(sigma2u)`	variance of u_{it}	`e(fact1)`	number of factors in first stage
`e(sigma2f)`	variance of factor component	`e(fact2)`	number of factors in second stage

Macros

`e(cmd)`	`xtivdfreg`	`e(properties)`	`b V`
`e(cmdline)`	command as typed	`e(estat_cmd)`	`xtivdfreg_estat`
`e(depvar)`	name of dependent variable	`e(predict)`	`xtivdfreg_p`
`e(ivar)`	variable denoting groups	`e(marginsok)`	predictions allowed by `margins`
`e(tvar)`	variable denoting time	`e(estimator)`	`fstage`, `sstage`, or `mg`
`e(vcetype)`	title used to label Std. Err.		

Matrices

`e(b)`	coefficient vector	`e(factnum)`	variable-specific number of factors in first stage
`e(V)`	variance–covariance matrix of the estimators		

Functions

`e(sample)`	marks estimation sample

4 Examples

4.1 Example 1: Estimation of the determinants of banks' capital adequacy ratios

In this example, we illustrate the `xtivdfreg` command by estimating the effect of main drivers behind capital adequacy ratios for banking institutions. We make use of panel data from a random sample of 300 U.S. banks, each one observed over 56 time periods, namely, 2006:Q1–2019:Q4.

We focus on the model

$$\begin{aligned} \mathtt{CAR}_{it} &= \alpha \mathtt{CAR}_{it-1} + \beta_1 \mathtt{size}_{it} + \beta_2 \mathtt{ROA}_{it} + \beta_3 \mathtt{liquidity}_{it} + u_{it} \\ u_{it} &= \eta_i + \tau_t + \boldsymbol{\gamma}_{y,i}' \mathbf{f}_{y,t} + \varepsilon_{it} \end{aligned} \tag{12}$$

where $i = 1, \ldots, 300$ and $t = 2, \ldots, 56$. All data are publicly available, and they have been downloaded from the Federal Deposit Insurance Corporation website.[13]

13. See https://www.fdic.gov/. The dataset is available as an ancillary file for the `xtivdfreg` package.

- $\texttt{CAR}_{it}$ stands for "capital adequacy ratio", which is proxied by the ratio of tier 1 (core) capital over risk-weighted assets.
- $\texttt{size}_{it}$ is proxied by the natural logarithm of banks' total assets.
- $\texttt{ROA}_{it}$ stands for the "return on assets", defined as annualized net income expressed as a percentage of average total assets. `ROA` is used as a measure of profitability.
- $\texttt{liquidity}_{it}$ is proxied by the loan-to-deposit ratio. Note that higher values of this variable imply a lower level of liquidity.

Finally, the error term is composite; η_i and τ_t capture bank-specific and time-specific effects, $\mathbf{f}_{y,t}$ is an $m_y \times 1$ vector of unobserved common shocks with corresponding loadings given by $\boldsymbol{\gamma}_{y,i}$, and ε_{it} is a purely idiosyncratic error. Note that m_y is unknown.

Some discussion on the interpretation of the parameters that characterize (12) is useful. The autoregressive coefficient, α, reflects costs of adjustment that prevent banks from achieving optimal levels of capital adequacy instantaneously. β_k, for $k = 1, \ldots, K(= 3)$, denote the slope coefficients of the model. β_1 measures the effect of size on capital adequacy behavior. Under the "too-big-to-fail hypothesis", large banks may count on public bailout during periods of financial distress, knowing that they are systematically very important (for example, Cui, Sarafidis, and Yamagata [2020b]). Essentially, this hypothesis reflects the classic moral hazard problem, where one party takes on excessive risk, knowing that it is protected against the risk and that another party will incur the cost. Under such a scenario, β_1 is expected to be negative.

β_2 measures the effect of profitability on capital adequacy. Standard theory suggests that higher bank profitability dissuades a bank's risk taking, and thus it is associated with larger capital reserves because profitable banks stand to lose more shareholder value if downside risks realize (Keeley 1990). On the other hand, more profitable banks can borrow more and engage in risky activities on a larger scale under the presence of leverage constraints (Martynova, Ratnovski, and Vlahu 2020). A positive (negative) value of β_2 is consistent with the former (latter) interpretation. Lastly, the direction of the effect of liquidity, β_3, is ultimately an empirical question as well. For instance, a positive value indicates that lower liquidity levels force banks to increase their capital reserves, arguably to reduce risk exposure.

