Norkutė, M., Sarafidis, V., Yamagata, T., & Cui, G. (2021). Instrumental variable estimation of dynamic linear panel data models with defactored regressors and a multifactor error structure. *Annals Issue: Celebrating 40 Years of Panel Data Analysis: Past, Present and Future*, *220*(2), 416–446. https://doi.org/10.1016/j.jeconom.2020.04.008

Instrumental Variable Estimation of Dynamic Linear Panel Data Models with Defactored Regressors and a Multifactor Error Structure*

Milda Norkutė† Lund University

Vasilis Sarafidis‡ Monash University

Takashi Yamagata§ University of York and Osaka University

Guowei Cui¶ Monash University

Abstract

This paper develops two instrumental variable (IV) estimators for dynamic panel data models with exogenous covariates and a multifactor error structure when both the cross-sectional and time series dimensions, N and T respectively, are large. The main idea is to project out the common factors from the exogenous covariates of the model, and to construct instruments based on defactored covariates. For models with homogeneous slope coefficients, we propose a two-step IV estimator. In the first step, the model is estimated consistently by employing defactored covariates as instruments. In the second step, the entire model is defactored based on estimated factors extracted from the residuals of the first-step estimation, after which an IV regression is implemented using the same instruments as in step one. For models with heterogeneous slope coefficients, we propose a mean-group-type estimator, which involves the averaging of first-step IV estimates of cross-section-specific slopes. The proposed estimators do not need to seek for instrumental variables outside the model. Furthermore, these estimators are linear, and therefore computationally robust and inexpensive. Notably, they require no bias correction. We investigate the finite sample performances of the proposed estimators and associated statistical tests, and the results show that the estimators and the tests perform well even for small N and T.

Key Words: method of moments; dynamic panel data; cross-sectional dependence; factor model

JEL Classification: C13, C15, C23.

*This paper is dedicated to the memory of our wonderful colleague Guowei Cui, who passed away while this paper was being peer-reviewed. We are grateful to David Edgerton, Kazuhiko Hayakawa, Cheng Hsiao, Artūras Juodis, Hashem Pesaran, Mervyn Silvapulle, Joakim Westerlund, three anonymous referees and an Associate Editor, Professor Tom Wansbeek, for comments and suggestions that led to significant improvements.

†Department of Economics, Lund University, P.O. Box 7082, S-220 07 Lund, Sweden. E-mail address: Milda.Norkute@nek.lu.se. This author thanks the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number W2016–0467:1.

‡Department of Econometrics and Business Statistics, Monash University, VIC 3145, Australia. E-mail address: vasilis.sarafidis@monash.edu. This author gratefully acknowledges financial support from the Australian Research Council, under research grant number DP-170103135.

§Corresponding author. Department of Economics and Related Studies, University of York, York YO10 5DD, UK; E-mail address: takashi.yamagata@york.ac.uk. This author gratefully acknowledges the partial support of JSPS KAKENHI JP15H05728 and JP18K01545.

¶Department of Econometrics and Business Statistics, Monash University, VIC 3145, Australia.

1 Introduction

The rapid increase in the availability of panel data over the last few decades has inspired considerable interest in the development of effective ways of modelling and analysing these data. In particular, the issue of characterising cross-sectional dependence, and subsequently developing estimation methods that are consistent and yield asymptotically-valid inferences, has proven both popular and challenging. The factor structure approach has been used widely for modelling cross-sectional dependence. It escapes from the curse of dimensionality by asserting that there exists a common component that is a linear combination of a finite number of time-varying common factors with individual-specific factor loadings. Different interpretations of this approach can be provided, depending on the application considered. In macroeconomic panels, the unobserved factors are frequently viewed as economy-wide shocks that affect all individuals, albeit with different intensities; see e.g. Favero et al. (2005). In microeconomic panels, the factor error structure may reflect distinct sources of unobserved individual-specific heterogeneity, the impact of which varies over time. For instance, in a model of wage determination the factor loadings may represent several unmeasured skills that are specific to each individual, while the factors may capture the price of these skills, which changes intertemporally in an arbitrary way; see e.g. Carneiro et al. (2003) and Heckman et al. (2006).

A large body of literature has focused on the development of statistical inferential methods for models with an error factor structure. Two estimation approaches have been popular for large panels: Pesaran (2006) proposed the Common Correlated Effects (CCE) estimator, which approximates the unobserved factors using linear combinations of cross-sectional averages of the dependent and explanatory variables, while Bai (2009a) proposed an iterative least squares estimator with bias corrections, which approximates the unobserved factors using a principal component (PC) estimator.[1] For both estimators it is assumed that the regressors are strictly exogenous with respect to the idiosyncratic error component, whereas possible correlation between the regressors and the error factor component is permitted. Under somewhat weaker assumptions, Moon and Weidner (2015) show that the estimator of Bai (2009a) can be interpreted as a quasi maximum likelihood estimator (QMLE), the consistency of which is maintained even when the number of factors is not specified correctly, as long as it is larger than or equal to the true number of factors.

This paper considers the estimation of linear dynamic panel data models with an error factor structure in large panels.[2] Recently, the CCE and PC estimators have been extended to accommodate this case as well. In particular, Chudik and Pesaran (2015a) propose mean group CCE (CCEMG) estimation for panel autoregressive distributed lag models. The dynamic structure that they consider is very general for two reasons. Firstly, it permits cross-sectionally heterogeneous slope coefficients. Secondly, their model can be seen as a structural transformation of a multivariate dynamic process, such as a vector autoregressive model. Chudik and Pesaran (2015a) employ a mean-group-type estimator and propose that the regression be augmented with the cross-sectional averages of dependent variables and covariates and their lags, in order to control for the common components.

On the other hand, Moon and Weidner (2017) propose a bias-corrected QMLE (BC-QMLE) estimator for dynamic panel data models with homogeneous slopes, and put forward bias-corrected likelihood-based tests. Unlike CCEMG and the approach proposed in the present paper, they allow covariates to be correlated with the common component in disturbances without imposing a linear factor structure. Furthermore, the precision of the estimator is expected to be higher than those of existing estimators in large samples under certain regularity conditions.

[1]See Westerlund and Urbain (2015) for a comparison analysis of the CCE and PC estimation methods. Chudik and Pesaran (2015b), Sarafidis and Wansbeek (2012) and Bai and Wang (2016) also provide excellent surveys on the related literature.

[2]The estimation of such models for short panel data is considered by Ahn et al. (2013), Robertson and Sarafidis (2015), and Juodis and Sarafidis (2018, 2020).

This paper develops two instrumental variable (IV) estimators for dynamic panel data models with exogenous covariates and a multifactor error structure when both the cross-sectional and time series dimensions, N and T respectively, are large. We consider models with homogeneous and heterogeneous slope coefficients. In both cases, the main idea of the proposed approach is to project out the common factors from the exogenous covariates of the model, and to construct instruments based on defactored covariates.[3] The assumption underlying our IV approach is that any sources of endogeneity of the covariates arise due solely to the non-zero correlation between the common components in the covariates and in the model disturbances. Notably, this assumption can be tested using an overidentifying restrictions test.

In particular, we propose a two-step IV estimator for models with homogeneous slope coefficients. The first-step IV estimator is obtained simply by employing the aforementioned instruments based on the defactored covariates. In the second step, the entire morel based is defactored based on the factors extracted from the residuals of the first-step estimation. Subsequently, an IV regression is implemented using the same instruments as in step one. We derive the $\sqrt{NT}$-consistency of the two-step estimator and establish its asymptotic normality. Although the proposed IV approach and the BC-QMLE approach of Moon and Weidner (2017) are both based on the PC estimator, there are important differences between them in practice. Firstly, since our estimator is an instrumental variable estimator, it is not subject to the "Nickell bias" that arises with least squares type estimators in dynamic panel data models when T is relatively small. Secondly, our estimator is linear, and therefore robust and computationally inexpensive. In comparison, the BC-QMLE estimator requires nonlinear optimisation, which can be more costly and could fail to reach the global minimum.[4] Thirdly, unlike the QMLE estimator, which requires bias-correction to re-centre the limiting distribution of the original estimator, the proposed IV estimator does not have an asymptotic bias.

For models with heterogeneous slope coefficients, we propose a mean group-type estimator, which is the cross-sectional average of first-step IV estimates of individual-specific slopes. We establish the $\sqrt{N}$-consistency of our estimator to the population average of the slopes and its asymptotic normality. Our estimator has some advantages over the CCEMG estimator of Chudik and Pesaran (2015a). Firstly, since we employ the PC approach for defactoring the exogenous covariates, there is no need to seek external variables for approximating the factors when the number of unobserved factors is larger than the number of covariates plus one. In contrast, the CCE estimation requires additional sets of variables in this situation that are not in the original model of interest but are expected to form a part of the dynamic system. In practice, this might not be a trivial exercise. Secondly, CCE is subject to "Nickell bias". Chudik and Pesaran (2015a) propose that the bias be adjusted using the jackknife method, which might not be very effective for small or moderate values of T, especially with persistent data. Simulation results reported in this paper tend to confirm this observation.

Using simulated data, it is shown that the proposed approach performs satisfactorily under all circumstances examined. In particular, unlike the aforementioned alternative methods, the two IV estimators proposed here appear to have little or negligible bias in most circumstances, and a correct size of the t-test even for small sample sizes. Furthermore, the overidentifying restrictions test appears to have high power when the key assumption of the model is violated, namely the exogeneity of the covariates with respect to the purely idiosyncratic disturbance. In addition, the test tends to have good power under slope parameter heterogeneity, unless the number of degrees of freedom of the test statistic is very small. In contrast, the CCEMG estimator can suffer from a non-negligible bias and large size distortions of the associated t-test. Similarly, although BC-QMLE tends to exhibit the smallest dispersion in most cases under slope homogeneity, it suffers from a large bias and substantial size distortions of the associated bias-corrected test, unless both N and T are large.

[3]This idea can be regarded as an extension of the approach taken by Sarafidis et al. (2009).

[4]See Moon and Weidner (2019) for more details.

It is worth mentioning that our approach can be regarded as being opposite to those employed by Bai and Ng (2010) and Kapetanios and Marcellino (2010). Specifically, in their model the idiosyncratic errors of the reduced form regression of the covariates cause endogeneity, and therefore no error factor structure is considered in the structural model of interest. They propose that instruments be constructed by extracting the common components from external variables and the endogenous covariates in the model. Our approach essentially complements theirs.[5]

The remainder of the paper is organised as follows. Section 2 focuses on the model with homogeneous slopes, and develops a consistent and asymptotically normal two-step IV estimator. Section 3 focuses on heterogeneous panels and puts forward consistent estimators of cross-sectionally heterogeneous slope coefficients and their averages. It also establishes the asymptotic normality of the mean group estimator. Section 4 studies the finite sample performance of the proposed estimators along with the CCE estimator of Chudik and Pesaran (2015a) and the BC-QMLE estimator of Moon and Weidner (2017). Section 5 contains concluding remarks. Proofs of propositions, theorems and corollaries, together with the lemmas used, are contained in Appendix A. Appendix B gives proofs of all of the lemmas, and Appendix C provides extra simulation results. Both these appendices are available as Supplemental Material to this paper.

2 Model and Estimation Method

Consider the following autoregressive distributed lag, ARDL$(1,0)$, panel data model with homogeneous slopes and a multifactor error structure:[6]

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

with

$$u_{it} = \boldsymbol{\gamma}_{yi}^{0\prime}\mathbf{f}_{y,t}^{0} + \varepsilon_{it}, \tag{2}$$

where $|\rho| < 1$; $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_k)'$ such that at least one of $\{\beta_\ell\}_{\ell=1}^{k}$ is non-zero; $\mathbf{x}_{it} = (x_{1it}, x_{2it}, ..., x_{kit})'$ is a $k \times 1$ vector of regressors, and $\mathbf{f}_{y,t}^0 = (f_{y,1t}^0, f_{y,2t}^0, ..., f_{y,m_y t}^0)'$ denotes an $m_y \times 1$ vector of true unobservable factors. The $m_y \times 1$ vector $\boldsymbol{\gamma}_{yi}^0$ contains the true factor loadings associated with $\mathbf{f}_{y,t}^0$, and ε_{it} is an idiosyncratic error. $\mathbf{x}_{it}$ is subject to the following process:

$$\mathbf{x}_{it} = \boldsymbol{\Gamma}_{xi}^{0\prime}\mathbf{f}_{x,t}^{0} + \mathbf{v}_{it}, \tag{3}$$

where $\boldsymbol{\Gamma}_{xi}^0 = (\boldsymbol{\gamma}_{1i}^0, \boldsymbol{\gamma}_{2i}^0, ..., \boldsymbol{\gamma}_{ki}^0)$ denotes the true $m_x \times k$ factor loading matrix, $\mathbf{f}_{x,t}^0 = (f_{x,1t}^0, f_{x,2t}^0, ..., f_{x,m_x t}^0)'$ denotes an $m_x \times 1$ vector of true factors, and $\mathbf{v}_{it} = (v_{1it}, v_{2it}, ..., v_{kit})'$ is an idiosyncratic error term that is independent of ε_{it}.

Remark 1 Our approach permits correlations between and within $\boldsymbol{\gamma}_{yi}^0$ and $\boldsymbol{\Gamma}_{xi}^0$. Moreover, (non)overlapping elements in $\mathbf{f}_{y,t}^0$ and $\mathbf{f}_{x,t}^0$ may be correlated to each other. Importantly, our approach controls for endogeneity of $\mathbf{x}_{it}$ that stems from the common components, but assumes that $\mathbf{x}_{it}$ is strongly exogenous with respect to ε_{it}.

Remark 2 The results presented in this paper remain valid when individual-specific and common time effects are present, provided that $\{y_{it}, \mathbf{x}_{it}'\}$ is replaced with the transformed variables $\{\dot{y}_{it}, \dot{\mathbf{x}}_{it}'\}$, where $\dot{y}_{it} = y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}$ and $\dot{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i - \bar{\mathbf{x}}_t + \bar{\mathbf{x}}$ with $\bar{y}_i = T^{-1}\sum_{t=0}^{T} y_{it}$, $\bar{y}_t = N^{-1}\sum_{i=1}^{N} y_{it}$ and $\bar{y} = N^{-1}\sum_{i=1}^{N} \bar{y}_i$, and $\bar{\mathbf{x}}_i$, $\bar{\mathbf{x}}_t$ and $\bar{\mathbf{x}}$ are defined analogously. Indeed, the experiments for our proposed estimators and the tests implemented in Section 4 are based on the transformed variables.

[5]Another important related work is that by Harding and Lamarche (2011), which proposes an instrumental variable estimator for a model with an error factor structure.

[6]The main results of this paper extend naturally to models with higher-order lags, i.e. ARDL(p, q) for $p > 0$ and $q \geq 0$. Models with heterogeneous slopes are considered in Section 3.

Stacking the T observations for each i yields

$$\mathbf{y}_i = \rho \mathbf{y}_{i,-1} + \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i \text{ with } \mathbf{u}_i = \mathbf{F}_y^0 \boldsymbol{\gamma}_{yi}^0 + \boldsymbol{\varepsilon}_i, \tag{4}$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{iT})'$ and $\mathbf{y}_{i,-1} = L^1 \mathbf{y}_i = (y_{i0}, y_{i1}, ..., y_{iT-1})'$, with L^j being the j^{th} lag operator, $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})'$, $\mathbf{u}_i = (u_{i1}, u_{i2}, ..., u_{iT})'$, $\mathbf{F}_y^0 = (\mathbf{f}_{y,1}^0, \mathbf{f}_{y,2}^0, ..., \mathbf{f}_{y,T}^0)'$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})'$. Similarly,

$$\mathbf{X}_i = \mathbf{F}_x^0 \boldsymbol{\Gamma}_{xi}^0 + \mathbf{V}_i, \tag{5}$$

where $\mathbf{F}_x^0 = (\mathbf{f}_{x,1}^0, \mathbf{f}_{x,2}^0, ..., \mathbf{f}_{x,T}^0)'$ and $\mathbf{V}_i = (\mathbf{v}_{i1}, \mathbf{v}_{i2}, ..., \mathbf{v}_{iT})'$.