We start by running the `xtivdfreg` command using two lags of the covariates as defactored instruments and up to a maximum of three factors. Thus, we use nine instruments in total, three for each covariate. There are four parameters, which implies that the degree of overidentification equals five. We control for bank-specific and time-specific effects by eliminating them prior to estimation. This baseline regression is obtained as follows:

```
. use xtivdfreg_example
(Capital Adequacy Ratios of U.S. Banking Institutions; Source: FDIC)
. 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)
```

CAR	Coefficient	Robust 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
```

To illustrate the specification of the command in terms of the notation used in the article, let $\mathbf{X}_i = (\mathbf{X}_i^{(1)}, \mathbf{X}_i^{(2)}, \mathbf{X}_i^{(3)})$, where $\mathbf{X}_i^{(k)}$ denotes the regressor corresponding to the coefficient β_k in (12), for $k = 1, 2, 3$. The matrix of instruments is given by

$$\widehat{\mathbf{Z}}_i = \left(\mathbf{M}_{\widehat{F}_x}\mathbf{X}_i, \mathbf{M}_{\widehat{F}_{x,-1}}\mathbf{X}_{i,-1}, \mathbf{M}_{\widehat{F}_{x,-2}}\mathbf{X}_{i,-2}\right)$$

which is of dimension $T \times 3K$, with $\widehat{\mathbf{F}}_{x,-\tau}$ defined as $\sqrt{T}$ times the eigenvectors corresponding to the m_x largest eigenvalues of the $T \times T$ matrix $\sum_{i=1}^{N} \mathbf{X}_{i,-\tau}\mathbf{X}_{i,-\tau}'/NT$, for $\tau = 1, 2$. The second-stage IV estimator is defined in (6).

All coefficients are statistically significant at the 1% level. Moreover, the p-value of the J-test statistic suggests that the overidentifying restrictions (instruments) are valid. The estimated number of factors in the first and second stages equals 1 in both cases; that is, $\widehat{m}_x = \widehat{m}_y = 1$.

`xtivdfreg` also reports the fraction of the variance of u_{it} that is explained by the factor component, denoted as `rho`. Because the value of `rho` is roughly equal to 3/4 in the present sample, it appears that most of the variation in the composite error term is due to the single unobserved factor, conditional on bank-specific and time-specific effects. Therefore, estimators that fail to control for common shocks are likely to be severely biased.

The estimated autoregressive coefficient equals about 0.373, which suggests medium persistence in the `CAR` time series. The estimated coefficient of `size` is highly negative, so it is consistent with the "too-big-to-fail hypothesis", providing evidence of moral hazard-type behavior of banking institutions. Profitability (`ROA`) appears to have a positive effect on capital adequacy, which is in line with Keeley (1990). The positive estimate for β_3 shows that lower levels of bank asset liquidity (that is, higher values of `liquidity`) lead to an increase in capital reserves, all other things being equal. This implies that banking institutions suffering from a liquidity crunch tend to respond by raising their equity.

Finally, note that `xtivdfreg` reports an estimate of a constant term (intercept). This is obtained as the mean of the residuals in a separate step after computing the slope coefficients.[14] Whether a constant term is estimated has no effect on the computation of the slope coefficients because the latter are computed for the demeaned model with or without the absorption of fixed effects. The standard error of the constant term is computed with the influence-function approach of Kripfganz and Schwarz (2019).

Next we fit the same model, except that the slope coefficients are allowed to be heterogeneous:

$$\texttt{CAR}_{it} = \alpha_i \texttt{CAR}_{it-1} + \beta_{1i}\texttt{size}_{it} + \beta_{2i}\texttt{ROA}_{it} + \beta_{3i}\texttt{liquidity}_{it} + u_{it}$$

u_{it} has the same structure as before. This regression is computed by adding the option `mg`. The results correspond to the MGIV estimator defined in (11):

```
. 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)
```

CAR	Coefficient	Robust 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

As we can see, the estimated coefficients are similar to those obtained from the model that pools the data and imposes slope parameter homogeneity. This is not surprising, because otherwise failure to account for slope parameter heterogeneity would invalidate

14. For the model with heterogeneous slopes, the intercept is also treated as heterogeneous.

the overidentifying restrictions, thus likely leading to a rejection of the null hypothesis for the J statistic. Thus, conditional on common factors, bank-specific and time-specific effects, slope parameter heterogeneity does not appear to be relevant in the present sample.

In what follows, we examine alternative specifications for `xtivdfreg` and use other estimators. For exposition, table 1 below includes the results for the previous two baseline specifications (columns 1–2).

Table 1. Estimation results

	2SIV	MGIV	2SIV (2)	2SIV (3)	MGIV (3)
`L.CAR`	0.373***	0.375***	0.379***	0.356***	0.358***
	(0.032)	(0.017)	(0.038)	(0.040)	(0.018)
`size`	−2.025***	−2.178***	−2.174***	−2.088***	−2.235***
	(0.177)	(0.168)	(0.210)	(0.198)	(0.169)
`ROA`	0.200***	0.214***	0.104***	0.212***	0.218***
	(0.030)	(0.038)	(0.027)	(0.037)	(0.038)
`liquidity`	1.998***	1.457***	2.053***	1.930***	1.071***
	(0.454)	(0.248)	(0.501)	(0.452)	(0.242)
N	16200	16200	16200	16200	16200
`e(fact1)`	1	1	1	1	1
`e(fact2)`	1		1	1	
`rho`	0.777		0.758	0.783	
`e(p_J)`	0.198		0.020	0.150	

NOTES: Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Columns 3–5 illustrate examples of IV estimators that allow for a more flexible specification of instruments than the baseline regression. In particular, column 3 shows results for a second-stage IV estimator that involves dropping `ROA` from the set of instruments and using an external variable instead, namely, `ROE`.[15]

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

(*output omitted*)

The results in column 3 are similar to the baseline specification in column 1, except for the coefficient of `ROA`, which is statistically different at the 5% level. Note also that in this case the J-test statistic rejects the null hypothesis because the p-value equals 0.020. This implies that `ROE` may not form a valid instrument.

15. `ROE` stands for the "return on equity", defined as annualized net income expressed as a percentage of total equity on a consolidated basis. `ROE` represents an alternative measure of bank profitability.

Column 4 corresponds to a second-stage IV estimator that we can compute by typing

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

In this specification, {`size`, `ROA`} are defactored based on a common set of factors estimated jointly, whereas `liquidity` is defactored separately, based on its own estimated factors. Such an instrumentation strategy can be particularly useful under three circumstances: first, when `size` and `ROA` are driven by entirely different factors than `liquidity`; second, when `size` and `ROA` have a different number of factors than `liquidity`; and third, when different lags of the covariates are used as instruments. Column 5 corresponds to the same specification as in column 4, although it refers to its MGIV version:

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

As we can see, the output of columns 4–5 is similar to that reported in columns 1–2, respectively. Therefore, the estimates appear to be fairly robust to different choices of instruments.

In terms of the notation used in the article, the choice of instruments corresponding to columns 4–5 is given by

$$\widehat{\mathbf{Z}}_i = \left(\mathbf{M}_{\widehat{F}_{x_{12}}}\mathbf{X}_i^{(1,2)}, \mathbf{M}_{\widehat{F}_{x_{12},-1}}\mathbf{X}_{i,-1}^{(1,2)}, \mathbf{M}_{\widehat{F}_{x_{12},-2}}\mathbf{X}_{i,-2}^{(1,2)}, \mathbf{M}_{\widehat{F}_{x_3}}\mathbf{X}_i^{(3)}, \mathbf{M}_{\widehat{F}_{x_3,-1}}\mathbf{X}_{i,-1}^{(3)}\right)$$

where $\mathbf{X}_i^{(1,2)} = (\mathbf{X}_i^{(1)}, \mathbf{X}_i^{(2)})$, $\widehat{F}_{x_{12},-\tau}$ is defined as $\sqrt{T}$ times the eigenvectors corresponding to the $m_{x_{12}}$ largest eigenvalues of the $T \times T$ matrix $\sum_{i=1}^{N} \mathbf{X}_{i,-\tau}^{(1,2)} \left(\mathbf{X}_{i,-\tau}^{(1,2)}\right)' / NT$, and so on. The column dimension of the matrix of instruments is $T \times 8$. Hence, two lags of $\mathbf{X}_i^{(1,2)}$ and one lag of $\mathbf{X}_i^{(3)}$ are used as instruments. Note also that the maximum numbers of factors specified to be estimated from $\mathbf{X}_i^{(1,2)}$ and $\mathbf{X}_i^{(3)}$ are different and equal 3 and 2, respectively.

Remark 7. For the MGIV estimator, although the matrix $\widehat{\mathbf{Z}}_i$ above is formulated by defactoring $\mathbf{X}_i^{(1,2)}$ and $\mathbf{X}_i^{(3)}$ separately, the empirical projection matrix $\mathbf{M}_{\widehat{F}_x}$ used to defactor the entire model[16] is computed by extracting factors jointly from the matrix of all covariates; that is, $\mathbf{X}_i = (\mathbf{X}_i^{(1)}, \mathbf{X}_i^{(2)}, \mathbf{X}_i^{(3)})$.

In practice, users can avoid extracting factors jointly from the matrix of all covariates. For motivation, suppose that $\mathbf{X}_i^{(3)}$ were a binary regressor that is not subject to a common factor structure. In that case, one may wish to 1) instrument $\mathbf{X}_i^{(3)}$ by itself (that is, without defactoring or lags) and 2) defactor the entire model by extracting factors only from $\mathbf{X}_i^{(1,2)}$, that is, to omit $\mathbf{X}_i^{(3)}$ from the construction of $\mathbf{M}_{\widehat{F}_x}$. This can be achieved by specifying

16. See the expressions in (9)–(10).