Let $\mathbf{W}_i = (\mathbf{y}_{i,-1}, \mathbf{X}_i)$ and $\boldsymbol{\theta} = (\rho, \boldsymbol{\beta}')'$. The model in Eq. (4) can be written more concisely as

$$\mathbf{y}_i = \mathbf{W}_i \boldsymbol{\theta} + \mathbf{u}_i. \tag{6}$$

Our estimation approach involves two steps. In the first step, we asymptotically eliminate the common factors in $\mathbf{X}_i$ by projecting them out. Subsequently, we use the defactored regressors as instruments for estimating the structural parameters of the model. To illustrate the first-step estimator, consider the following projection matrices:

$$\mathbf{M}_{F_x^0} = \mathbf{I}_T - \mathbf{F}_x^0 \left(\mathbf{F}_x^{0\prime} \mathbf{F}_x^0\right)^{-1} \mathbf{F}_x^{0\prime}; \ \mathbf{M}_{F_{x,-1}^0} = \mathbf{I}_T - \mathbf{F}_{x,-1}^0 \left(\mathbf{F}_{x,-1}^{0\prime} \mathbf{F}_{x,-1}^0\right)^{-1} \mathbf{F}_{x,-1}^{0\prime}, \tag{7}$$

where $\mathbf{F}_{x,-1}^0 = L^1 \mathbf{F}_x^0$. Suppose for the moment that $\mathbf{F}_x^0$ is observed. Premultiplying $\mathbf{X}_i$ by $\mathbf{M}_{F_x^0}$ would yield $\mathbf{M}_{F_x^0} \mathbf{X}_i = \mathbf{M}_{F_x^0} \mathbf{V}_i$. Assuming that $\mathbf{V}_i$ is independent of $\boldsymbol{\varepsilon}_i, \mathbf{F}_x^0, \mathbf{F}_y^0$ and $\boldsymbol{\gamma}_{yi}^0$, it is easy to see that $E(\mathbf{X}_i' \mathbf{M}_{F_x^0} \mathbf{u}_i) = E(\mathbf{V}_i' \mathbf{M}_{F_x^0} \mathbf{u}_i) = \mathbf{0}$. Furthermore, let $\mathbf{X}_{i,-j} = L^j \mathbf{X}_i$. So long as $\{y_{it}, \mathbf{x}_{it}'\}$, $t = 0, 1, ..., T$ is observed, and the $T \times k$ matrix $\mathbf{X}_{i,-1}$ is also observed. Using similar assumptions, one can show that $E(\mathbf{X}_{i,-1}' \mathbf{M}_{F_{x,-1}^0} \mathbf{u}_i) = E(\mathbf{V}_{i,-1}' \mathbf{M}_{F_{x,-1}^0} \mathbf{u}_i) = \mathbf{0}$. Collect the set of instruments:

$$\mathbf{Z}_i = \left(\mathbf{M}_{F_x^0} \mathbf{X}_i, \mathbf{M}_{F_{x,-1}^0} \mathbf{X}_{i,-1}\right) \ (T \times 2k). \tag{8}$$

Given the model in Eq. (6), it is clear that $\mathbf{Z}_i$ satisfies $E(\mathbf{Z}_i' \mathbf{u}_i) = \mathbf{0}$, and also $E(\mathbf{Z}_i' \mathbf{W}_i) \neq \mathbf{0}$. Thus, $\mathbf{Z}_i$ is a valid instrument set.[7]

Having obtained a consistent first-step estimator, the second step of our approach involves estimating the factors in the error term, $\mathbf{F}_y^0$, using the residuals in the first-step IV regression. Then, we asymptotically eliminate $\mathbf{F}_y^0$ from the entire model by projecting them out from $\{\mathbf{y}_i, \mathbf{W}_i\}$, and use the instruments $\mathbf{Z}_i$ to obtain the second-step estimator. To portray the second-step estimator, suppose for the moment that $\mathbf{F}_y^0$ is observable and define the projection matrix

$$\mathbf{M}_{F_y^0} = \mathbf{I}_T - \mathbf{F}_y^0 \left(\mathbf{F}_y^{0\prime} \mathbf{F}_y^0\right)^{-1} \mathbf{F}_y^{0\prime}. \tag{9}$$

Premultiplying the model in Eq. (6) by $\mathbf{M}_{F_y^0}$, we obtain

$$\mathbf{M}_{F_y^0} \mathbf{y}_i = \mathbf{M}_{F_y^0} \mathbf{W}_i \boldsymbol{\theta} + \mathbf{M}_{F_y^0} \boldsymbol{\varepsilon}_i, \tag{10}$$

where the factor component $\mathbf{F}_y^0 \boldsymbol{\gamma}_{yi}^0$ in the error term is swept away. Based on similar reasoning as in the earlier discussion, we can see easily that $E(\mathbf{Z}_i' \mathbf{M}_{F_y^0} \boldsymbol{\varepsilon}_i) = \mathbf{0}$ and $E(\mathbf{Z}_i' \mathbf{M}_{F_y^0} \mathbf{W}_i) \neq \mathbf{0}$. Thus, it is straightforward to apply instrumental variable (IV) estimation using $\mathbf{Z}_i$ to the transformed model in Eq. (10).[8]

[7] In general, for ARDL(p, q) models, the usual order condition for IV identification requires using $(s+1)k$ instruments of the form $\left\{\mathbf{M}_{F_{x,-r}^0} \mathbf{X}_{i,-r}\right\}_{r=0}^{s}$, where $s = q + \lceil p/k \rceil$ and $\lceil . \rceil$ is the ceiling function.

[8] This IV estimation is equivalent to that using the transformed instrument set, $\mathbf{M}_{F_y^0} \mathbf{Z}_i$, for the original model in Eq. (6).

In practice, the factors $\mathbf{F}_x^0$, $\mathbf{F}_{x,-1}^0$ and $\mathbf{F}_y^0$ usually are not observed. As will be discussed in detail below, we replace these factors with the ones estimated based on the principal components approach, as advanced by Bai (2003) and Bai (2009a), among many others.[9]

This section and the next treat the number of factors, m_x and m_y, as given. In practice, though, these should be estimated. m_x can be estimated from the raw data $\mathbf{x}_{it}$, $t = 0, ..., T$, $i = 1, ..., N$, using methods that have been proposed in the literature, such as the information criterion approach of Bai and Ng (2002) or the eigenvalue methods of Ahn and Horenstein (2013). m_y can be estimated from the residual covariance matrix using the methods mentioned above.[10] The Monte Carlo section below uses the various existing methods to determine the number of factors, and show that these provide quite an accurate determination of the number of factors in our experimental design.

Remark 3 Since our approach makes use of the transformed x's as instruments, the identification of ρ requires that *at least one element of $\boldsymbol{\beta}$ is not equal to zero*, given the model in Eq. (3). We believe that this is a mild restriction, especially compared to the restriction that *all of the elements in $\boldsymbol{\beta}$ be non-zero*. Specifically, the identification of the autoregressive parameter can be achieved based on the covariate(s) and lagged value(s) that correspond to the non-zero slope coefficient(s). Notably, it is not necessary to know which covariates have non-zero coefficients, since by construction the IV estimation procedure does not require that all instruments be relevant to all endogenous regressors.

Remark 4 More instruments may be available when further lags of $\mathbf{x}_{it}$ are observed. In particular, given the model in Eq. (3), $(j + 1)k$ instruments can be used instead of Eq. (8) when $\{\mathbf{x}_{it}\}_{t=1-j}^T$ for $j \geq 1$ are observable:

$$\mathbf{Z}_i = \left(\mathbf{M}_{F_x^0}\mathbf{X}_i, \mathbf{M}_{F_{x,-1}^0}\mathbf{X}_{i,-1}, ..., \mathbf{M}_{F_{x,-j}^0}\mathbf{X}_{i,-j}\right) \ (T \times (j+1)k). \tag{11}$$

It is well documented in the literature that including larger numbers of instruments makes the estimator more efficient but also more biased. This paper assumes a small finite number $j \geq 1$ that does not depend on T.[11] Without loss of generality, we set $j = 1$ for the theoretical analysis in Sections 2 and 3. Section 4 conducts a finite sample experiment with different values of j.[12]

To obtain our results it is sufficient to make the following assumptions, where $\text{tr}[\mathbf{A}]$ and $||\mathbf{A}|| = \sqrt{\text{tr}[\mathbf{A}'\mathbf{A}]}$ denote the trace and Frobenius (Euclidean) norm of the matrix $\mathbf{A}$, respectively, and Δ is a finite positive constant.

Assumption 1 (idiosyncratic error in y): ε_{it} is distributed independently across i and t, with $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}^2) = \sigma_{\varepsilon,it}^2$, and $E|\varepsilon_{it}|^{8+\delta} \leq \Delta < \infty$ for a small positive constant δ.

Assumption 2 (idiosyncratic error in x): (i) $v_{\ell it}$ is distributed independently across i and group-wise independent from ε_{it}; (ii) $E(v_{\ell it}) = 0$ and $E|v_{\ell it}|^{8+\delta} \leq \Delta < \infty$; (iii) $T^{-1}\sum_{s=1}^T\sum_{t=1}^T E|v_{\ell is}v_{\ell it}|^{1+\delta} \leq \Delta < \infty$; (iv) $E\left|N^{-1/2}\sum_{i=1}^N[v_{\ell is}v_{\ell it} - E(v_{\ell is}v_{\ell it})]\right|^4 \leq \Delta < \infty$ for every ℓ, t and s; (v) $N^{-1}T^{-2}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T\sum_{r=1}^T\sum_{w=1}^T|\text{cov}(v_{\ell is}v_{\ell it}, v_{\ell ir}v_{\ell iw})| \leq \Delta < \infty$; and (vi) the largest eigenvalue of $E(\mathbf{v}_{\ell i}\mathbf{v}_{\ell i}')$ is bounded uniformly for every ℓ, i and T.

[9] One could also employ Pesaran's (2006) approach for estimating the common factors in the regressors.

[10] See Bai (2009b, C.3) for a discussion on the estimation of the number of factors in disturbances.

[11] The limit behaviour of the estimators when the number of instruments increases with T might be of theoretical interest, but is beyond the scope of this paper. See Alvarez and Arellano (2003), among others, for a related analysis.

[12] The simulation results confirm that different values of j are subject to the well-known trade-off between bias and efficiency. In principle, one could devise a lag selection procedure for optimising the bias-variance trade-off for the GMM estimator, as per Okui (2009); however, we leave this avenue for future research.

Assumption 3 (stationary factors): $\mathbf{f}^0_{x,t} = \boldsymbol{C}_x(L)\mathbf{e}_{f_x,t}$ and $\mathbf{f}^0_{y,t} = \boldsymbol{C}_y(L)\mathbf{e}_{f_y,t}$, where $\boldsymbol{C}_x(L)$ and $\boldsymbol{C}_y(L)$ are absolutely summable, $\mathbf{e}_{f_x,t} \sim i.i.d.(\mathbf{0}, \boldsymbol{\Sigma}_{fx})$ and $\mathbf{e}_{f_y,t} \sim i.i.d.(\mathbf{0}, \boldsymbol{\Sigma}_{fy})$, with $\boldsymbol{\Sigma}_{fx}$ and $\boldsymbol{\Sigma}_{fy}$ being positive definite matrices. Each element of $\mathbf{e}_{f_x,t}$ and $\mathbf{e}_{f_y,t}$ has finite fourth-order moments and all are group-wise independent from $\mathbf{v}_{it}$ and ε_{it}.

Assumption 4 (random factor loadings): $\boldsymbol{\Gamma}^0_{xi} \sim i.i.d.(\mathbf{0}, \boldsymbol{\Sigma}_{\Gamma x})$, $\boldsymbol{\gamma}^0_{yi} \sim i.i.d.(\mathbf{0}, \boldsymbol{\Sigma}_{\gamma y})$, where $\boldsymbol{\Sigma}_{\Gamma x}$ and $\boldsymbol{\Sigma}_{\gamma y}$ are positive definite matrices, and each element of $\boldsymbol{\Gamma}^0_{xi}$ and $\boldsymbol{\gamma}^0_{yi}$ has finite fourth-order moments. $\boldsymbol{\Gamma}^0_{xi}$ and $\boldsymbol{\gamma}^0_{yi}$ are independent groups from ε_{it}, $\mathbf{v}_{it}$, $\mathbf{e}_{f_x,t}$ and $\mathbf{e}_{f_y,t}$.

Assumption 5 (identification of $\boldsymbol{\theta}$): (i) $\tilde{\mathbf{A}}_{i,T} = T^{-1}\mathbf{Z}_i'\mathbf{W}_i$, $\tilde{\mathbf{B}}_{i,T} = T^{-1}\mathbf{Z}_i'\mathbf{Z}_i$, $\mathbf{A}_{i,T} = T^{-1}\mathbf{Z}_i'\mathbf{M}_{F^0_y}\mathbf{W}_i$ and $\mathbf{B}_{i,T} = T^{-1}\mathbf{Z}_i'\mathbf{M}_{F^0_y}\mathbf{Z}_i$ have full column rank for all i for a sufficiently large value of T; (ii) $E\left\|\tilde{\mathbf{A}}_{i,T}\right\|^{2+2\delta} \leq \Delta < \infty$, $E\left\|\tilde{\mathbf{B}}_{i,T}\right\|^{2+2\delta} \leq \Delta < \infty$, $E\left\|\mathbf{A}_{i,T}\right\|^{2+2\delta} \leq \Delta < \infty$ and $E\left\|\mathbf{B}_{i,T}\right\|^{2+2\delta} \leq \Delta < \infty$ for all i for a sufficiently large value of T; and (iii) $E\left\|\boldsymbol{\varphi}_{FiT}\right\|^{2+\delta} \leq \Delta < \infty$ for all i for a sufficiently large value of T, where $\boldsymbol{\varphi}_{FiT} = T^{-1/2}\mathbf{Z}_i{}'\mathbf{M}_{F^0_y}\boldsymbol{\varepsilon}_i$, and $E(\boldsymbol{\varphi}_{FiT}\boldsymbol{\varphi}'_{FiT})$ is a positive definite matrix for all i for a sufficiently large value of T. In addition, $\lim_{N,T\to\infty} N^{-1}\sum_{i=1}^N E\left(\boldsymbol{\varphi}_{FiT}\boldsymbol{\varphi}'_{FiT}\right) = \boldsymbol{\Omega}$, which is a fixed positive definite matrix.

These assumptions merit some discussion. First of all, note that Assumption 1 allows non-normality and (unconditional) time series and cross-sectional heteroskedasticity in the idiosyncratic errors in the equation for y. Assumptions 2 and 3 allow for serial correlation in the idiosyncratic errors in the equation for x and the factors. Assumption 2 is in line with Bai (2003), but assumes independence across i, which can be relaxed such that the factors and $(\varepsilon_{it}, \mathbf{v}_{it})$ and/or ε_{jt} and ε_{is} are weakly dependent, provided that higher-order moments exist; see Assumptions D–F of Bai (2003).[13] Assumptions 3 and 4 are standard in the principal components literature; see e.g. Bai (2003), among others. Assumption 3 permits correlations between $\mathbf{f}^0_{x,t}$ and $\mathbf{f}^0_{y,t}$, and within each of them. Assumption 4 allows for possible non-zero correlations between $\boldsymbol{\gamma}^0_{yi}$ and $\boldsymbol{\Gamma}^0_{xi}$, and within each of them. Since the variables y_{it} and $\mathbf{x}_{it}$ of the same individual unit i can be affected by the common shocks in a related manner, it is potentially important in practice to allow for this possibility. Finally, Assumption 5(i)–(ii) is common in overidentified instrumental variable (IV) estimation; for example, see Wooldridge (2002, Ch. 5). Assumption 5(iii) is required for the identification of the estimator, the consistency property of the variance-covariance estimator and the asymptotic normality of the estimator, as N and T tend to infinity jointly.

Let us begin with a discussion of our approach's first-step IV estimator. Given m_x, the factors are extracted from $\{\mathbf{X}_i\}_{i=1}^N$ using principal components (PC). Define $\hat{\mathbf{F}}_x$ as $\sqrt{T}$ times the eigenvectors that correspond to the m_x largest eigenvalues of the $T \times T$ matrix $\sum_{i=1}^N \mathbf{X}_i\mathbf{X}_i'/NT$. $\hat{\mathbf{F}}_{x,-1}$ is defined in the same way, but this time based on $\sum_{i=1}^N \mathbf{X}_{i,-1}\mathbf{X}'_{i,-1}/NT$.

Remark 5 Note that $\mathbf{F}^0_x$, $\boldsymbol{\Gamma}^0_{xi}$, $\mathbf{F}^0_{x,-1}$, ($\mathbf{F}^0_y$ and $\boldsymbol{\gamma}^0_{yi}$) can be identified and estimated up to an invertible $m_x \times m_x$ (and $m_y \times m_y$) matrix transformation; see Bai and Ng (2013), among others. For example, $\hat{\mathbf{F}}_x$ is a consistent estimator of $\mathbf{F}_x = \mathbf{F}^0_x\mathbf{G}_x$, where $\mathbf{G}_x$ is an invertible matrix such that $\mathbf{F}_x'\mathbf{F}_x/T = \mathbf{I}_T$, $\boldsymbol{\gamma}_{\ell i} = \mathbf{G}_x^{-1}\boldsymbol{\gamma}^0_{\ell i}$, with $\sum_{\ell=1}^k\sum_{i=1}^N \boldsymbol{\gamma}_{\ell i}\boldsymbol{\gamma}'_{\ell i}$ being a diagonal matrix. We define $\mathbf{F}_{x,-1}$, ($\mathbf{F}_y$ and $\boldsymbol{\gamma}_{yi}$) in an analogous manner.

The empirical counterparts of the projection matrices that are defined in Eqs. (7) and (9) are given by

$$\mathbf{M}_{\hat{F}_x} = \mathbf{I}_T - \hat{\mathbf{F}}_x\left(\hat{\mathbf{F}}_x'\hat{\mathbf{F}}_x\right)^{-1}\hat{\mathbf{F}}_x';\ \mathbf{M}_{\hat{F}_{x,-1}} = \mathbf{I}_T - \hat{\mathbf{F}}_{x,-1}\left(\hat{\mathbf{F}}'_{x,-1}\hat{\mathbf{F}}_{x,-1}\right)^{-1}\hat{\mathbf{F}}'_{x,-1}. \tag{12}$$

The associated transformed instrument matrix discussed above is

$$\hat{\mathbf{Z}}_i = \left(\mathbf{M}_{\hat{F}_x}\mathbf{X}_i, \mathbf{M}_{\hat{F}_{x,-1}}\mathbf{X}_{i,-1}\right). \tag{13}$$

[13] This includes conditional heteroskedasticity, such as ARCH or GARCH processes.