```
. 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                          Number of obs         =     16200
Time variable: t                            Number of groups      =       300

Number of instruments  =      7             Obs per group     min =        54
Number of factors in X =      *                               avg =        54
                                                              max =        54

Mean-group estimator (model with heterogeneous slope coefficients)
```

CAR	Coefficient	Robust 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)
```

Defactoring of $\mathbf{X}_i^{(3)}$ is avoided by specifying `factmax(0)`. The omission of $\mathbf{X}_i^{(3)}$ in the construction of $\mathbf{M}_{\widehat{F}_x}$ is achieved by specifying `nodoubledefact`. Note that in this case the number of estimated factors differs across covariates. Therefore, `xtivdfreg` reports `Number of factors in X =` * and provides detailed results on the estimated number of factors at the bottom of the output.

The columns in table 2 report results for several alternative popular estimators. To begin with, columns 1–2 correspond to the standard fixed-effects and 2SLS estimators, both of which accommodate a two-way error-components model, but they do not allow for common shocks:

```
. xtreg l(0/1).CAR size ROA liquidity i.t, fe vce(cluster id)
```

(*output omitted*)

```
. ivregress 2sls CAR size ROA liquidity (L.CAR = l(0/2).(size ROA liquidity))
> i.id i.t, vce(cluster id)
```

(*output omitted*)

Table 2. Estimation results (continued)

	Fixed effect	2SLS	IV no DF	MGIV no DF	CCEP	CCEMG
`L.CAR`	0.878***	0.651***	0.651***	0.410***	0.526***	0.356***
	(0.018)	(0.207)	(0.207)	(0.020)	(0.039)	(0.013)
`size`	−0.085***	−0.220*	−0.220*	−1.096***	−0.423***	−1.396***
	(0.023)	(0.124)	(0.124)	(0.206)	(0.058)	(0.245)
`ROA`	0.101*	0.142	0.142	0.256***	0.130**	0.286***
	(0.055)	(0.131)	(0.131)	(0.047)	(0.056)	(0.040)
`liquidity`	0.205	0.503	0.503	1.581***	1.992***	1.740***
	(0.163)	(0.460)	(0.460)	(0.283)	(0.412)	(0.253)
N	16500	16200	16200	16200	15900	15900
`e(fact1)`			0	0		
`e(fact2)`						
`rho`	0.061					
`e(p_J)`			0.000			

NOTES: Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

It is apparent that the estimated coefficients differ substantially compared with those obtained based on the 2SIV approach. In particular, the autoregressive coefficient appears to be biased upward, and—for the case of 2SLS (column 2)—the standard error of the estimate is much larger compared with the second-stage IV (see column 1 of table 1). On the other hand, the coefficients of `ROA` and `liquidity` seem to be biased in the opposite direction. Moreover, in three out of four cases, these coefficients are not statistically significant. This outcome is indicative of the importance of controlling for common shocks in the present example.

Column 3 reproduces the results of 2SLS using the `xtivdfreg` command. This is achieved by setting the number of factors equal to 0 and requesting the first-stage estimator:

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

Thus, the popular 2SLS estimator can be viewed as a special case of the 2SIV approach and arises by imposing zero number of factors (that is, setting `factmax(0)`) and fitting the model in a single stage (`fstage`). Column 4 yields 2SLS-type results for a model with heterogeneous slopes. Note that this option is not allowed in `ivregress`.

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

Finally, the last two columns correspond to the CCE estimator of Pesaran (2006). CCEP in column 5 denotes the pooled CCE estimator, and CCEMG in column 6 is the

mean-group CCE version. These have been computed using the `xtdcce2` command developed by Ditzen (2018).

```
. xtdcce2 CAR L.CAR size ROA liquidity,
> crosssectional(CAR L.CAR size ROA liquidity)
> cr_lags(2) pooled(L.CAR size ROA liquidity) pooledvce(nw)
  (output omitted)
. xtdcce2 CAR L.CAR size ROA liquidity,
> crosssectional(CAR L.CAR size ROA liquidity)
> cr_lags(2)
  (output omitted)
```

As we can see, the estimates of CCEP are smaller than those obtained by the second-stage IV estimator (columns 1 and 4 of table 1), and the differences are statistically significant. On the other hand, the estimates of CCEMG are fairly close to those of the MGIV estimator in most cases. The main exception is the coefficient of `size`, which appears to be much smaller and less precise for CCEMG.

Remark 8. As pointed out by a referee, when the covariates are strictly exogenous, one can also use leads (as opposed to lags) as instruments. To see this, notice that the following two specifications are equivalent, provided that the number of estimated factors across all different lags of instruments is the same:

```
. xtivdfreg l(0/1).CAR size ROA liquidity, absorb(id t)
> iv(size ROA liquidity, lags(2)) factmax(3)
  (output omitted)
. xtivdfreg l(0/1).CAR size ROA liquidity, absorb(id t) iv(size ROA liquidity)
> iv(L.(size ROA liquidity)) iv(L2.(size ROA liquidity)) factmax(3)
  (output omitted)
```

Hence, one can also use leads by typing

```
. xtivdfreg l(0/1).CAR size ROA liquidity, absorb(id t) iv(size ROA liquidity)
> iv(F.(size ROA liquidity)) iv(F2.(size ROA liquidity)) factmax(3)
  (output omitted)
```

where `F.()` is the lead operator. Note that this equivalence does not hold when a second defactorization step is applied in stage one, that is, either for the model with heterogeneous slope coefficients or when the option `doubledefact` is declared:

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

To achieve equivalence in that case, exclude the lags or leads from the second defactorization, as follows:

```
. xtivdfreg L(0/1).CAR size ROA liquidity, absorb(id t) iv(size ROA liquidity)
> iv(L.(size ROA liquidity), nodoubledefact)
> iv(L2.(size ROA liquidity), nodoubledefact) factmax(3) doubledefact
  (output omitted)
```

4.2 Example 2: Estimation of cross-country production functions

To illustrate additional features of the `xtivdfreg` command for unbalanced panels, we use the macropanel dataset of Eberhardt and Teal (2010) for estimating cross-country production functions in the manufacturing sector. The dataset contains observations on 48 developing and developed countries during the period 1970 to 2002. These data are available as an ancillary file for the `xtmg` package, developed by Eberhardt (2012):

```
. use http://www.stata-journal.com/software/sj12-1/st0246/manu_prod
(Manufacturing productivity analysis (1970-2002))
```

Following Eberhardt and Teal (2010), we focus on the following model, which imposes constant returns to scale:

$$\ln\left(\frac{Y_{it}}{L_{it}}\right) = \beta \ln\left(\frac{K_{it}}{L_{it}}\right) + u_{it} \tag{13}$$
$$u_{it} = \eta_i + \tau_t + \boldsymbol{\gamma}'_{yi}\mathbf{f}_{y,t} + \varepsilon_{it}$$

The dependent and independent variables denote the log value added per worker and the log capital stock per worker, respectively, for $i = 1, \ldots, 48$, with each country observed over T_i observations.