The first-step instrumental variable (IV) estimator is given by $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}_{IV} = \left(\hat{\hat{\mathbf{A}}}'_{NT}\hat{\hat{\mathbf{B}}}^{-1}_{NT}\hat{\hat{\mathbf{A}}}_{NT}\right)^{-1}\hat{\hat{\mathbf{A}}}'_{NT}\hat{\hat{\mathbf{B}}}^{-1}_{NT}\hat{\hat{\mathbf{g}}}_{NT}, \tag{14}$$

where

$$\hat{\hat{\mathbf{A}}}_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathbf{Z}}'_i\mathbf{W}_i,\ \hat{\hat{\mathbf{B}}}_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathbf{Z}}'_i\hat{\mathbf{Z}}_i,\ \hat{\hat{\mathbf{g}}}_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\hat{\mathbf{Z}}'_i\mathbf{y}_i. \tag{15}$$

Firstly, we will derive consistency for the above estimator. To begin with, from Eqs. (6) and (14) we obtain

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{IV} - \boldsymbol{\theta}\right) = \left(\hat{\hat{\mathbf{A}}}'_{NT}\hat{\hat{\mathbf{B}}}^{-1}_{NT}\hat{\hat{\mathbf{A}}}_{NT}\right)^{-1}\hat{\hat{\mathbf{A}}}'_{NT}\hat{\hat{\mathbf{B}}}^{-1}_{NT}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\hat{\mathbf{Z}}'_i\mathbf{u}_i\right). \tag{16}$$

Since the asymptotic properties of the estimator are determined primarily by those of $\sum_{i=1}^{N}\hat{\mathbf{Z}}'_i\mathbf{u}_i/\sqrt{NT}$, we focus on this term. The formal analysis is provided as a proposition below, where $(N,T)\xrightarrow{j}\infty$ signifies that N and T tend to infinity jointly.

Proposition 1 *Under Assumptions 1–5, as $(N,T)\xrightarrow{j}\infty$ such that $N/T \to c$ with $0 < c < \infty$,*

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\hat{\mathbf{Z}}'_i\mathbf{u}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\tilde{\mathbf{Z}}'_i\mathbf{u}_i + \sqrt{\frac{T}{N}}\mathbf{b}_{1NT} + \sqrt{\frac{N}{T}}\mathbf{b}_{2NT} + o_p(1),$$

where $\hat{\mathbf{Z}}_i$ is defined by Eq. (13), $\tilde{\mathbf{Z}}_i = (\mathbf{M}_{F_x^0}\tilde{\mathbf{X}}_i, \mathbf{M}_{F_{x,-1}^0}\tilde{\mathbf{X}}_{i,-1})$, $\tilde{\mathbf{X}}_i = \mathbf{X}_i - \frac{1}{N}\sum_{n=1}^{N}\mathbf{X}_n\boldsymbol{\Gamma}^{0\prime}_{xn}\left(\boldsymbol{\Upsilon}^0_{xkN}\right)^{-1}\boldsymbol{\Gamma}^0_{xi}$, $\tilde{\mathbf{X}}_{i,-1} = \mathbf{X}_{i,-1} - \frac{1}{N}\sum_{n=1}^{N}\mathbf{X}_{n,-1}\boldsymbol{\Gamma}^{0\prime}_{xn}\left(\boldsymbol{\Upsilon}^0_{xkN}\right)^{-1}\boldsymbol{\Gamma}^0_{xi}$, $\boldsymbol{\Upsilon}^0_{xkN} = \frac{1}{N}\sum_{\ell=1}^{k}\sum_{i=1}^{N}\boldsymbol{\gamma}^0_{\ell i}\boldsymbol{\gamma}^{0\prime}_{\ell i}$ and $\mathbf{b}_{1NT} = [\mathbf{b}'_{11NT}, \mathbf{b}'_{12NT}]'$, $\mathbf{b}_{2NT} = [\mathbf{b}'_{21NT}, \mathbf{b}'_{22NT}]'$, with

$$\mathbf{b}_{11NT} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\tilde{\mathbf{V}}'_i\mathbf{V}_j}{T}\boldsymbol{\Gamma}^{0\prime}_{xj}\left(\boldsymbol{\Upsilon}^0_{xkN}\right)^{-1}\left(\frac{\mathbf{F}'_x\mathbf{F}^0_x}{T}\right)^{-1}\frac{\mathbf{F}^{0\prime}_x\mathbf{u}_i}{T};$$

$$\mathbf{b}_{12NT} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\tilde{\mathbf{V}}'_{i,-1}\mathbf{V}_{j,-1}}{T}\boldsymbol{\Gamma}^{0\prime}_{xj}\left(\boldsymbol{\Upsilon}^0_{xkN}\right)^{-1}\left(\frac{\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}^0_{x,-1}}{T}\right)^{-1}\frac{\mathbf{F}^{0\prime}_{x,-1}\mathbf{u}_i}{T};$$

$$\mathbf{b}_{21NT} = -\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Gamma}^{0\prime}_{xi}\left(\boldsymbol{\Upsilon}^0_{xkN}\right)^{-1}\left(\frac{\mathbf{F}^{0\prime}_x\mathbf{F}^0_x}{T}\right)^{-1}\mathbf{F}^{0\prime}_x\bar{\boldsymbol{\Sigma}}_{kNT}\mathbf{M}_{F_x^0}\mathbf{u}_i;$$

$$\mathbf{b}_{22NT} = -\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Gamma}^{0\prime}_{xi}\left(\boldsymbol{\Upsilon}^0_{xkN}\right)^{-1}\left(\frac{\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}^0_{x,-1}}{T}\right)^{-1}\mathbf{F}^{0\prime}_{x,-1}\bar{\boldsymbol{\Sigma}}_{kNT,-1}\mathbf{M}_{F_{x,-1}^0}\mathbf{u}_i,$$

$\tilde{\mathbf{V}}_i = \mathbf{V}_i - \frac{1}{N}\sum_{n=1}^{N}\mathbf{V}_n\boldsymbol{\Gamma}^{0\prime}_{xn}\left(\boldsymbol{\Upsilon}^0_{xkN}\right)^{-1}\boldsymbol{\Gamma}^0_{xi}$, $\tilde{\mathbf{V}}_{i,-1} = \mathbf{V}_{i,-1} - \frac{1}{N}\sum_{n=1}^{N}\mathbf{V}_{n,-1}\boldsymbol{\Gamma}^{0\prime}_{xn}\left(\boldsymbol{\Upsilon}^0_{xkN}\right)^{-1}\boldsymbol{\Gamma}^0_{xi}$, $\bar{\boldsymbol{\Sigma}}_{kNT} = \frac{1}{N}\sum_{\ell=1}^{k}\sum_{j=1}^{N}E\left(\mathbf{v}_{\ell j}\mathbf{v}'_{\ell j}\right)$ *and* $\bar{\boldsymbol{\Sigma}}_{kNT,-1} = \frac{1}{N}\sum_{\ell=1}^{k}\sum_{j=1}^{N}E\left(\mathbf{v}_{\ell j,-1}\mathbf{v}'_{\ell j,-1}\right)$.

Remark 6 The source of the bias term in Proposition 1 differs from those of the bias terms reported by Bai (2009a) and Moon and Weidner (2017). In particular, the bias term of our estimator arises primarily due to the correlation between the factor loadings associated with $\mathbf{F}_x$ in x and the error term in the equation of y, $\mathbf{u}_i$. On the other hand, the two bias terms in Bai (2009a) and Moon and Weidner (2017) arise from error serial dependence and weak cross-sectional dependence. In our case, error serial correlation in the idiosyncratic part of the x process, $v_{\ell it}$, does not result in bias because $v_{\ell it}$ is not correlated with the error term in the y equation, ε_{it}. Also note that Moon and Weidner (2017) report that an additional bias term that generalises the small T bias, called the "Nickell bias", typically occurs in the least squares estimation of dynamic panel models. Our estimator is not subject to such a bias, as it is based on instrumental variables.

It can be shown from the result stated in Proposition 1 that $\sum_{i=1}^{N} \tilde{\mathbf{Z}}_i' \mathbf{u}_i / \sqrt{NT}$ is $O_p(1)$ and tends to a multivariate distribution. In addition, $\sqrt{T/N}\mathbf{b}_{1NT}$ and $\sqrt{N/T}\mathbf{b}_{2NT}$ are $O_p(1)$ as $(N,T) \xrightarrow{j} \infty$ such that $N/T \to c$ with $0 < c < \infty$. Therefore, the IV estimator is $\sqrt{NT}$-consistent in such situations.

The above discussion is summarised formally in the following theorem:

Theorem 1 *Consider the model in Eqs. (1)–(3) and suppose that Assumptions 1–5 hold true. Then,*

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{IV} - \boldsymbol{\theta}\right) = O_p\left(1\right)$$

as $(N,T) \xrightarrow{j} \infty$ *such that* $N/T \to c$ *with* $0 < c < \infty$*, where* $\hat{\boldsymbol{\theta}}_{IV}$ *is defined in Eq. (14).*

Even though the estimator $\hat{\boldsymbol{\theta}}_{IV}$ is $\sqrt{NT}$-consistent, under our assumptions the limiting distribution of $\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{IV} - \boldsymbol{\theta}\right)$ will contain asymptotic bias terms such as the limits of $\mathbf{b}_{1NT}$ and $\mathbf{b}_{2NT}$, which are defined in Proposition 1.[14] Rather than bias-correcting this estimator, we put forward a potentially more efficient second-step estimator, by projecting $\mathbf{F}_y$ out from the model asymptotically using $\hat{\boldsymbol{\theta}}_{IV}$.

We compute the second-step estimator by estimating the factors $\mathbf{F}_y$ using principal components from $\{\hat{\mathbf{u}}_i\}_{i=1}^{N}$, where $\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{W}_i \hat{\boldsymbol{\theta}}_{IV}$, with $\hat{\boldsymbol{\theta}}_{IV}$ being a first-step IV estimator defined in Eq. (14). We define $\hat{\mathbf{F}}_y$ as $\sqrt{T}$ times the eigenvectors that correspond to the m_y largest eigenvalues of the $T \times T$ matrix $\sum_{i=1}^{N} \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' / NT$.

The second-step IV estimator is defined as

$$\hat{\hat{\boldsymbol{\theta}}}_{IV} = \left(\hat{\mathbf{A}}_{NT}' \hat{\mathbf{B}}_{NT}^{-1} \hat{\mathbf{A}}_{NT}\right)^{-1} \hat{\mathbf{A}}_{NT}' \hat{\mathbf{B}}_{NT}^{-1} \hat{\mathbf{g}}_{NT}, \tag{17}$$

where

$$\hat{\mathbf{A}}_{NT} = \frac{1}{NT}\sum_{i=1}^{N} \hat{\mathbf{Z}}_i' \mathbf{M}_{\hat{F}_y} \mathbf{W}_i,\ \hat{\mathbf{B}}_{NT} = \frac{1}{NT}\sum_{i=1}^{N} \hat{\mathbf{Z}}_i' \mathbf{M}_{\hat{F}_y} \hat{\mathbf{Z}}_i,\ \hat{\mathbf{g}}_{NT} = \frac{1}{NT}\sum_{i=1}^{N} \hat{\mathbf{Z}}_i' \mathbf{M}_{\hat{F}_y} \mathbf{y}_i, \tag{18}$$

with

$$\mathbf{M}_{\hat{F}_y} = \mathbf{I}_T - \hat{\mathbf{F}}_y \left(\hat{\mathbf{F}}_y' \hat{\mathbf{F}}_y\right)^{-1} \hat{\mathbf{F}}_y'. \tag{19}$$

In order to derive the consistency of $\hat{\hat{\boldsymbol{\theta}}}_{IV}$, we use again Eqs. (6) and (17) to obtain:

$$\sqrt{NT}\left(\hat{\hat{\boldsymbol{\theta}}}_{IV} - \boldsymbol{\theta}\right) = \left(\hat{\mathbf{A}}_{NT}' \hat{\mathbf{B}}_{NT}^{-1} \hat{\mathbf{A}}_{NT}\right)^{-1} \hat{\mathbf{A}}_{NT}' \hat{\mathbf{B}}_{NT}^{-1} \left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \hat{\mathbf{Z}}_i' \mathbf{M}_{\hat{F}_y} \mathbf{u}_i\right). \tag{20}$$

The asymptotic property of the key term $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \hat{\mathbf{Z}}_i' \mathbf{M}_{\hat{F}_y} \mathbf{u}_i$ in Eq. (20) is stated in the following proposition.

Proposition 2 *Under Assumptions 1–5, as* $(N,T) \xrightarrow{j} \infty$ *such that* $N/T \to c$ *with* $0 < c < \infty$*,*

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \hat{\mathbf{Z}}_i' \mathbf{M}_{\hat{F}_y} \mathbf{u}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{M}_{F_y^0} \boldsymbol{\varepsilon}_i + o_p\left(1\right),$$

where $\hat{\mathbf{Z}}_i$ *is defined by Eq. (13).*

[14] Of course, we could estimate these bias terms consistently.

We see from Proposition 2 that the estimation effect in $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y}\mathbf{u}_i$ can be ignored asymptotically. Since $\boldsymbol{\varepsilon}_i$ is independent of $\mathbf{Z}_i$ and $\mathbf{F}_y^0$ with zero mean, the limiting distribution of $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y}\mathbf{u}_i$ is centred at zero. The following theorem provides asymptotic normality of the distribution of $\hat{\hat{\boldsymbol{\theta}}}_{IV}$, based on Hansen's (2007) law of large numbers and central limit theorem, which are restated as Lemmas 1 and 2 in Appendix A.

Theorem 2 *Suppose that Assumptions 1–5 hold true under the model in Eqs. (1)–(3). Then, as* $(N,T)\xrightarrow{j}\infty$ *such that* $N/T\to c$ *with* $0<c<\infty$,

(i)

$$\sqrt{NT}\left(\widehat{\hat{\boldsymbol{\theta}}}_{IV}-\boldsymbol{\theta}\right)\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Psi}\right),$$

where $\hat{\hat{\boldsymbol{\theta}}}_{IV}$ *is defined by Eq. (17) and*

$$\boldsymbol{\Psi}=\left(\mathbf{A}'\mathbf{B}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}'\mathbf{B}^{-1}\boldsymbol{\Omega}\mathbf{B}^{-1}\mathbf{A}\left(\mathbf{A}'\mathbf{B}^{-1}\mathbf{A}\right)^{-1},$$

is a positive definite matrix, where $\mathbf{A}=\text{plim}_{N,T\to\infty}\hat{\mathbf{A}}_{NT}$ *and* $\mathbf{B}=\text{plim}_{N,T\to\infty}\hat{\mathbf{B}}_{NT}$ *with* $\hat{\mathbf{A}}_{NT}$ *and* $\hat{\mathbf{B}}_{NT}$ *defined in Eq. (18), and* $\boldsymbol{\Omega}$ *is defined in Assumption 5.*

(ii) $\widehat{\boldsymbol{\Psi}}_{NT}-\boldsymbol{\Psi}\xrightarrow{p}\mathbf{0}$, *where*

$$\widehat{\boldsymbol{\Psi}}_{NT}=\left(\hat{\mathbf{A}}_{NT}'\hat{\mathbf{B}}_{NT}^{-1}\hat{\mathbf{A}}_{NT}\right)^{-1}\hat{\mathbf{A}}_{NT}'\hat{\mathbf{B}}_{NT}^{-1}\hat{\boldsymbol{\Omega}}_{NT}\hat{\mathbf{B}}_{NT}^{-1}\hat{\mathbf{A}}_{NT}\left(\hat{\mathbf{A}}_{NT}'\hat{\mathbf{B}}_{NT}^{-1}\hat{\mathbf{A}}_{NT}\right)^{-1}, \tag{21}$$

with

$$\hat{\boldsymbol{\Omega}}_{NT}=\frac{1}{NT}\sum_{i=1}^{N}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{M}_{\hat{F}_y}\hat{\mathbf{Z}}_i \tag{22}$$

and $\hat{\mathbf{u}}_i=\mathbf{y}_i-\mathbf{W}_i\hat{\boldsymbol{\theta}}_{IV}$.

Observe that the estimator above is asymptotically unbiased.

Finally, we propose the optimal second-step estimator, which we recommend be used:[15]

$$\hat{\hat{\boldsymbol{\theta}}}_{IV2}=\left(\hat{\mathbf{A}}_{NT}'\hat{\boldsymbol{\Omega}}_{NT}^{-1}\hat{\mathbf{A}}_{NT}\right)^{-1}\hat{\mathbf{A}}_{NT}'\hat{\boldsymbol{\Omega}}_{NT}^{-1}\hat{\mathbf{g}}_{NT}, \tag{23}$$

where $\hat{\mathbf{A}}_{NT}$ and $\hat{\mathbf{B}}_{NT}$ are defined in Eq. (18) and $\hat{\boldsymbol{\Omega}}_{NT}$ is given in Eq. (22). The following corollary describes the asymptotic properties of the estimator:

Corollary 1 *Suppose that Assumptions 1–5 hold true under the model in Eqs. (1)–(3). Then, as* $(N,T)\xrightarrow{j}\infty$ *such that* $N/T\to c$ *with* $0<c<\infty$,

$$\sqrt{NT}\left(\hat{\hat{\boldsymbol{\theta}}}_{IV2}-\boldsymbol{\theta}\right)\xrightarrow{d}N\left(\mathbf{0},\left(\mathbf{A}'\boldsymbol{\Omega}^{-1}\mathbf{A}\right)^{-1}\right)$$

and

$$\hat{\mathbf{A}}_{NT}'\hat{\boldsymbol{\Omega}}_{NT}^{-1}\hat{\mathbf{A}}_{NT}-\mathbf{A}'\boldsymbol{\Omega}^{-1}\mathbf{A}\xrightarrow{p}\mathbf{0},$$

where $\hat{\hat{\boldsymbol{\theta}}}_{IV2}$ *is defined by Eq. (23),* $\mathbf{A}=\text{plim}_{N,T\to\infty}\hat{\mathbf{A}}_{NT}$ *and* $\boldsymbol{\Omega}$ *is defined in Assumption 5.*

[15] The optimality of the two-step estimator is conditioned upon the chosen finite set of instruments. Our second-step estimator can be seen as sub-optimal in that it does not exploit all available instruments. See Remark 4 for a related discussion.