We start by running the `xtivdfreg` command using two lags of the covariates as defactored instruments and up to a maximum of three factors. We use three instruments, and the degree of overidentification equals two. We control for bank-specific and time-specific effects by eliminating them prior to estimation. This baseline regression is computed by typing

```
. xtivdfreg ly lk, absorb(list year) iv(lk, lags(2)) factmax(3)
Defactored instrumental variables estimation
Factor estimation:
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
...                                                    3
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
...                                                    3
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
...                                                    3
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
...                                                    3
Group variable: list                        Number of obs         =      1084
Time variable: year                         Number of groups      =        48
Number of instruments  =      3             Obs per group     min =         9
Number of factors in X =      1                               avg =  22.58333
Number of factors in u =      1                               max =        31
Second-stage estimator (model with homogeneous slope coefficients)
```

ly	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
lk	.4835692	.0733879	6.59	0.000	.3397315	.6274069
_cons	4.569137	.7842979	5.83	0.000	3.031942	6.106333
sigma_f	.07592639	(std. dev. of factor error component)				
sigma_e	.1151073	(std. dev. of idiosyncratic error component)				
rho	.30317989	(fraction of variance due to factors)				

```
Hansen test of the overidentifying restrictions        chi2(2)     =    4.1397
H0: overidentifying restrictions are valid             Prob > chi2 =    0.1262
```

Because the panel is unbalanced, the `xtivdfreg` command estimates the factors based on the iterative procedure described in section 2.3. The first line of dots reports the number of iterations required to estimate the factors in $\ln(K/L)$. The second (third) line of dots reports the number of iterations required to estimate the number of factors in the lagged (second lagged) value of $\ln(K/L)$. Finally, the last line corresponds to factor estimation from the first-stage residuals, which is relevant for the 2SIV estimator. In all cases, three iterations turn out to be sufficient for convergence.

The estimated coefficient of $\ln(K/L)$ is approximately equal to 0.5 and is statistically significant. The p-value of the J-test statistic indicates that the overidentifying restrictions are supported by the data. Moreover, $\widehat{m}_x = \widehat{m}_y = 1$, whereas the fraction of the variance of u_{it} that is explained by the factor component appears to be around 0.3.

Note that the number of lines of dots not only is a function of the number of lags used as instruments but also depends on whether factors are extracted jointly or individually for each regressor separately. For illustration, consider a similar model as in (13) but without imposing constant returns to scale:

$$\ln(Y_{it}) = \beta_1 \ln(L_{it}) + \beta_2 \ln(K_{it}) + u_{it}$$

We specify the `iv()` option twice, one for each individual regressor. This yields

```
. xtivdfreg lY lL lK, absorb(list year) iv(lL, lags(2)) iv(lK, lags(2))
> factmax(3)

Defactored instrumental variables estimation

Factor estimation:
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
....                                                  4
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
....                                                  4
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
....                                                  4
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
....                                                  4
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
....                                                  4
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
....                                                  4
────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
...                                                   3

Group variable: list                      Number of obs         =      1084
Time variable: year                       Number of groups      =        48

Number of instruments  =      6           Obs per group     min =         9
Number of factors in X =      1                             avg =  22.58333
Number of factors in u =      1                             max =        31

Second-stage estimator (model with homogeneous slope coefficients)
──────────────────────────────────────────────────────────────────────────────
             |               Robust
          lY | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
─────────────┼────────────────────────────────────────────────────────────────
          lL |   .5286618   .0668596     7.91   0.000     .3976194    .6597042
          lK |   .4688145   .0802439     5.84   0.000     .3115394    .6260896
       _cons |   4.761711   1.287278     3.70   0.000     2.238693     7.28473
─────────────┼────────────────────────────────────────────────────────────────
     sigma_f |  .07657777   (std. dev. of factor error component)
     sigma_e |  .11538927   (std. dev. of idiosyncratic error component)
         rho |  .30576172   (fraction of variance due to factors)
──────────────────────────────────────────────────────────────────────────────
Hansen test of the overidentifying restrictions        chi2(4)     =    9.0610
H0: overidentifying restrictions are valid             Prob > chi2 =    0.0596
```

This time, the number of dotted lines corresponding to factor estimation from the covariates has doubled. This is because the factors are extracted separately for $\ln(L)$ and $\ln(K)$, and therefore the algorithm performs twice the number of iteration loops.