The associated overidentifying restrictions test statistic is given by

$$S_{NT} = \frac{1}{NT}\left(\sum_{i=1}^{N} \hat{\hat{\mathbf{u}}}_i' \mathbf{M}_{\hat{F}_y} \hat{\mathbf{Z}}_i\right) \hat{\boldsymbol{\Omega}}_{NT}^{-1} \left(\sum_{i=1}^{N} \hat{\mathbf{Z}}_i' \mathbf{M}_{\hat{F}_y} \hat{\hat{\mathbf{u}}}_i\right), \tag{24}$$

where $\hat{\hat{\mathbf{u}}}_i = \mathbf{y}_i - \mathbf{W}_i \hat{\hat{\boldsymbol{\theta}}}_{IV2}$, and $\hat{\boldsymbol{\Omega}}_{NT}$ is defined by Eq. (22). The limit distribution of the overidentifying restrictions test statistic is established in the following theorem:

Theorem 3 *Suppose that Assumptions 1–5 hold true under the model in Eqs. (1)–(3). Then, as* $(N, T) \xrightarrow{j} \infty$ *such that* $N/T \to c$ *with* $0 < c < \infty$*,*

$$S_{NT} \xrightarrow{d} \chi^2_{k-1} \tag{25}$$

for $k > 1$*, where* S_{NT} *is defined in Eq. (24).*

Remark 7 The overidentifying restrictions test is particularly useful in our approach. Firstly, 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 y. Secondly, the orthogonality condition of the instruments is violated if the slope vector, $\boldsymbol{\theta}$, is cross-sectionally heterogeneous, meaning that the estimators proposed in this section may become inconsistent. In such cases, the test is expected to reject the null hypothesis.

The next section discusses the estimation of models with heterogeneous slope coefficients.

3 The Model with Heterogeneous Coefficients

We now turn our focus to a model with heterogeneous coefficients. Let

$$\mathbf{y}_i = \mathbf{W}_i \boldsymbol{\theta}_i + \mathbf{u}_i, \tag{26}$$

where $\mathbf{W}_i = (\mathbf{y}_{i,-1}, \mathbf{X}_i)$, $\mathbf{X}_i$ follows the factor structure defined in Eq. (5), $\boldsymbol{\theta}_i = (\rho_i, \boldsymbol{\beta}_i')'$ with $\sup_{1 \le i \le N} |\rho_i| < 1$, and $\mathbf{u}_i$ is defined by Eq. (4). It is known widely that the pooled estimator, including $\hat{\hat{\boldsymbol{\theta}}}_{IV2}$, will be inconsistent for dynamic panel data models with, say, $\boldsymbol{\theta} = E(\boldsymbol{\theta}_i)$, if the slopes are cross-sectionally heterogeneous.[16] Henceforth, we introduce an estimator of $\boldsymbol{\theta}_i$ and propose a mean group IV estimator of the population average of $\boldsymbol{\theta}_i$. Thus, consistency and asymptotic normality are both established.

To begin with, we employ the following additional assumptions about the heterogeneous slopes, $\boldsymbol{\theta}_i$:

Assumption 6 (random coefficients): (i) $\boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{\eta}_i$, $\boldsymbol{\eta}_i \sim i.i.d.(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$, where $\boldsymbol{\Sigma}_\eta$ is a fixed positive definite matrix; (ii) $\boldsymbol{\eta}_i$ is independent of $\boldsymbol{\Gamma}_{xi}^0$, $\boldsymbol{\gamma}_{yi}^0$, ε_{it}, $\mathbf{v}_{it}$, $\mathbf{e}_{f_x,t}$ and $\mathbf{e}_{f_y,t}$; and (iii) $\boldsymbol{\eta}_i$ satisfies the tail bound:

$$P(|\eta_{ir}| > z) \le 2\exp\left(-\frac{1}{2} \times \frac{z^2}{a + bz}\right)$$

for all z (and all i) and fixed $a, b > 0$, where η_{ir} is the rth element of $\boldsymbol{\eta}_i$ for $2 \le r \le k+1$.

Assumption 7 (moment condition): (i) $E\|\boldsymbol{\eta}_i\|^4 \le \Delta$; (ii) $E\|T^{-1/2}\mathbf{V}_i'\mathbf{F}_x^0\|^4 \le \Delta$; and (iii) $E\|N^{-1/2}T^{-1/2}\sum_{\ell=1}^{k}\sum_{j=1}^{N}(\mathbf{V}_i'\mathbf{v}_{\ell j} - E(\mathbf{V}_i'\mathbf{v}_{\ell j}))\boldsymbol{\gamma}_{\ell j}^{0\prime}\|^4 \le \Delta$. In addition, (iv) $E(T^{-1/2}\sum_{\ell=1}^{k}\sum_{t=1}^{T}(v_{\ell it}^2 - Ev_{\ell it}^2))^2 \le \Delta$.

[16] See Pesaran and Smith (1995).

Assumption 8 (identification of $\boldsymbol{\theta}_i$): $\mathbf{A}_i = p\lim_{T\to\infty} \tilde{\mathbf{A}}_{i,T}$ has full column rank, $\mathbf{B}_i = p\lim_{T\to\infty} \tilde{\mathbf{B}}_{i,T}$, and $\boldsymbol{\Sigma}_i = p\lim_{T\to\infty} T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i\mathbf{u}_i'\mathbf{M}_{F_x^0}\mathbf{Z}_i'$ are positive definite, uniformly.

Assumptions 6(i)–(ii) are standard in the random coefficients literature; see for example Pesaran (2006). Assumptions 6(iii), 7 and 8 are required in order for the estimators of $\boldsymbol{\theta}_i$ to tend to their limiting distributions, uniformly.

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

$$\hat{\boldsymbol{\theta}}_{IV,i} = \left(\hat{\tilde{\mathbf{A}}}_{i,T}'\hat{\tilde{\mathbf{B}}}_{i,T}^{-1}\hat{\tilde{\mathbf{A}}}_{i,T}\right)^{-1}\hat{\tilde{\mathbf{A}}}_{i,T}'\hat{\tilde{\mathbf{B}}}_{i,T}^{-1}\hat{\tilde{\mathbf{g}}}_{i,T}, \tag{27}$$

where

$$\hat{\tilde{\mathbf{A}}}_{i,T} = \frac{1}{T}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_x}\mathbf{W}_i,\ \hat{\tilde{\mathbf{B}}}_{i,T} = \frac{1}{T}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_x}\hat{\mathbf{Z}}_i,\ \hat{\tilde{\mathbf{g}}}_{i,T} = \frac{1}{T}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_x}\mathbf{y}_i. \tag{28}$$

We can see from Eq. (28) that the instrument set here is $\mathbf{M}_{\hat{F}_x}\hat{\mathbf{Z}}_i$. This is tantamount to making use of $\hat{\mathbf{Z}}_i$ for the model in Eq. (26) premultiplied by $\mathbf{M}_{\hat{F}_x}$, which is expected to lead to a more efficient first-step IV estimator of $\boldsymbol{\theta}_i$ if the span of $\mathbf{F}_y^0$ includes a subset of $\mathbf{F}_x^0$.[17] Using Eqs. (26) and (27), we have

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{IV,i} - \boldsymbol{\theta}_i\right) = \left(\hat{\tilde{\mathbf{A}}}_{i,T}'\hat{\tilde{\mathbf{B}}}_{i,T}^{-1}\hat{\tilde{\mathbf{A}}}_{i,T}\right)^{-1}\hat{\tilde{\mathbf{A}}}_{i,T}'\hat{\tilde{\mathbf{B}}}_{i,T}^{-1}\left(T^{-1/2}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_x}\mathbf{u}_i\right). \tag{29}$$

The limiting property of $T^{-1/2}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_x}\mathbf{u}_i$ is given by the following proposition.

Proposition 3 *Consider the model in Eq. (26). Under Assumptions 1–6, as $(N,T) \xrightarrow{j} \infty$ such that $N/T \to c$ with $0 < c < \infty$, we have*

$$T^{-1/2}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_x}\mathbf{u}_i = T^{-1/2}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i + \sqrt{T}O_p\left(\delta_{NT}^{-2}\right),$$

where $\hat{\mathbf{Z}}_i$, $\mathbf{M}_{\hat{F}_x}$ and $\mathbf{Z}_i$ are defined by Eqs. (8), (12) and (13), respectively, and $\delta_{NT} = \min\left\{\sqrt{T}, \sqrt{N}\right\}$.

Using the result stated in Proposition 3 we see that $T^{-1/2}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_x}\mathbf{u}_i$ is $O_p(1)$ and tends to a random vector as $(N,T) \xrightarrow{j} \infty$, such that $N/T \to c$ with $0 < c < \infty$. The formal result is summarised in Theorem 4.

Theorem 4 *Consider the model in Eq. (26) and suppose that Assumptions 1–8 hold true. Then, as $(N,T) \xrightarrow{j} \infty$ such that $N/T \to c$ with $0 < c < \infty$, for each i,*

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{IV,i} - \boldsymbol{\theta}_i\right) \xrightarrow{d} N\left(\mathbf{0}, (\mathbf{A}_i'\mathbf{B}_i^{-1}\mathbf{A}_i)^{-1}\mathbf{A}_i'\mathbf{B}_i^{-1}\boldsymbol{\Sigma}_i\mathbf{B}_i^{-1}\mathbf{A}_i(\mathbf{A}_i'\mathbf{B}_i^{-1}\mathbf{A}_i)^{-1}\right), \tag{30}$$

where $\hat{\boldsymbol{\theta}}_{IV,i}$ is defined in Eq. (27), and $\mathbf{A}_i$, $\mathbf{B}_i$ and $\boldsymbol{\Sigma}_i$ are defined in Assumption 8.

Therefore, the estimator $\hat{\boldsymbol{\theta}}_{IV,i}$ is $\sqrt{T}$-consistent with $\boldsymbol{\theta}_i$.

Using a similar line of argument as in the discussion of the IV estimator in Section 2, we could consider a mean group IV estimator using the second-step estimator in an attempt to project $\mathbf{F}_y^0$ out from the model asymptotically, i.e. $\mathbf{M}_{F_y^0}\mathbf{y}_i = \mathbf{M}_{F_y^0}\mathbf{W}_i\boldsymbol{\theta}_i + \mathbf{M}_{F_y^0}\mathbf{u}_i$, and then use our IV method to estimate $\boldsymbol{\theta}_i$. However, the need to deal with heterogeneous slopes means that $\mathbf{F}_y$ should be estimated using the residuals from the time series IV regression, $\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{W}_i\hat{\boldsymbol{\theta}}_{IV,i}$. Since $\hat{\boldsymbol{\theta}}_{IV,i}$

[17] We could construct the first-step IV estimator of $\boldsymbol{\theta}$ in Section 2 using $\mathbf{M}_{\hat{F}_x}\hat{\mathbf{Z}}_i$ instead of $\hat{\mathbf{Z}}_i$, but the second-step estimator would be asymptotically equivalent to the proposed one that is based on $\mathbf{M}_{\hat{F}_y}\hat{\mathbf{Z}}_i$ when the span of $\mathbf{F}_y^0$ includes a subset of $\mathbf{F}_x^0$.

is $\sqrt{T}$-consistent rather than $\sqrt{NT}$-consistent, the estimation of $\mathbf{F}_y$ may become very inefficient. As a result, we will not pursue such an estimator here. Note that the estimation of $\mathbf{F}_x$ for the first-step IV estimator does not suffer from a similar problem, because it can be estimated using the raw data $\{\mathbf{X}_i\}_{i=1}^N$.

The mean group estimator of $\boldsymbol{\theta}$ is defined as

$$\hat{\boldsymbol{\theta}}_{IVMG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_{IV,i}, \tag{31}$$

where $\hat{\boldsymbol{\theta}}_{IV,i}$ is given in Eq. (27). It can be shown from Eqs. (26) and (29) and Assumptions 1–8 that[18]

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{IVMG} - \boldsymbol{\theta}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\eta}_i + o_p(1). \tag{32}$$

It is easy to see that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\eta}_i \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$ as $N \to \infty$, which implies that $\hat{\boldsymbol{\theta}}_{IVMG}$ is $\sqrt{N}$-consistent. The asymptotic normality of $\hat{\boldsymbol{\theta}}_{IVMG}$ and the consistency of an estimator of $\boldsymbol{\Sigma}_\eta$ are summarised in the following theorem:

Theorem 5 *Consider the model in Eq. (26) combined with Eq. (5), and suppose that Assumptions 1–7 hold true. Then, as $(N,T) \xrightarrow{j} \infty$ such that $N/T \to c$ with $0 < c < \infty$,*

(i)

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{IVMG} - \boldsymbol{\theta}\right) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \tag{33}$$

where $\hat{\boldsymbol{\theta}}_{IVMG}$ is defined in Eq. (31); and

(ii)

$$\widehat{\boldsymbol{\Sigma}}_\eta - \boldsymbol{\Sigma}_\eta \xrightarrow{p} \mathbf{0}, \tag{34}$$

where

$$\widehat{\boldsymbol{\Sigma}}_\eta = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{IV,i} - \hat{\boldsymbol{\theta}}_{IVMG}\right)\left(\hat{\boldsymbol{\theta}}_{IV,i} - \hat{\boldsymbol{\theta}}_{IVMG}\right)', \tag{35}$$

and $\hat{\boldsymbol{\theta}}_{IV,i}$ and $\hat{\boldsymbol{\theta}}_{IVMG}$ are given by Eqs. (27) and (31), respectively.

4 Monte Carlo Experiments

This section investigates the finite sample behaviour of the proposed estimators by means of Monte Carlo experiments, based on the bias, the root mean squared error (RMSE), and the empirical size and power of the t-test. In particular, we examine the optimal two-step IV estimator (IV2), which is defined in Eq. (23), and the mean group IV estimator (IVMG), defined in Eq. (31). We investigate the effects of the choice of the number of instruments (see Remark 4) by considering two sets of instruments for IV2 and IVMG, $\hat{\mathbf{Z}}_i$:

$$\text{IV set } a\text{: } \left(\mathbf{M}_{\hat{F}_x}\mathbf{X}_i; \mathbf{M}_{\hat{F}_{x,-1}}\mathbf{X}_{i,-1}\right)\ (T\times 2k)$$

$$\text{IV set } b\text{: } \left(\mathbf{M}_{\hat{F}_x}\mathbf{X}_i, \mathbf{M}_{\hat{F}_{x,-1}}\mathbf{X}_{i,-1}, \mathbf{M}_{\hat{F}_{x,-2}}\mathbf{X}_{i,-2}\right)\ (T\times 3k). \tag{36}$$

The instrument sets used for the IV estimators are denoted by the superscripts a and b (e.g. IV2^b makes use of IV set b).

[18] See the proof of Theorem 5.

For the purposes of comparison, we also investigate the performance of the bias-corrected quasi maximum likelihood estimator (BC-QMLE) that was proposed recently by Moon and Weidner (2017), as well as the CCE mean group (CCEMG) estimator and its bias-corrected version (BC-CCEMG), put forward by Chudik and Pesaran (2015a).

The bias-corrected QMLE estimator, $\hat{\boldsymbol{\theta}}_{BC-QMLE}$, is defined as[19]

$$\hat{\boldsymbol{\theta}}_{BC-QMLE} = \hat{\boldsymbol{\theta}}_{QMLE} - \hat{\boldsymbol{b}}_{QMLE}, \tag{37}$$

where

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{QMLE} &= \arg\min_{\boldsymbol{\theta}\in\Theta} L_{NT}(\boldsymbol{\theta}); \\ L_{NT}(\boldsymbol{\theta}) &= \min_{\boldsymbol{\Gamma}_y, \mathbf{F}_y} \mathcal{L}_{NT}(\boldsymbol{\theta}, \boldsymbol{\Gamma}_y, \mathbf{F}_y); \\ \mathcal{L}_{NT}(\boldsymbol{\theta}, \boldsymbol{\Gamma}_y, \mathbf{F}_y) &= \min_{\mathbf{F}_y} \frac{1}{NT} \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{W}_i\boldsymbol{\theta})' \mathbf{M}_{F_y} (\mathbf{y}_i - \mathbf{W}_i\boldsymbol{\theta}), \end{aligned}$$

with $\boldsymbol{\Gamma}_y = \left(\boldsymbol{\gamma}_{y,1}, \ldots, \boldsymbol{\gamma}_{y,N}\right)'$, whereas the estimator of the bias, $\hat{\boldsymbol{b}}_{QMLE}$, is defined in Definition 1 of Moon and Weidner (2017). The t-test is computed using the estimator of the variance-covariance matrix for $\hat{\boldsymbol{\theta}}_{BC-QMLE}$ (Moon and Weidner, 2017, p. 174). It should be highlighted that Moon and Weidner (2017) do not assume a linear factor process in $\mathbf{x}_{it}$, which is specified by Eq. (3), and therefore they may permit more general processes for the covariates.

The CCEMG estimator is given by

$$\hat{\boldsymbol{\theta}}_{CCEMG} = N^{-1} \sum_{i=1}^{N} \hat{\boldsymbol{\theta}}_{CCE,i}, \tag{38}$$

where $\hat{\boldsymbol{\theta}}_{CCE,i} = \left(\mathbf{W}_i'\mathbf{M}_{\bar{H}}\mathbf{W}_i\right)^{-1} \mathbf{W}_i'\mathbf{M}_{\bar{H}}\mathbf{y}_i$, $\mathbf{M}_{\bar{H}} = \mathbf{I}_T - \bar{\mathbf{H}}\left(\bar{\mathbf{H}}'\bar{\mathbf{H}}\right)^{-1}\bar{\mathbf{H}}'$, $\bar{\mathbf{H}} = N^{-1}\sum_{i=1}^{N}\mathbf{H}_i$. $\mathbf{H}_i$ contains $(\mathbf{y}_i; \mathbf{X}_i)$ and their lags:

$$\mathbf{H}_i = \left(\mathbf{y}_i, \mathbf{y}_{i,-1}, ..., \mathbf{y}_{i,-p_y}, \mathbf{X}_i, \mathbf{X}_{i,-1}, ..., \mathbf{X}_{i,-p_x}, \boldsymbol{\iota}_T\right), \tag{39}$$

where $\boldsymbol{\iota}_T$ is a $T\times 1$ vector of ones, $\mathbf{y}_{i,-j} = L^j\mathbf{y}_i$ and $\mathbf{X}_{i,-j} = L^j\mathbf{X}_i$. In view of the strict exogeneity of $\mathbf{X}_i$ in our experimental design, which is discussed shortly below, we include $p_y = p$ lags of $\mathbf{y}_i$ but no lags of $\mathbf{X}_i$, namely $p_x = 0$ in $\mathbf{H}_i$; see Chudik and Pesaran (2015a, equation 38).[20] Following Chudik and Pesaran (2015a), we choose the integer part of $T^{1/3}$ as the value of p. The t-test is computed using the estimated variance-covariance matrix, $\widehat{\boldsymbol{\Sigma}}_{MGCCE} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_{CCE,i} - \hat{\boldsymbol{\theta}}_{MGCCE}\right)\left(\hat{\boldsymbol{\theta}}_{CCE,i} - \hat{\boldsymbol{\theta}}_{MGCCE}\right)'$. The bias-corrected CCEMG estimator, $\hat{\boldsymbol{\theta}}_{BC-CCEMG}$, is given by

$$\hat{\boldsymbol{\theta}}_{BC-CCEMG} = 2\hat{\boldsymbol{\theta}}_{CCEMG} - \frac{1}{2}\left(\hat{\boldsymbol{\theta}}_{CCEMG}^{(1)} + \hat{\boldsymbol{\theta}}_{CCEMG}^{(2)}\right), \tag{40}$$

where $\hat{\boldsymbol{\theta}}_{MGCCE}^{(1)}$ denotes the mean group CCE estimator computed from the first half of the available time period and $\hat{\boldsymbol{\theta}}_{MGCCE}^{(2)}$ that computed from the second half. See Chudik and Pesaran (2015a) for more details.[21]

Following Remark 2, the data are demeaned using the within transformation before computing the proposed IV estimators, in order to eliminate individual-specific effects. $\hat{m}_x$ and $\hat{m}_y$ are

[19] We are grateful to Martin Weidner for providing to us the computational algorithm for the BC-QMLE estimator.

[20] We have also considered $p_y = p_x = p$, such that $\mathbf{y}_i$ and $\mathbf{X}_i$ have the same number of lags in $\mathbf{H}_i$. The performance of the CCEMG estimator is slightly worse in this case. The results are reported in Tables C11–C12 in Appendix C.

[21] We are grateful to Alex Chudik for sharing with us his code for computing the (BC-)CCEMG estimator.

obtained in each replication, based on the eigenvalue ratio (ER) statistic proposed by Ahn and Horenstein (2013, p. 1207). In our experiment we set $m_x = 2$ and $m_y = 3$, as will be shown shortly. For the estimation, we set the maximum number of factors equal to three for $\hat{m}_x$ and four for $\hat{m}_y$. For the CCEMG estimator, we use the untransformed data, $(\mathbf{y}_i, \mathbf{W}_i)$, but include a $T\times 1$ vector of ones along with the cross-sectional averages, as described above. Finally, for the computation of BC-QMLE, we follow the practice of Moon and Weidner (2015) and use the within-transformed data, as in our IV estimators. To avoid introducing further uncertainty by estimating the number of factors in u_{it}, the BC-QMLE is computed using the true number of factors, m_y.

4.1 Design

We consider the following dynamic panel data model:

$$y_{it} = \alpha_i + \rho_i y_{it-1} + \sum_{\ell=1}^{k} \beta_{\ell i} x_{\ell it} + u_{it}; \quad u_{it} = \sum_{s=1}^{m_y} \gamma_{si}^0 f_{s,t}^0 + \varepsilon_{it}, \tag{41}$$

$i = 1, ..., N$, $t = -49, ..., T$, where

$$f_{s,t}^0 = \rho_{f,s} f_{s,t-1}^0 + (1 - \rho_{fs}^2)^{1/2} \zeta_{s,t}, \tag{42}$$

with $\zeta_{s,t} \sim i.i.d.N(0, 1)$ for $s = 1, ..., m_y$. We set $k = 2$ and $m_y = 3$, and set $\rho_{f,s} = 0.5$ for all s.

The idiosyncratic error, ε_{it}, is non-normal and heteroskedastic across both i and t, such that $\varepsilon_{it} = \varsigma_\varepsilon \sigma_{it}(\epsilon_{it} - 1)/\sqrt{2}$, $\epsilon_{it} \sim i.i.d.\chi_1^2$, with $\sigma_{it}^2 = \eta_i \varphi_t$, $\eta_i \sim i.i.d.\chi_2^2/2$, and $\varphi_t = t/T$ for $t = 0, 1, ..., T$ and unity otherwise.

It is straightforward to see that the average variance of ε_{it} depends only on ς_ε^2. Let π_u denote the proportion of the average variance of u_{it} that is due to ε_{it}. That is, we define $\pi_u := \varsigma_\varepsilon^2 / (m_y + \varsigma_\varepsilon^2)$. Thus, for example, $\pi_u = 3/4$ means that the variance of the idiosyncratic error accounts for 75% of the total variance in u. In this case, most of the variation in the total error is due to the idiosyncratic component, and the factor structure has relatively minor contribution. Solving in terms of ς_ε^2 yields

$$\varsigma_\varepsilon^2 = \frac{\pi_u}{(1 - \pi_u)} m_y. \tag{43}$$

We set ς_ε^2 such that $\pi_u \in \{1/4, 3/4\}$.

The process for the covariates is given by

$$x_{\ell it} = \mu_{\ell i} + \sum_{s=1}^{m_x} \gamma_{\ell si}^0 f_{s,t}^0 + v_{\ell it}; \ i = 1, 2, ..., N; t = -49, -48, ..., T, \tag{44}$$

for $\ell = 1, 2$.

We set $m_x = 2$. This implies that the first two factors in u_{it}, f_{1t}^0, f_{2t}^0, are also contained in $x_{\ell it}$, for $\ell = 1, 2$, whilst f_{3t}^0 is included in u_{it} only. Observe that, using the notation from earlier sections, $\mathbf{f}_{y,t}^0 = (f_{1t}^0, f_{2t}^0, f_{3t}^0)'$ and $\mathbf{f}_{x,t}^0 = (f_{1t}^0, f_{2t}^0)'$.

The idiosyncratic error of the process for the covariates are serially correlated, such that

$$v_{\ell it} = \rho_{v,\ell} v_{\ell it-1} + (1 - \rho_{v,\ell}^2)^{1/2} \varpi_{\ell it}, \quad \varpi_{\ell it} \sim i.i.d.N(0, \varsigma_v^2), \tag{45}$$

for $\ell = 1, 2$. We set $\rho_{v,\ell} = 0.5$ for all ℓ.

Initially, all individual-specific effects and factor loadings are generated as correlated and *mean-zero* random variables; these are distinguished using the superscript "*". In particular, the *mean-zero* individual-specific effects are drawn as

$$\alpha_i^* \sim i.i.d.N(0, (1 - \rho_i)^2), \ \ \mu_{\ell i}^* = \rho_{\mu,\ell} \alpha_i^* + (1 - \rho_{\mu,\ell}^2)^{1/2} \omega_{\ell i}, \tag{46}$$

where $\omega_{\ell i} \sim i.i.d.\ N(0,(1-\rho_i)^2)$, for $\ell = 1,2$. We set $\rho_{\mu,\ell} = 0.5$ for $\ell = 1,2$.
Moreover, the *mean-zero* factor loadings in u_{it} are generated as $\gamma_{si}^{0*} \sim i.i.d. N(0,1)$ for $s = 1,...,m_y = 3$, and the factor loadings in x_{1it} and x_{2it} are drawn as

$$\gamma_{1si}^{0*} = \rho_{\gamma,1s}\gamma_{3i}^{0*} + (1-\rho_{\gamma,1s}^2)^{1/2}\xi_{1si}; \quad \xi_{1si} \sim i.i.d. N(0,1), \tag{47}$$

$$\gamma_{2si}^{0*} = \rho_{\gamma,2s}\gamma_{si}^{0*} + (1-\rho_{\gamma,2s}^2)^{1/2}\xi_{2si}; \quad \xi_{2si} \sim i.i.d. N(0,1), \tag{48}$$

respectively, for $s = 1,...,m_x = 2$. The process in Eq. (47) allows the factor loadings on $f_{1,t}^0$ and $f_{2,t}^0$ in x_{1it} to be correlated with the factor loadings that correspond to the factor that is specific to u_{it}, $f_{3,t}^0$. On the other hand, Eq. (48) ensures that the factor loadings on $f_{1,t}^0$ and $f_{2,t}^0$ in x_{2it} are allowed to be correlated with the factor loadings that correspond to the same factors in u_{it}, $f_{1,t}^0$ and $f_{2,t}^0$. We consider $\rho_{\gamma,11} = \rho_{\gamma,12} \in \{0,\ 0.5\}$, while $\rho_{\gamma,21} = \rho_{\gamma,22} = 0.5$.

Finally, the factor loadings that enter the model are generated such that

$$\mathbf{\Gamma}_i^0 = \mathbf{\Gamma}^0 + \mathbf{\Gamma}_i^{0*}, \tag{49}$$

where

$$\mathbf{\Gamma}_i^0 = \begin{pmatrix} \gamma_{1i}^0 & \gamma_{11i}^0 & \gamma_{21i}^0 \\ \gamma_{2i}^0 & \gamma_{12i}^0 & \gamma_{22i}^0 \\ \gamma_{3i}^0 & 0 & 0 \end{pmatrix} \text{ and } \mathbf{\Gamma}_i^{0*} = \begin{pmatrix} \gamma_{1i}^{0*} & \gamma_{11i}^{0*} & \gamma_{21i}^{0*} \\ \gamma_{2i}^{0*} & \gamma_{12i}^{0*} & \gamma_{22i}^{0*} \\ \gamma_{3i}^{0*} & 0 & 0 \end{pmatrix}.$$

Observe that, using the notation from earlier sections, $\boldsymbol{\gamma}_{yi}^0 = (\gamma_{1i}^0, \gamma_{2i}^0, \gamma_{3i}^0)'$ and $\mathbf{\Gamma}_{x,i}^0 = (\boldsymbol{\gamma}_{1i}^0, \boldsymbol{\gamma}_{2i}^0)'$, with $\boldsymbol{\gamma}_{\ell i}^0 = (\gamma_{\ell 1i}^0, \gamma_{\ell 2i}^0)'$ for $\ell = 1,2$. Also, it is easy to see that the average of the factor loadings is given by $E\left(\mathbf{\Gamma}_i^0\right) = \mathbf{\Gamma}^0$. To ensure that the rank condition for CCEMG is satisfied, we set[22]

$$\mathbf{\Gamma}^0 = \begin{pmatrix} 1/4 & 1/4 & -1 \\ 1/2 & -1 & 1/4 \\ 1/2 & 0 & 0 \end{pmatrix}. \tag{50}$$

We note that our estimators and BC-QMLE do not require this condition, and we also consider the experiment with $\mathbf{\Gamma}^0 = \mathbf{0}$.[23]

In a similar manner, the individual effects that enter the data generating process are such that

$$\alpha_i = \alpha + \alpha_i^*, \quad \mu_{\ell i} = \mu_\ell + \mu_{\ell i}^*, \tag{51}$$

for $\ell = 1,2$, setting $\alpha = 1/2$, $\mu_1 = 1$, $\mu_2 = -1/2$.

The slope coefficients are generated as

$$\rho_i = \rho + \eta_{\rho i}, \quad \beta_{1i} = \beta_1 + \eta_{\beta_1 i} \text{ and } \beta_{2i} = \beta_2 + \eta_{\beta_2 i}. \tag{52}$$

We consider $\rho \in \{0.5, 0.8\}$. Following Bai (2009a), we set $\beta_1 = 3$ and $\beta_2 = 1$ as a benchmark case, though we also consider $\beta_1 = 3$ and $\beta_2 = 0$ in order to investigate the properties of the estimator when one of the slope coefficients is equal to zero.

For the homogeneous slopes design, we set $\rho_i = \rho$, $\beta_{1i} = \beta_1$ and $\beta_{2i} = \beta_2$. For the heterogeneous slopes design, we specify $\eta_{\rho i} \sim i.i.d.\ U[-c,+c]$ and

$$\eta_{\beta_\ell i} = \left[(2c)^2/12\right]^{1/2} \rho_\beta \xi_{\beta_\ell i} + \left(1-\rho_\beta^2\right)^{1/2} \eta_{\rho i},$$

where $\xi_{\beta_\ell i}$ is the standardised squared idiosyncratic error in $x_{\ell it}$, computed as

$$\xi_{\beta_\ell i} = \frac{\overline{v_{\ell i}^2} - \overline{v_\ell^2}}{\left[N^{-1}\sum_{i=1}^N \left(\overline{v_{\ell i}^2} - \overline{v_\ell^2}\right)^2\right]^{1/2}},$$

[22] See Assumption 6 of Chudik and Pesaran (2015a).
[23] The results are reported in Tables C7 and C8 in Appendix C.

with $\overline{v_{\ell i}^2} = T^{-1}\sum_{t=1}^{T} v_{\ell it}^2$, $\overline{v_\ell^2} = N^{-1}\sum_{i=1}^{N} \overline{v_{\ell i}^2}$, for $\ell = 1, 2$. We set $c = 1/5$, $\rho_\beta = 0.4$ for $\ell = 1, 2$.

Denoting $\rho_\upsilon = \rho_{\upsilon,\ell}$, $\ell = 1, 2$, we define the signal-to-noise ratio (SNR) for the homogeneous model, conditional on the factor structure and the individual-specific effects, as follows:

$$SNR := \frac{\text{var}\left[(y_{it} - \varepsilon_{it})\,|\mathcal{L}\right]}{\overline{\text{var}}\,(\varepsilon_{it})} = \frac{\left(\frac{\beta_1^2+\beta_2^2}{1-\rho_\upsilon^2}\right)\varsigma_\upsilon^2 + \frac{\varsigma_\varepsilon^2}{1-\rho_\upsilon^2} - \varsigma_\varepsilon^2}{\varsigma_\varepsilon^2}, \tag{53}$$

where $\mathcal{L}$ is the information set that contains the factor structure and the individual-specific effects,[24] and $\overline{\text{var}}\,(\varepsilon_{it})$ is the overall average of $E\left(\varepsilon_{it}^2\right)$ over i and t. Solving for ς_υ^2 yields

$$\varsigma_\upsilon^2 = \varsigma_\varepsilon^2 \left[SNR - \frac{\rho_\upsilon^2}{1-\rho_\upsilon^2}\right]\left(\frac{\beta_1^2+\beta_2^2}{1-\rho_\upsilon^2}\right)^{-1}. \tag{54}$$

We set $SNR = 4$, which lies within the range $\{3, 9\}$ considered by the simulation study of Bun and Kiviet (2006). We consider all combinations of (T, N) for $T \in \{25, 50, 100, 200\}$ and $N \in \{25, 50, 100, 200\}$.

We investigate the power of the overidentifying restrictions test, which is defined in Eq. (24), by considering violations of the null due to slope heterogeneity and endogeneity as a result of the contemporaneous correlation between $\mathbf{x}_{it}$ and ε_{it}. For the slope heterogeneity, we use the DGP specified in Eq. (52). For the case of endogeneity, we replace the DGP given by Eq. (45) with $\upsilon_{\ell it} = \rho_\upsilon \upsilon_{\ell it-1} + (1-\rho_\upsilon^2)^{1/2}\varpi_{\ell it} + \varepsilon_{it}$, where $\varpi_{\ell it}$ is as defined previously and $\ell = 1, 2$.

All of our results are obtained based on 2,000 replications, and all tests are conducted at the 5% significance level. For the size of the t-test, $H_0 : \rho = \rho^0$ (or $H_0 : \beta_\ell = \beta_\ell^0$ for $\ell = 1, 2$), where $\rho^0, \beta_1^0, \beta_2^0$ are the true parameter values. For the power of the test, we consider $H_0 : \rho = \rho^0 + 0.1$ (or $H_0 : \beta_\ell = \beta_\ell^0 + 0.1$ for $\ell = 1, 2$) against two-sided alternatives. The power of the t-test reported below is the size-corrected power, for which the 5% critical values used are obtained as the 2.5% and 97.5% quantiles of the empirical distribution of the t-ratio under the null hypothesis.[25]

4.2 Results

Tables 1–4 report the bias ($\times 100$) and RMSE results of IV2^b, the BC-QMLE of Moon and Weidner (2017), IVMG^b and the CCEMG of Chudik and Pesaran (2015a), as well as the size (nominal level is 5%) and power (size-adjusted) of the associated t-tests for the panel ARDL(1,0) model with $\rho = 0.5$, $\beta_1 = 3$, $\beta_2 = 1$ and $\pi_u = 3/4$.[26] We compare the sensitivity of the estimators to the correlation structure of the factor loadings in $\mathbf{x}_{it}$ and u_{it} by considering independent factor loadings in $\mathbf{x}_{it}$ and u_{it} in Tables 1 and 2, and correlated loadings in Tables 3 and 4.