MGIV estimation of the baseline regression in (13) is computed by typing

```
. xtivdfreg ly lk, absorb(list year) iv(lk, lags(2)) factmax(3) mg
Defactored instrumental variables estimation
Factor estimation:
────┼──── 1 ────┼──── 2 ────┼──── 3 ────┼──── 4 ────┼──── 5
...                                                        3
────┼──── 1 ────┼──── 2 ────┼──── 3 ────┼──── 4 ────┼──── 5
...                                                        3
────┼──── 1 ────┼──── 2 ────┼──── 3 ────┼──── 4 ────┼──── 5
...                                                        3
────┼──── 1 ────┼──── 2 ────┼──── 3 ────┼──── 4 ────┼──── 5
...                                                        3
Group variable: list                        Number of obs         =      1084
Time variable: year                         Number of groups      =        48
Number of instruments  =      3             Obs per group     min =         9
Number of factors in X =      1                               avg =  22.58333
                                                              max =        31
Mean-group estimator (model with heterogeneous slope coefficients)
------------------------------------------------------------------------------
                   |               Robust
          ly       | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------------+----------------------------------------------------------
          lk       |   .4498443   .0778308     5.78   0.000     .2972987      .60239
        _cons      |   5.089032   .8010497     6.35   0.000     3.519003     6.65906
------------------------------------------------------------------------------
```

As we can see, the estimate of the coefficient of $\ln(K/L)$ is similar to that of the homogeneous model.[17] For further analysis using this example, see Eberhardt and Teal (2010).

5 Conclusion

`xtivdfreg` is useful for estimating large panel-data models with unobserved common factors or interactive effects. The slope coefficients can be either homogeneous or heterogeneous. The command accommodates a flexible specification of instruments and incorporates the two-way error-components model as a special case. Results obtained from the popular `ivregress` command can be reproduced using `xtivdfreg` by imposing zero factors.

6 Acknowledgments

We are grateful to an anonymous referee for providing useful comments and suggestions. Vasilis Sarafidis gratefully acknowledges financial support from the Australian Research Council under research grant number DP-170103135.

17. Also, this estimate is similar to the value obtained from the CCE estimator, which is computed by typing either `xtmg ly lk, cce` or `xtdcce2 ly lk, crosssectional(_all)`.