We have investigated two different sets of instruments for our estimators, as Eq. (36) explains. IV2^a (IVMG^a) uses $2k$ instruments and IV2^b (IVMG^b) $3k$. As one might expect, the former has a smaller bias but the latter has a smaller dispersion. In terms of RMSE, the latter always performs better. Therefore, we only report results for IV2^b and IVMG^b.[27] Moreover, we do not report results for BC-CCEMG, since it did not reduce the bias of CCEMG in our experiments, nor did it mitigate the size-distortion of the associated t-tests.[28]

Table 1 reports results for the model under slope homogeneity. Panel A corresponds to ρ and Panel B to β_1. The results for β_2 are not reported because they are qualitatively similar to those

[24] Our reason for conditioning on these variables is that they influence both the composite error in the equation for the dependent variable and the covariates.

[25] The size-adjusted power is employed in this experiment because the finite T bias of the CCEMG and BC-QMLE estimators, and the size distortion of the associated statistical tests, often makes the power comparison too confusing.

[26] The results for the specifications where $\{\rho, \beta_1, \beta_2\} = \{0.8, 3, 1\}, \{0.5, 3, 0\}$ and $\pi_u = 1/4$ are very similar qualitatively. See Tables C1–C6 in Appendix C.

[27] The results for IV2^a and IVMG^a are reported in Tables C9 and C10 in Appendix C.

[28] The results for BC-CCEMG are provided in Tables C11 and C12 in Appendix C.

for β_1.[29] As can be seen, IV2^b appears to have virtually no bias. In particular, the largest reported value of the absolute bias ($\times 100$) is 0.1 for $T = 50, N = 25$. On the other hand, the absolute bias of BC-QMLE appears to be much larger, perhaps indicating that bias correction is not able to remove the bias completely under these circumstances. However, the absolute bias declines steadily with larger values of N and T. In terms of RMSE, BC-QMLE outperforms IV2^b and other estimators, which reflects the higher efficiency of the maximum likelihood approach over IV and least-squares. For larger values of N or T (especially N), though, the RMSE values of IV2^b are very close to those of BC-QMLE. The bias of IVMG^b is similar to that of BC-QMLE, whereas that of CCEMG tends to be much larger, especially when $N = 25$ or $T = 25, 50$. IVMG^b mostly outperforms CCEMG in terms of the RMSE.

In regard to inference, the size of the t-test that is associated with IV2^b is close to the nominal value in most cases, though moderate size distortions are observed for $N = 25$. The size of IVMG^b appears to be very accurate unless T is much smaller than N. In contrast, both BC-QMLE and CCEMG exhibit substantial size distortions, which may be attributable in part to the relatively large biases of these estimators. In view of these size distortions, we report the size-adjusted power. As expected, under slope parameter homogeneity, the power of the IV2^b and BC-QMLE estimators is higher than that of the MG-type estimators, at least when N and T are both relatively small.

Next, we turn our attention to Panel B of Table 1, which reports results for β_1. The bias of IV2^b is slightly larger for small N and T than in Panel A but it remains smaller than that of other estimators. For instance, the absolute bias of BC-QMLE is large when $N = 25$, although it declines steadily as the sample size increases. As a result, IV2^b mostly outperforms BC-QMLE in terms of the RMSE. Moreover, the size of the IV2^b is close to its nominal level, with moderate distortion for $N = 25$. In contrast, BC-QMLE suffers from large size distortions. In regards to heterogeneous estimators, the relative properties of the absolute bias of both IVMG^b and CCEMG are similar to those for ρ, in the sense that IVMG^b has a smaller bias than CCEMG. The size of the t-test of IVMG^b is very close to 5% for all combinations of N and T, whilst CCEMG exhibits moderate size distortions even for large values of N or T.

Table 2 reports results for the model with heterogeneous slopes. Note that IV2^b and BC-QMLE are not justified asymptotically in this case. This is confirmed in finite samples. In particular, it is evident that IV2^b exhibits a systematic bias, fluctuating around 0.01 across all combinations of N and T. The bias of BC-QMLE is much larger, reaching values close to 0.03 for ρ (Panel A), for large values of N and T. This outcome is accompanied by large size distortions for both estimators. In contrast, for IVMG^b and CCEMG the bias appears to behave in a similar manner to the homogeneous case in Table 1. IVMG^b continues to perform well in terms of size, whereas the size properties for CCEMG improve substantially compared to the homogeneous case, although they still deviate significantly from the nominal value at least for small values of N or T. Similar conclusions apply to Panel B, with the main difference being that the size of the CCEMG estimator is closer to its nominal level for all combinations of N and T, and the power of the t-test appears to be smaller across all estimators.

Now let us turn our attention to the case where the factor loadings in x_{1it} are correlated with those in u_{it}. The results for homogeneous slopes are reported in Table 3, while those for heterogeneous slopes are shown in Table 4. The performances of IV2^b and IVMG^b are very similar to those shown in Tables 1 and 2, which suggests that our approach is robust to such correlations in factor loadings. In contrast, for β_1, the performances of BC-QMLE and CCEMG appear to deteriorate when the factor loadings are correlated. For example, for $T = 100$ and $N = 25, 50, 100, 200$, the bias ($\times 100$) values for BC-QMLE are equal to -4.0, 2.1, -0.5 and 0.4 respectively, whereas the corresponding values in the uncorrelated loadings design (Table 1, Panel B) are -1.6, -0.5, 0.2 and 0.4. Consequently, IV2^b outperforms BC-QMLE in terms of RMSE and the size of the test.

[29] These are available from the authors upon request.

For the models with heterogeneous slopes, it is interesting to note that the bias of the CCEMG estimator for β_1 does not decrease as the sample size increases. For example, in the correlated loadings design with heterogeneous slopes (Table 4, Panel B), the bias($\times 100$) values of CCEMG for β_1 with $T = N = 25, 50, 100, 200$ are 0.7, -0.8, -1.3 and -1.5, whereas in the uncorrelated loadings design (Table 2, Panel B) they are 2.2, 1.3, 0.1 and -0.3. As a consequence, IVMG^b mostly outperforms CCEMG by a substantial margin in terms of bias, RMSE and size.

Finally, we look at the finite sample behaviour for the overidentifying restrictions test based on the IV2^b estimator, which is summarised in Table 5. As was emphasised in Remark 7, we would like the test to reject the null when the exogeneity assumption on $\mathbf{x}_{it}$ is violated and/or when the slope coefficients are cross-sectionally heterogeneous. Table 5 contains two column blocks: the left one, entitled IV2^a, shows results using $2k$ instruments, while the right block, entitled IV2^b, shows results using additional instruments that raise the total number of instruments to $3k$. The latter case provides for more degrees of freedom of the overidentifying restrictions test. As can be seen, the size of the test is sufficiently close to its nominal level for both sets of instruments. On the other hand, there appear to be substantial differences in terms of the power of the test against slope heterogeneity. In particular, when $2k$ instruments are employed, such that the degree of overidentification equals 1, the overidentifying restrictions test lacks power and has rejection frequencies of 4.7% and 5.0% for $N = T = 100$ and $N = T = 200$. This outcome may be related to the results of Newey (1985), which show that overidentifying restrictions tests can lack power for some directions when the number of degrees of freedom is too small compared to the dimension of the misspecification. Indeed, the power appears to increase dramatically when we add two more instruments, such that it rises to 23.9% and 77.2% for $N = T = 100$ and $N = T = 200$, respectively. Therefore, the overidentifying restrictions test statistics that are associated with the optimal second-step IV estimator appear to have satisfactory power to reject the null of slope homogeneity, unless the degrees of freedom of the test are very small. Finally, the test has substantial power for both sets of instruments when the exogeneity of $\mathbf{x}_{it}$ is violated; specifically, ε_{it} is correlated with $\mathbf{v}_{it}$. For example, the power of the test with $2k$ instruments is 45.2% and 95.9% for $N = T = 100$ and $N = T = 200$, respectively, whilst that with $3k$ instruments is 36.7% and 91.4%.

In conclusion, we recommend the use of the (optimal) second-step IV estimator, $\widehat{\hat{\boldsymbol{\theta}}}_{IV2}$, defined by Eq. (23), for slope homogeneous models, and the mean group IV estimator, $\hat{\boldsymbol{\theta}}_{IVMG}$, defined by Eq. (31), for slope heterogeneous models with a moderate number of degrees of freedom. This is because $\widehat{\hat{\boldsymbol{\theta}}}_{IV2}$ is more efficient than $\hat{\boldsymbol{\theta}}_{IVMG}$ in models with homogeneous slopes, but becomes unreliable for models with heterogeneous slopes. We also note that $\hat{\boldsymbol{\theta}}_{IVMG}$ and the associated t-test seem reliable for the models in our experiment with either heterogeneous or homogeneous slope coefficients. Both estimators appear to be reasonably precise, and, notably, robust in cases where factor loadings are mutually correlated. Typically, the size of the associated tests is far more accurate than those of BC-QMLE and CCEMG, and they have sound power. The choice between the two estimators depends on the assertion of heterogeneity in the slope coefficients. The overidentifying restrictions test associated with the optimal second-step IV estimator has good power to reject the null under slope heterogeneity with sufficient degrees of overidentification, which could be used as a guide.

5 Concluding Remarks

This paper develops two instrumental variable estimators for the consistent estimation of homogeneous and heterogeneous dynamic panel data models with a multifactor error structure, when both N and T are large. For models with homogeneous slope coefficients, we put forward a two-step IV estimator that is $\sqrt{NT}$-consistent. The proposed estimator requires no bias correction, unlike that of Moon and Weidner (2017). Similarly, for models with heterogeneous coefficients, we develop a mean group IV estimator that does not require any small-T bias correction, unlike Chudik and Pesaran (2015a).

The finite sample evidence reported here suggests that the proposed estimators perform reasonably well under all circumstances examined, and therefore form a good alternative method of estimation to existing approaches. In particular, relative to the alternative methods examined, both IV estimators appear to have little or negligible bias in most circumstances, and a correct size of the t-test. Furthermore, the experimental results of the overidentifying restrictions test show that it has high power when a key assumption of the model is violated, namely the exogeneity of x.

Naturally, it is recommended that the optimal two-step IV estimator be employed for slope homogeneous models and the mean group IV estimator be employed for slope heterogeneous models. This is because $\hat{\hat{\boldsymbol{\theta}}}_{IV2}$ is more efficient than $\hat{\boldsymbol{\theta}}_{IVMG}$ for homogeneous slope specification but becomes unreliable in models with heterogeneous slopes. We also note that $\hat{\boldsymbol{\theta}}_{IVMG}$ and the associated t-test seem reliable in our experiment for models with either heterogeneous or homogeneous slope coefficients. The choice of the estimators depends on the assertion of heterogeneity in the slope coefficients. The experimental results show that, in general, the overidentifying restriction test associated with the optimal second-step IV estimator has good power to reject the null of slope homogeneity, unless the degrees of freedom of the test are very small. Thus, the development of a direct test for slope heterogeneity is of importance. We leave this as an avenue for future research.

This paper has assumed that the covariates in the model are strongly exogenous with respect to the idiosyncratic errors. This assumption may not be too restrictive for many applications, but the possibility of relaxing it to weak exogeneity, such that $\mathbf{x}_{it} = \boldsymbol{\Gamma}_{xi}^{0\prime}\mathbf{f}_{x,t}^{0} + \boldsymbol{\kappa}\varepsilon_{i,t-1} + \mathbf{v}_{it}$ with $\boldsymbol{\kappa} = (\kappa_1, ..., \kappa_k)'$, may be of interest and it merits further investigation.

Finally, we note that our approach is quite general and can actually be applied to a large class of linear panel data models. For example, our method is applicable to the model considered by Pesaran (2006), Bai and Li (2014), and Westerlund and Urbain (2015), among others: $y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \boldsymbol{\gamma}_{yi}^{0\prime}\mathbf{f}_{y,t}^{0} + \varepsilon_{it}$ with $\mathbf{x}_{it} = \boldsymbol{\Gamma}_{xi}^{0\prime}\mathbf{f}_{x,t}^{0} + \mathbf{v}_{it}$. A comparison of our approach with the existing approaches mentioned above may be an interesting research theme.

Table 1: Bias, root mean squared error (RMSE) of IV2^b, bias-corrected QMLE, IVMG^b and CCEMG estimates and size and power of the associated t-tests, for the panel ARDL(1,0) model with homogeneous slopes with $\{\rho, \beta_1, \beta_2\} = \{0.5, 3, 1\}$, $\pi_u = 3/4$, independent factor loadings in x_{1it} & u_{it}

PANEL A: Results for ρ, homogeneous slopes with $\{\rho, \beta_1, \beta_2\} = \{0.5, 3, 1\}$ and $\pi_u = 3/4$

	IV2^b				QMLE				IVMG^b				CCEMG			
T,N	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
BIAS (×100)																
25	0.0	0.0	0.0	0.0	-0.5	-0.6	-0.8	-1.0	-0.5	-0.7	-0.6	-0.7	-3.2	-3.4	-3.7	-3.9
50	-0.1	0.0	0.0	0.0	0.0	-0.3	-0.4	-0.5	-0.5	-0.4	-0.4	-0.4	-0.8	-1.0	-1.2	-1.5
100	0.0	0.0	0.0	0.0	0.1	-0.1	-0.2	-0.3	-0.2	-0.2	-0.2	-0.2	0.4	0.1	-0.1	-0.4
200	0.0	0.0	0.0	0.0	0.2	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	0.9	0.7	0.4	0.1
RMSE (×100)																
25	3.2	2.2	1.7	1.1	1.5	1.2	1.2	1.2	3.3	2.4	1.8	1.4	4.1	4.0	4.0	4.1
50	2.1	1.4	1.0	0.7	1.0	0.8	0.6	0.6	2.3	1.6	1.2	0.9	1.8	1.5	1.5	1.6
100	1.4	1.0	0.7	0.4	0.7	0.5	0.4	0.3	1.4	1.1	0.7	0.5	1.1	0.9	0.6	0.6
200	1.0	0.7	0.4	0.3	0.5	0.4	0.3	0.2	1.0	0.7	0.5	0.4	1.2	0.9	0.7	0.4
SIZE: $H_0 : \rho = 0.5$ against $H_1 : \rho \neq 0.5$, at the 5% level																
25	9.5	7.4	6.9	4.8	18.3	22.7	37.4	59.1	5.5	6.1	7.5	10.7	28.2	51.1	81.3	97.7
50	10.2	6.0	6.0	5.7	13.5	15.1	22.8	45.0	6.5	6.4	6.9	9.2	13.8	22.7	49.8	81.8
100	8.4	6.4	6.3	5.2	13.5	14.8	17.4	27.0	5.6	5.8	6.4	7.3	12.4	14.4	18.4	39.0
200	9.8	6.6	5.7	5.6	16.8	11.6	12.2	17.2	6.2	4.9	4.7	7.3	32.8	36.1	34.5	24.7
POWER (size-adjusted) : $H_0 : \rho = 0.6$ against $H_1 : \rho \neq 0.6$, at the 5% level																
25	87.5	98.2	99.9	100.0	99.8	100.0	100.0	100.0	80.7	94.8	99.5	100.0	48.2	54.0	57.4	63.2
50	98.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.1	99.9	100.0	100.0	99.7	100.0	100.0	100.0
100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

PANEL B: Results for β_1, homogeneous slopes $\{\rho, \beta_1, \beta_2\} = \{0.5, 3, 1\}$ and $\pi_u = 3/4$

	IV2^b				QMLE				IVMG^b				CCEMG			
T,N	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
BIAS (×100)																
25	-0.2	-0.2	-0.1	0.0	-2.3	-1.1	-0.5	0.0	1.3	1.0	1.5	1.4	2.9	2.9	3.3	3.2
50	0.2	0.1	0.1	0.0	-1.8	-0.5	-0.1	0.4	0.9	0.7	0.8	0.7	1.1	1.4	1.9	2.0
100	-0.1	0.0	0.1	0.0	-1.6	-0.5	0.2	0.4	0.3	0.4	0.4	0.4	-0.6	-0.2	0.3	0.7
200	-0.2	0.0	0.0	0.0	-1.5	-0.3	0.1	0.2	0.0	0.2	0.1	0.2	-1.8	-1.2	-0.8	-0.2
RMSE (×100)																
25	12.1	8.6	6.1	4.4	14.2	9.9	7.1	5.3	16.9	11.7	8.5	6.1	17.1	11.9	9.0	6.7
50	8.2	5.6	4.0	2.9	12.4	8.0	5.4	3.3	10.2	6.9	5.1	3.6	9.7	6.7	5.1	3.9
100	5.7	3.9	2.8	1.9	11.1	6.6	3.6	2.0	6.4	4.3	3.2	2.3	6.4	4.2	3.2	2.2
200	4.1	2.8	1.9	1.4	9.8	4.8	2.3	1.3	4.4	3.1	2.2	1.6	4.7	3.3	2.3	1.5
SIZE: $H_0 : \beta_1 = 3$ against $H_1 : \beta_1 \neq 3$, at the 5% level																
25	9.1	7.0	5.9	5.8	37.8	30.8	27.3	23.2	6.4	5.8	6.3	6.1	7.2	6.4	8.1	8.8
50	8.7	6.1	5.7	5.8	44.6	32.3	22.7	14.6	5.9	5.2	6.2	6.9	6.8	6.0	6.8	10.1
100	8.6	6.7	6.3	6.2	48.2	32.1	16.0	10.2	6.7	5.3	5.6	5.9	7.0	6.5	7.2	7.3
200	8.8	6.1	6.7	6.2	51.6	25.5	10.5	7.7	5.8	5.0	5.7	5.7	7.8	9.2	9.0	8.7
POWER (size-adjusted) : $H_0 : \beta_1 = 3.1$ against $H_1 : \beta_1 \neq 3.1$, at the 5% level																
25	17.6	25.4	43.2	66.8	6.9	12.4	19.6	40.4	11.1	18.2	30.7	49.4	12.5	21.9	32.4	54.0
50	27.0	47.4	72.2	92.6	8.1	15.6	38.5	88.2	22.3	38.6	59.5	85.0	21.9	41.4	70.1	89.8
100	47.8	73.6	94.0	100.0	8.4	24.8	83.4	99.8	36.3	65.8	90.0	99.5	34.7	62.9	89.8	99.7
200	71.3	95.0	99.8	100.0	9.4	49.0	99.0	100.0	62.7	90.7	99.2	100.0	45.3	76.4	97.5	100.0

Notes: The data generating process is $y_{it} = \alpha_i + \rho_i y_{it-1} + \sum_{\ell=1}^{2} \beta_{\ell i} x_{\ell it} + u_{it}$, $u_{it} = \sum_{s=1}^{3} \gamma_{si}^0 f_{st}^0 + \varepsilon_{it}$, $x_{\ell it} = \mu_{\ell i} + \sum_{s=1}^{2} \gamma_{\ell si}^0 f_{st}^0 + v_{\ell it}$ $\ell = 1, 2$; $i = 1,..,N$; $t = -50,..,T$ and the first 50 observations are discarded; $f_{st}^0 = \rho_{fs} f_{st-1}^0 + (1 - \rho_{fs}^2)^{1/2} \zeta_{st}$, $\zeta_{st} \sim i.i.d.N(0,1)$, $\gamma_{si}^0 = \gamma_s + \gamma_{si}^*$, $\gamma_{si}^* \sim i.i.d.N(0,1)$ for $s = 1, 2, 3$, $\varepsilon_{it} = \varsigma_\varepsilon \sigma_{it}(\epsilon_{it} - 1)/\sqrt{2}$, $\epsilon_{it} \sim i.i.d.\chi_1^2$ with $\sigma_{it}^2 = \eta_i \varphi_t$, $\eta_i \sim i.i.d.\chi_2^2/2$, and $\varphi_t = t/T$ for $t = 0, 1, ..., T$ and unity otherwise; $\gamma_{\ell si}^0 = \gamma_{\ell s} + \gamma_{\ell si}^{0*}$, $\gamma_{1si}^{0*} = \rho_{\gamma,1s} \gamma_{3i}^{0*} + (1 - \rho_{\gamma,1s}^2)^{1/2} \xi_{1si}$, $\gamma_{2si}^{0*} = \rho_{\gamma,2s} \gamma_{si}^{0*} + (1 - \rho_{\gamma,2s}^2)^{1/2} \xi_{2si}$, $\xi_{\ell si} \sim i.i.d.N(0,1)$, $v_{\ell it} = \rho_{v,\ell} v_{\ell it-1} + (1 - \rho_{v,\ell}^2)^{1/2} \varpi_{\ell it}$, $\varpi_{\ell it} \sim i.i.d.N(0, \varsigma_v^2 \sigma_{\varpi_{\ell,i}}^2)$, $\sigma_{\varpi_{\ell,i}}^2 \sim i.i.d.U[0.5, 1.5]$ for $\ell = 1, 2$, $s = 1, 2$. We set $\rho_i = \rho$, $\beta_{1i} = \beta_1$ and $\beta_{2i} = \beta_2$, $\rho_{fs} = \rho_{\gamma,2s} = \rho_{v,\ell} = 0.5$ and $\rho_{\gamma,1s} = 0.0$ for all ℓ, s. IV2^b and IVMG^b are given by (23), (31) with (36), BC-QMLE and CCEMG by (37), (38). The rank condition for CCEMG is met.

Table 2: Bias, root mean squared error (RMSE) of IV2^b, bias-corrected QMLE, IVMG^b and CCEMG estimates and size and power of the associated t-tests, for the panel ARDL(1,0) model with heterogeneous slopes with $\{\rho, \beta_1, \beta_2\} = \{0.5, 3, 1\}$, $\pi_u = 3/4$, independent factor loadings in x_{1it} & u_{it}

PANEL A: Results for ρ, homogeneous slopes with $\{\rho, \beta_1, \beta_2\} = \{0.5, 3, 1\}$ and $\pi_u = 3/4$																
	IV2^b				QMLE				IVMG^b				CCEMG			
T,N	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
	BIAS ($\times$100)															
25	0.7	0.8	0.9	0.8	1.1	0.8	0.5	0.3	-0.7	-0.7	-0.6	-0.7	-3.1	-3.3	-3.6	-3.8
50	1.1	1.1	1.1	1.1	2.0	1.9	1.8	1.8	-0.3	-0.3	-0.4	-0.4	-0.6	-0.9	-1.2	-1.5
100	1.0	1.2	1.2	1.2	2.5	2.5	2.3	2.3	-0.2	-0.2	-0.3	-0.2	0.4	0.3	-0.1	-0.3
200	1.1	1.2	1.3	1.3	2.7	2.6	2.6	2.6	-0.1	-0.1	-0.1	-0.1	1.0	0.8	0.5	0.2
	RMSE ($\times$100)															
25	4.4	3.2	2.3	1.7	4.1	2.9	2.2	1.7	4.2	3.0	2.1	1.6	4.8	4.3	4.1	4.1
50	3.4	2.6	2.0	1.5	4.0	3.1	2.5	2.1	3.1	2.2	1.6	1.2	2.8	2.2	1.9	1.8
100	3.0	2.3	1.8	1.6	4.1	3.3	2.8	2.6	2.7	1.9	1.4	1.0	2.5	1.9	1.3	1.0
200	2.9	2.1	1.8	1.6	4.1	3.4	3.0	2.8	2.5	1.7	1.2	0.9	2.6	1.8	1.3	0.9
	SIZE: $H_0 : \rho = 0.5$ against $H_1 : \rho \neq 0.5$, at the 5% level															
25	10.7	9.9	9.5	10.8	52.0	47.3	47.4	48.6	6.2	7.0	7.0	9.2	18.6	30.7	56.7	84.2
50	11.6	12.0	13.1	17.1	63.6	67.5	73.1	79.7	5.2	6.3	5.9	6.2	8.3	10.2	16.2	36.4
100	12.5	13.3	17.1	27.7	75.0	80.2	87.9	94.4	5.9	6.3	5.9	5.9	7.3	8.0	8.0	9.6
200	13.6	13.8	20.1	34.8	82.7	86.4	94.4	98.9	5.9	5.4	5.0	5.2	9.3	8.0	9.0	7.1
	POWER (size-adjusted) : $H_0 : \rho = 0.6$ against $H_1 : \rho \neq 0.6$, at the 5% level															
25	72.2	91.8	99.9	100.0	69.8	92.7	98.0	99.7	65.1	85.9	99.1	100.0	31.0	40.6	46.9	51.8
50	89.7	98.6	100.0	100.0	82.4	97.6	99.9	100.0	84.5	98.3	99.9	100.0	89.2	98.6	100.0	100.0
100	95.9	100.0	100.0	100.0	82.9	99.2	100.0	100.0	93.1	99.8	100.0	100.0	98.1	100.0	100.0	100.0
200	97.7	100.0	100.0	100.0	88.6	99.6	100.0	100.0	96.4	100.0	100.0	100.0	99.2	100.0	100.0	100.0
PANEL B: Results for β_1, homogeneous slopes $\{\rho, \beta_1, \beta_2\} = \{0.5, 3, 1\}$ and $\pi_u = 3/4$																
	IV2^b				QMLE				IVMG^b				CCEMG			
T,N	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
	BIAS ($\times$100)															
25	-1.9	-1.1	-1.3	-1.3	-4.9	-4.1	-4.4	-3.7	1.1	1.7	1.5	1.6	2.2	3.3	2.9	3.2
50	-1.6	-1.3	-1.1	-1.3	-5.4	-4.0	-3.6	-3.7	0.5	0.7	0.9	0.4	0.8	1.3	1.8	1.9
100	-1.0	-1.2	-1.2	-1.2	-4.7	-4.0	-3.5	-3.4	0.5	0.2	0.3	0.3	-0.4	-0.4	0.1	0.6
200	-1.0	-1.1	-1.1	-1.2	-5.0	-4.0	-3.7	-3.6	0.2	0.1	0.1	0.2	-1.6	-1.4	-0.9	-0.3
	RMSE ($\times$100)															
25	13.8	9.7	7.2	5.0	16.0	12.1	9.5	7.3	16.8	11.9	8.6	6.1	16.8	12.2	9.0	6.7
50	9.6	6.9	4.8	3.6	14.5	9.8	7.3	5.7	10.4	7.2	5.2	3.6	9.8	7.2	5.3	3.8
100	7.0	5.0	3.5	2.6	12.6	8.7	6.0	4.8	7.0	4.9	3.3	2.4	6.7	4.7	3.2	2.3
200	5.1	3.6	2.8	2.1	11.7	7.7	5.4	4.4	4.9	3.4	2.5	1.7	5.3	3.8	2.6	1.7
	SIZE: $H_0 : \beta_1 = 3$ against $H_1 : \beta_1 \neq 3$, at the 5% level															
25	11.0	8.0	8.1	7.3	41.6	37.3	37.6	37.0	5.6	5.4	6.1	6.8	6.7	6.5	7.6	7.8
50	9.8	7.4	6.4	8.0	47.5	39.9	35.6	41.2	5.9	5.6	6.0	4.9	6.1	6.3	7.5	9.1
100	9.5	8.8	7.3	8.6	52.8	47.4	42.9	50.3	5.4	5.5	5.0	5.4	5.7	6.4	5.3	6.3
200	9.5	7.3	9.5	11.6	60.4	54.7	56.0	67.6	6.1	5.8	5.3	4.9	8.5	9.4	8.6	6.8
	POWER (size-adjusted) : $H_0 : \beta_1 = 3.1$ against $H_1 : \beta_1 \neq 3.1$, at the 5% level															
25	10.5	15.2	24.2	40.5	4.1	5.5	4.5	7.4	12.1	19.4	29.7	47.6	12.8	19.7	30.9	53.4
50	14.7	28.5	46.5	70.1	3.3	5.6	7.5	13.0	19.9	33.6	56.4	83.5	19.8	38.1	62.6	89.9
100	28.1	46.7	73.7	94.5	4.4	6.4	12.7	22.4	36.1	57.9	86.6	99.1	33.4	51.3	87.7	99.5
200	45.3	68.2	92.4	99.8	4.5	6.4	13.7	27.5	52.2	81.2	97.9	100.0	34.6	64.0	94.4	99.9

Notes: The DGP is the same as that for Table 1 except that the slope coefficients are heterogeneous. Specifically, $\rho_i = \rho + \eta_{\rho i}$; $\beta_{\ell i} = \beta_\ell + \eta_{\beta_\ell i}$, $\eta_{\rho i} \sim i.i.d.U\left[-1/5, +1/5\right]$, and $\eta_{\beta_\ell i} = \left[(2/5)^2/12\right]^{1/2} \rho_\beta \xi_{\beta_\ell i} + (1-\rho_\beta^2)^{1/2} \zeta_{\beta_\ell i}$, where $\xi_{\beta_\ell i}$ is the standardised squared idiosyncratic errors in $x_{\ell i t}$, computed as $\xi_{\beta_\ell i} = (\overline{v_{\ell i}^2} - \overline{v_\ell^2})/[N^{-1}\sum_{i=1}^N (\overline{v_{\ell i}^2} - \overline{v_\ell^2})^2]^{1/2}$ with $\overline{v_{\ell i}^2} = T^{-1}\sum_{t=1}^T v_{\ell i t}^2$, $\overline{v_\ell^2} = N^{-1}\sum_{i=1}^N \overline{v_{\ell i}^2}$, for $\ell = 1, 2$, whereas $\zeta_{\beta_\ell i} \sim i.i.d.U\left(-\sqrt{3}, \sqrt{3}\right)$ for $\ell = 1, 2$.

Table 3: Bias, root mean squared error (RMSE) of IV2^b, bias-corrected QMLE, IVMG^b and CCEMG estimates and size and power of the associated t-tests, for the panel ARDL(1,0) model with homogeneous slopes with $\{\rho, \beta_1, \beta_2\} = \{0.5, 3, 1\}$, $\pi_u = 3/4$, correlated factor loadings in x_{1it} & u_{it}

PANEL A: Results for ρ, homogeneous slopes with $\{\rho, \beta_1, \beta_2\} = \{0.5, 3, 1\}$ and $\pi_u = 3/4$

	IV2^b				QMLE				IVMG^b				CCEMG			
T,N	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
	BIAS (×100)															
25	0.1	0.0	0.0	0.0	-0.4	-0.7	-0.9	-1.0	-0.7	-0.7	-0.6	-0.7	-3.2	-3.5	-3.6	-3.8
50	0.0	0.0	0.0	0.0	0.0	-0.3	-0.5	-0.6	-0.5	-0.4	-0.4	-0.3	-0.9	-1.0	-1.2	-1.4
100	0.0	0.0	0.0	0.0	0.1	-0.1	-0.2	-0.3	-0.3	-0.2	-0.2	-0.2	0.3	0.2	-0.1	-0.4
200	0.0	0.0	0.0	0.0	0.2	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	0.9	0.7	0.4	0.1
	RMSE (×100)															
25	3.1	2.2	1.6	1.1	1.6	1.4	1.3	1.3	3.4	2.6	1.9	1.5	4.3	4.1	4.0	4.0
50	2.1	1.4	1.0	0.7	1.1	0.8	0.7	0.7	2.3	1.5	1.1	0.9	1.8	1.6	1.5	1.6
100	1.4	1.0	0.6	0.4	0.7	0.5	0.4	0.4	1.5	1.0	0.7	0.6	1.1	0.9	0.7	0.6
200	1.0	0.6	0.4	0.3	0.6	0.4	0.3	0.2	1.0	0.7	0.5	0.4	1.2	0.9	0.6	0.4
	SIZE: $H_0 : \rho = 0.5$ against $H_1 : \rho \neq 0.5$, at the 5% level															
25	7.5	6.4	6.1	5.4	18.5	26.0	39.1	59.2	5.8	7.0	7.9	11.5	30.2	52.8	77.8	94.9
50	8.7	7.0	6.7	4.9	16.4	17.8	25.5	46.9	5.9	5.4	6.5	8.7	14.3	25.5	49.1	76.6
100	8.9	6.5	5.2	4.8	13.4	14.2	18.7	30.2	5.9	6.4	5.2	6.6	13.5	17.1	23.5	40.7
200	8.6	5.4	6.1	5.3	16.0	10.7	12.3	17.1	6.2	5.4	5.5	6.6	32.9	35.7	33.4	28.1
	POWER (size-adjusted) : $H_0 : \rho = 0.6$ against $H_1 : \rho \neq 0.6$, at the 5% level															
25	89.0	98.0	100.0	100.0	99.8	100.0	100.0	100.0	78.9	93.7	99.0	99.9	45.3	48.1	55.7	60.7
50	99.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.5	99.6	100.0	100.0	99.5	100.0	100.0	100.0
100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0
200	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

PANEL B: Results for β_1, homogeneous slopes $\{\rho, \beta_1, \beta_2\} = \{0.5, 3, 1\}$ and $\pi_u = 3/4$

	IV2^b				QMLE				IVMG^b				CCEMG			
T,N	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
	BIAS (×100)															
25	-0.2	0.2	0.0	-0.1	-4.4	-3.2	-2.9	-1.6	1.3	1.7	1.4	1.5	0.6	1.2	1.3	1.9
50	0.0	0.0	0.0	-0.1	-4.8	-2.9	-1.1	-0.1	0.9	0.7	0.8	0.5	-0.9	-0.4	0.2	0.7
100	0.2	-0.1	0.0	0.0	-4.0	-2.1	-0.5	0.4	0.6	0.2	0.3	0.4	-2.6	-2.3	-1.4	-0.5
200	0.0	0.0	0.0	0.0	-3.6	-1.6	-0.1	0.3	0.2	0.2	0.2	0.2	-3.9	-3.3	-2.4	-1.4
	RMSE (×100)															
25	11.8	8.7	6.1	4.3	16.6	12.4	9.9	7.4	16.9	11.9	8.5	6.2	16.6	12.4	8.8	6.4
50	8.1	5.6	4.0	2.7	14.5	10.7	7.3	4.7	10.0	6.8	5.0	3.5	10.1	7.1	4.9	3.5
100	5.8	3.9	2.8	1.9	13.4	8.6	4.9	2.5	6.4	4.4	3.2	2.2	7.2	5.2	3.6	2.3
200	3.9	2.8	1.9	1.4	12.2	6.8	3.1	1.4	4.3	3.1	2.1	1.5	6.3	4.8	3.2	2.1
	SIZE: $H_0 : \beta_1 = 3$ against $H_1 : \beta_1 \neq 3$, at the 5% level															
25	8.5	7.5	6.2	6.1	44.5	40.8	37.9	32.2	5.8	5.9	5.9	6.3	6.8	7.4	6.9	7.5
50	8.8	5.4	6.3	4.4	50.4	40.0	27.4	18.9	5.1	4.9	6.0	5.0	8.3	7.8	5.9	6.8
100	8.0	6.3	6.4	5.0	52.6	34.7	17.9	9.9	5.1	4.4	5.4	4.3	10.8	12.5	11.1	7.9
200	8.2	5.7	5.0	7.0	55.5	30.8	13.0	8.0	5.7	5.0	4.6	5.5	18.2	24.8	24.2	21.5
	POWER (size-adjusted) : $H_0 : \beta_1 = 3.1$ against $H_1 : \beta_1 \neq 3.1$, at the 5% level															
25	19.6	26.2	41.4	66.3	5.4	4.8	6.0	9.0	13.0	20.3	29.9	50.2	10.2	16.4	26.2	48.3
50	31.8	47.7	73.1	95.4	3.6	4.4	9.7	67.4	22.3	37.6	60.8	86.3	15.5	25.5	57.9	87.0
100	50.4	75.8	94.4	99.9	4.9	5.5	58.6	99.2	43.2	66.0	89.7	99.6	17.1	29.8	67.6	98.8
200	73.5	94.8	99.7	100.0	3.8	8.6	98.0	100.0	65.9	91.2	99.2	100.0	14.9	27.9	73.8	99.7

Notes: The DGP is the same as that for Table 1 except that the factor loadings in x_{1it} & u_{it} are correlated: $\rho_{\gamma,1s} = 0.5$ in $\gamma_{1si}^{0*} = \rho_{\gamma,1s}\gamma_{3i}^{0*} + (1 - \rho_{\gamma,1s}^2)^{1/2}\xi_{1si}$.