7 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 21-3
. net install st0650      (to install program files, if available)
. net get st0650          (to install ancillary files, if available)
```

To update the `xtivdfreg` package to the latest version, type

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

8 References

Ahn, S. C., and A. R. Horenstein. 2013. Eigenvalue ratio test for the number of factors. *Econometrica* 81: 1203–1227. https://doi.org/10.3982/ECTA8968.

Bai, J. 2009. Panel data models with interactive fixed effects. *Econometrica* 77: 1229–1279. https://doi.org/10.3982/ECTA6135.

Bai, J., Y. Liao, and J. Yang. 2015. Unbalanced panel data models with interactive effects. In *The Oxford Handbook of Panel Data*, ed. B. H. Baltagi, 149–170. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199940042.013.0005.

Bai, J., and S. Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70: 191–221. https://doi.org/10.1111/1468-0262.00273.

Baltagi, B. H., C. Ka, and F. Wang. 2021. Estimating and testing high dimensional factor models with multiple structural changes. *Journal of Econometrics* 220: 349–365. https://doi.org/10.1016/j.jeconom.2020.04.005.

Chudik, A., and M. H. Pesaran. 2015. Large panel data models with cross-sectional dependence: A survey. In *The Oxford Handbook Of Panel Data*, ed. B. H. Baltagi, 3–45. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199940042.013.0001.

Correia, S. 2016. Estimating multi-way fixed effect models with `reghdfe`. Presented July 28–29, 2016, at the Stata Conference 2016, Chicago. https://www.stata.com/meeting/chicago16/slides/chicago16_correia.pdf.

———. 2017. ftools: A faster Stata for large data sets. Presented July 27–28, 2017, at the Stata Conference 2017, Baltimore. https://www.stata.com/meeting/baltimore17/slides/Baltimore17_Correia.pdf.

Cui, G., M. Norkuté, V. Sarafidis, and T. Yamagata. 2020a. Two-stage instrumental variable estimation of linear panel data models with interactive effects. ISER Discussion Paper 1101, Institute of Social and Economic Research, Osaka University. https://ideas.repec.org/p/dpr/wpaper/1101.html.

Cui, G., V. Sarafidis, and T. Yamagata. 2020b. IV estimation of spatial dynamic panels with interactive effects: Large sample theory and an application on bank attitude toward risk. Monash Business School, Working Paper 11/20. https://doi.org/10.2139/ssrn.3642451.

Ditzen, J. 2018. Estimating dynamic common-correlated effects in Stata. *Stata Journal* 18: 585–617. https://doi.org/10.1177/1536867X1801800306.

Eberhardt, M. 2012. Estimating panel time-series models with heterogeneous slopes. *Stata Journal* 12: 61–71. https://doi.org/10.1177/1536867X1201200105.

Eberhardt, M., and F. Teal. 2010. Productivity analysis in global manufacturing production. Discussion Paper 515, Department of Economics, University of Oxford. https://ora.ox.ac.uk/objects/uuid:ea831625-9014-40ec-abc5-516ecfbd2118.

Harding, M., C. Lamarche, and M. H. Pesaran. 2020. Common correlated effects estimation of heterogeneous dynamic panel quantile regression models. *Journal of Applied Econometrics* 35: 294–314. https://doi.org/10.1002/jae.2753.

Jiang, B., Y. Yang, J. Gao, and C. Hsiao. Forthcoming. Recursive estimation in large panel data models: Theory and practice. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2020.07.055.

Juodis, A., H. Karabiyik, and J. Westerlund. 2021. On the robustness of the pooled CCE estimator. *Journal of Econometrics* 220: 325–348. https://doi.org/10.1016/j.jeconom.2020.06.002.

Juodis, A., and V. Sarafidis. 2018. Fixed T dynamic panel data estimators with multifactor errors. *Econometric Reviews* 37: 893–929. https://doi.org/10.1080/00927872.2016.1178875.

———. Forthcoming. A linear estimator for factor-augmented fixed-T panels with endogenous regressors. *Journal of Business & Economic Statistics* . https://doi.org/10.1080/07350015.2020.1766469.

Kapetanios, G., L. Serlenga, and Y. Shin. 2021. Estimation and inference for multidimensional heterogeneous panel datasets with hierarchical multi-factor error structure. *Journal of Econometrics* 220: 504–531. https://doi.org/10.1016/j.jeconom.2020.04.011.

Keeley, M. C. 1990. Deposit insurance, risk, and market power in banking. *American Economic Review* 80: 1183–1200.

Kripfganz, S., and C. Schwarz. 2019. Estimation of linear dynamic panel data models with time-invariant regressors. *Journal of Applied Econometrics* 34: 526–546. https://doi.org/10.1002/jae.2681.

Li, K., G. Cui, and L. Lu. 2020. Efficient estimation of heterogeneous coefficients in panel data models with common shocks. *Journal of Econometrics* 216: 327–353. https://doi.org/10.1016/j.jeconom.2019.08.011.

Martynova, N., L. Ratnovski, and R. Vlahu. 2020. Bank profitability, leverage constraints, and risk-taking. *Journal of Financial Intermediation* 44: 100821. https://doi.org/10.1016/j.jfi.2019.03.006.

Moon, H. R., and M. Weidner. 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* 83: 1543–1579. https://doi.org/10.3982/ECTA9382.

———. 2017. Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory* 33: 158–195. https://doi.org/10.1017/S0266466615000328.

Norkute, 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: 416–446. https://doi.org/10.1016/j.jeconom.2020.04.008.

Pesaran, M. H. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74: 967–1012. https://doi.org/10.1111/j.1468-0262.2006.00692.x.

Sarafidis, V., and T. Wansbeek. 2012. Cross-sectional dependence in panel data analysis. *Econometric Reviews* 31: 483–531. https://doi.org/10.1080/07474938.2011.611458.

———. 2021. Celebrating 40 years of panel data analysis: Past, present and future. *Journal of Econometrics* 220: 215–226. https://doi.org/10.1016/j.jeconom.2020.06.001.

Stock, J. H., and M. W. Watson. 1998. Diffusion indexes. NBER Working Paper No. 6702, The National Bureau of Economic Research. https://doi.org/10.3386/w6702.

Su, L., and S. Jin. 2012. Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics* 169: 34–47. https://doi.org/10.1016/j.jeconom.2012.01.006.

Westerlund, J., and J.-P. Urbain. 2015. Cross-sectional averages versus principal components. *Journal of Econometrics* 185: 372–377. https://doi.org/10.1016/j.jeconom.2014.09.014.

About the authors

Sebastian Kripfganz is an assistant professor at the Department of Economics in the University of Exeter Business School.

Vasilis Sarafidis is a professor of econometrics at the Department of Economics in the BI Norwegian Business School. At the time of writing this article, he was also a professor of econometrics at the Department of Econometrics and Business Statistics at Monash University.