Table 4: Bias, root mean squared error (RMSE) of IV2^b, bias-corrected QMLE, IVMG^b and CCEMG estimates and size and power of the associated t-tests, for the panel ARDL(1,0) model with heterogeneous slopes with $\{\rho, \beta_1, \beta_2\} = \{0.5, 3, 1\}$, $\pi_u = 3/4$, correlated factor loadings in x_{1it} & u_{it}

PANEL A: Results for ρ, homogeneous slopes with $\{\rho, \beta_1, \beta_2\} = \{0.5, 3, 1\}$ and $\pi_u = 3/4$																
	IV2^b				QMLE				IVMG^b				CCEMG			
T,N	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
	BIAS (×100)															
25	0.8	0.9	0.9	0.9	0.9	0.8	0.5	0.2	-0.7	-0.7	-0.7	-0.7	-3.1	-3.3	-3.5	-3.8
50	0.9	1.1	1.2	1.2	2.2	2.1	1.9	1.9	-0.4	-0.3	-0.3	-0.3	-0.7	-0.9	-1.1	-1.3
100	1.1	1.2	1.2	1.2	2.8	2.7	2.6	2.6	-0.2	-0.3	-0.2	-0.2	0.4	0.2	-0.1	-0.3
200	1.1	1.3	1.3	1.3	3.1	2.9	2.8	2.9	-0.1	-0.1	-0.1	-0.1	0.9	0.8	0.4	0.2
	RMSE (×100)															
25	4.4	3.1	2.3	1.7	4.3	3.2	2.5	2.0	4.1	3.0	2.2	1.7	4.8	4.3	4.1	4.1
50	3.5	2.6	2.0	1.6	4.2	3.3	2.6	2.3	3.2	2.2	1.6	1.2	2.9	2.3	1.9	1.8
100	3.0	2.3	1.8	1.6	4.4	3.5	3.0	2.8	2.7	2.0	1.4	1.0	2.6	1.9	1.4	1.1
200	2.8	2.2	1.8	1.6	4.5	3.7	3.2	3.1	2.5	1.8	1.3	0.9	2.6	1.9	1.4	1.0
	SIZE: $H_0 : \rho = 0.5$ against $H_1 : \rho \neq 0.5$, at the 5% level															
25	11.7	9.2	9.3	10.5	53.2	50.6	51.5	53.7	6.3	6.6	7.4	8.9	17.3	30.0	54.9	81.9
50	12.7	11.5	12.9	19.8	66.2	67.6	72.3	81.5	6.2	5.1	6.8	6.8	9.0	11.6	17.3	35.3
100	11.6	13.1	15.9	26.3	77.5	82.0	87.6	95.8	5.9	5.6	5.5	6.2	8.6	8.2	8.3	13.0
200	12.3	14.1	21.9	33.3	84.4	89.3	93.7	99.0	5.5	5.3	4.9	5.4	9.4	10.1	10.9	9.6
	POWER (size-adjusted) : $H_0 : \rho = 0.6$ against $H_1 : \rho \neq 0.6$, at the 5% level															
25	71.9	92.8	99.6	100.0	64.8	90.1	96.6	99.0	61.2	87.9	98.3	100.0	30.9	41.5	44.3	46.1
50	87.2	98.9	100.0	100.0	78.4	97.3	99.9	100.0	80.7	98.2	100.0	100.0	89.1	98.8	100.0	100.0
100	96.3	100.0	100.0	100.0	83.5	99.0	100.0	100.0	93.6	99.7	100.0	100.0	97.6	100.0	100.0	100.0
200	98.6	100.0	100.0	100.0	84.8	99.7	100.0	100.0	97.3	100.0	100.0	100.0	99.2	100.0	100.0	100.0
PANEL B: Results for β_1, homogeneous slopes $\{\rho, \beta_1, \beta_2\} = \{0.5, 3, 1\}$ and $\pi_u = 3/4$																
	IV2^b				QMLE				IVMG^b				CCEMG			
T,N	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
	BIAS (×100)															
25	-1.2	-1.3	-1.4	-1.5	-7.6	-7.4	-6.5	-6.2	1.7	1.4	1.5	1.4	0.7	1.2	1.5	2.0
50	-1.0	-1.3	-1.2	-1.2	-8.5	-7.0	-5.9	-4.9	0.9	0.4	0.7	0.7	-0.8	-0.8	0.1	0.9
100	-1.2	-1.0	-1.1	-1.2	-8.0	-6.7	-5.2	-4.5	0.2	0.5	0.4	0.3	-2.7	-2.0	-1.3	-0.6
200	-1.3	-1.0	-1.1	-1.2	-7.9	-6.0	-4.9	-4.6	0.1	0.3	0.2	0.1	-3.9	-3.2	-2.3	-1.5
	RMSE (×100)															
25	13.4	9.8	6.8	5.1	18.9	15.2	12.1	10.3	16.9	11.9	8.5	6.1	17.0	12.3	9.0	6.4
50	9.6	6.9	4.8	3.5	17.5	13.5	10.0	7.6	10.2	7.1	5.2	3.7	10.2	7.4	5.1	3.8
100	6.9	5.1	3.5	2.6	15.9	11.4	8.1	6.0	6.7	4.8	3.4	2.3	7.7	5.5	3.8	2.6
200	5.3	3.8	2.7	2.1	15.1	9.7	6.7	5.5	5.0	3.6	2.5	1.7	6.9	5.0	3.5	2.4
	SIZE: $H_0 : \beta_1 = 3$ against $H_1 : \beta_1 \neq 3$, at the 5% level															
25	9.6	7.5	6.5	6.6	48.1	47.6	46.3	51.7	6.3	5.8	6.2	6.7	7.6	6.8	7.0	8.0
50	8.8	7.5	6.3	7.7	54.7	50.7	50.2	49.8	6.2	4.8	5.7	5.7	8.2	7.2	6.7	8.4
100	9.0	9.0	7.3	8.2	59.4	54.9	51.3	59.1	5.5	5.5	5.3	4.8	10.7	11.3	10.6	8.9
200	10.8	8.6	8.6	12.0	67.7	58.4	62.0	76.4	6.2	6.5	6.0	4.6	16.9	20.5	20.0	18.9
	POWER (size-adjusted) : $H_0 : \beta_1 = 3.1$ against $H_1 : \beta_1 \neq 3.1$, at the 5% level															
25	14.2	16.6	25.0	39.6	3.1	1.9	0.9	1.5	11.6	16.5	29.8	47.4	10.7	16.6	26.4	46.5
50	20.5	28.3	46.9	72.3	1.5	1.6	2.6	3.0	19.2	35.0	55.0	82.9	16.1	25.6	51.8	81.8
100	28.5	45.7	74.3	95.4	2.0	1.0	2.7	5.8	33.0	60.2	86.1	99.1	14.8	29.4	62.9	96.1
200	40.8	66.5	92.6	99.8	2.2	1.8	2.6	7.3	50.3	80.5	98.0	100.0	11.0	24.4	65.5	97.3

Notes: The DGP is the same as that for Table 2 except that the factor loadings in x_{1it} & u_{it} are correlated: $\rho_{\gamma,1s} = 0.5$ in $\gamma_{1si}^{0*} = \rho_{\gamma,1s}\gamma_{3i}^{0*} + (1 - \rho_{\gamma,1s}^2)^{1/2}\xi_{1si}$.

Table 5: Size and power of the overidentifying restrictions test for the panel ARDL(1,0) model with $\{\rho, \beta_1, \beta_2\} = \{0.5, 3, 1\}$, $\pi_u = 3/4$, correlated factor loadings in x_{1it} & u_{it}

	IV2^a				IV2^b			
T,N	25	50	100	200	25	50	100	200
	Slope Homogeneity (Size)							
25	7.2	6.9	5.7	6.4	6.8	6.0	5.7	5.5
50	7.7	6.6	6.5	4.8	7.3	6.0	5.5	4.4
100	6.7	7.3	5.9	4.8	7.4	6.2	5.9	5.1
200	7.7	6.6	6.5	5.6	7.0	5.9	6.1	4.7
	Slope Heterogeneity (Power)							
25	7.5	6.2	6.0	5.4	8.3	7.9	8.8	10.1
50	6.9	5.9	5.4	5.7	8.6	10.4	12.6	22.0
100	7.4	5.6	4.7	4.9	11.2	13.6	23.9	44.3
200	7.0	5.9	5.8	5.0	14.6	24.0	45.3	77.2
	Endogeneous Idiosyncratic Error of X (Power)							
25	10.1	10.4	14.6	18.8	10.4	11.0	14.8	18.9
50	12.3	16.1	23.5	37.8	11.9	15.9	20.0	31.4
100	17.2	27.6	45.2	70.3	14.4	20.7	36.7	60.2
200	28.2	46.4	73.2	95.9	22.5	37.6	62.8	91.4

Notes: The table reports the size and the power of overidentifying restrictions tests based on the IV2 estimator using different set of instruments. IV2^a uses $(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_{i,-1})$ and IV2^b $(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_{i,-1}, \hat{\mathbf{X}}_{i,-2})$, where $\hat{\mathbf{X}}_i = \mathbf{M}_{F_x}\mathbf{X}_i$ and $\hat{\mathbf{X}}_{i,-j} = \mathbf{M}_{F_{x,-j}}\mathbf{X}_{i,-j}$ for $j = 1, 2$. The test statistic is defined by (24). The tests for IV2^a and IV2^b are referenced to the 95% quantiles of χ^2_1 and χ^2_3 distributions, respectively. The DGP for Slope Homogeneity is of Table 3, for Slope Heterogeneity is of Table 4, and for Endogeneous Idiosyncratic Error of X, the DGP of Table 3 is changed such that $v_{\ell it} = \rho_{v,\ell} v_{\ell it-1} + (1-\rho^2_\ell)^{1/2}\varpi_{\ell it}$, $\varpi_{\ell it} = \tau_\ell \varepsilon_{it} + (1-\tau^2_\ell)^{1/2}\varrho_{\ell it}$ with $\varrho_{\ell it} \sim i.i.d.N(0,1)$, $\ell = 1, 2$ (see notes to Table 1). We set $\tau_1 = 0.5$ and $\tau_2 = 0$ so that the idiosyncratic error of x_{1it} is contemporaneously correlated with ε_{it}

Appendix A: Proofs of Main Results

Lemma 1 *Under Assumptions 1-5, as $(N,T) \xrightarrow{j} \infty$ such that $N/T \to c$ with $0 < c < \infty$,*

$$T^{-1}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y}\hat{\mathbf{Z}}_i - T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_y^0}\mathbf{Z}_i = o_p(1), \tag{A.1}$$

$$T^{-1}\left(\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y} - \mathbf{Z}_i'\mathbf{M}_{F_y^0}\right)\mathbf{W}_i = o_p(1). \tag{A.2}$$

Lemma 2 *Under Assumptions 1-5, as $(N,T) \xrightarrow{j} \infty$ such that $N/T \to c$ with $0 < c < \infty$,*

$$\frac{1}{NT}\sum_{i=1}^N \hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y}\hat{\mathbf{Z}}_i - \mathbf{Z}_i'\mathbf{M}_{F_y^0}\mathbf{Z}_i = o_p(1), \tag{A.3}$$

$$\frac{1}{NT}\sum_{i=1}^N \left(\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y} - \mathbf{Z}_i'\mathbf{M}_{F_y^0}\right)\mathbf{W}_i = o_p(1). \tag{A.4}$$

Lemma 3 *Under Assumptions 1-5, as $(N,T) \xrightarrow{j} \infty$ such that $N/T \to c$ with $0 < c < \infty$,*

$$\begin{aligned}&\frac{1}{\sqrt{N}T^{3/2}}\sum_{i=1}^N \mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^0\right)^{-1}\left(\frac{\hat{\mathbf{F}}_x'\mathbf{F}_x^0}{T}\right)^{-1}\hat{\mathbf{F}}_x'\left(\mathbf{\Sigma}_{kNT} - \bar{\mathbf{\Sigma}}_{kNT}\right)\mathbf{M}_{\hat{F}_x}\mathbf{u}_i \\ &= O_p\left(T^{-1/2}\right) + O_p\left(\delta_{NT}^{-1}\right) + \sqrt{T}O_p\left(\delta_{NT}^{-2}\right),\end{aligned} \tag{A.5}$$

and

$$\begin{aligned}&\frac{1}{\sqrt{N}T^{3/2}}\sum_{i=1}^N \mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^0\right)^{-1}\left(\frac{\hat{\mathbf{F}}_{x,-1}'\mathbf{F}_{x,-1}^0}{T}\right)^{-1}\hat{\mathbf{F}}_{x,-1}'\left(\mathbf{\Sigma}_{kNT,-1} - \bar{\mathbf{\Sigma}}_{kNT,-1}\right)\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{u}_i \\ &= O_p\left(T^{-1/2}\right) + O_p\left(\delta_{NT}^{-1}\right) + \sqrt{T}O_p\left(\delta_{NT}^{-2}\right),\end{aligned} \tag{A.6}$$

where $\mathbf{\Sigma}_{kNT}=\frac{1}{N}\sum_{\ell=1}^{k}\sum_{j=1}^{N}\mathbf{v}_{\ell j}\mathbf{v}_{\ell j}'$, $\mathbf{\Sigma}_{kNT,-1}=\frac{1}{N}\sum_{\ell=1}^{k}\sum_{j=1}^{N}\mathbf{v}_{\ell j,-1}\mathbf{v}_{\ell j,-1}'$ *and* $\bar{\mathbf{\Sigma}}_{kNT}=\frac{1}{N}\sum_{\ell=1}^{k}\sum_{j=1}^{N}E\left(\mathbf{v}_{\ell j}\mathbf{v}_{\ell j}'\right)$, $\bar{\mathbf{\Sigma}}_{kNT,-1}=\frac{1}{N}\sum_{\ell=1}^{k}\sum_{j=1}^{N}E\left(\mathbf{v}_{\ell j,-1}\mathbf{v}_{\ell j,-1}'\right)$.

Lemma 4 *Under Assumptions 1-5, as* $(N,T)\overset{j}{\to}\infty$ *such that* $N/T\to c$ *with* $0<c<\infty$,

$$\begin{aligned}&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\mathbf{F}_{x}^{0\prime}\mathbf{M}_{\hat{F}_x}\mathbf{u}_i\\&=-\frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\mathbf{\Gamma}_{xj}^{0}\mathbf{V}_j'\mathbf{M}_{\hat{F}_x}\mathbf{u}_i\\&-\frac{1}{\sqrt{N}T^{3/2}}\sum_{i=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\left(\frac{\hat{\mathbf{F}}_x'\mathbf{F}_x^0}{T}\right)^{-1}\hat{\mathbf{F}}_x'\mathbf{\Sigma}_{kNT}\mathbf{M}_{\hat{F}_x}\mathbf{u}_i+o_p(1),\end{aligned}\tag{A.7}$$

and

$$\begin{aligned}&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\mathbf{F}_{x,-1}^{0\prime}\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{u}_i\\&=-\frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\mathbf{\Gamma}_{xj}^{0}\mathbf{V}_{j,-1}'\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{u}_i\\&-\frac{1}{\sqrt{N}T^{3/2}}\sum_{i=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\left(\frac{\hat{\mathbf{F}}_{x,-1}'\mathbf{F}_{x,-1}^0}{T}\right)^{-1}\hat{\mathbf{F}}_{x,-1}'\mathbf{\Sigma}_{kNT,-1}\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{u}_i+o_p(1).\end{aligned}\tag{A.8}$$

Lemma 5 *Under Assumptions 1-5, as* $(N,T)\overset{j}{\to}\infty$ *such that* $N/T\to c$ *with* $0<c<\infty$,

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\mathbf{F}_{x}^{0\prime}\mathbf{M}_{\hat{F}_x}\mathbf{M}_{\hat{F}_y}\mathbf{u}_i=o_p(1),\tag{A.9}$$

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\mathbf{F}_{x,-1}^{0\prime}\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{M}_{\hat{F}_y}\mathbf{u}_i=o_p(1).\tag{A.10}$$

Lemma 6 *Under Assumptions 1-5, as* $(N,T)\overset{j}{\to}\infty$ *such that* $N/T\to c$ *with* $0<c<\infty$,

$$\begin{aligned}&\frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\mathbf{\Gamma}_{xj}^{0}\mathbf{V}_j'\mathbf{M}_{\hat{F}_x}\mathbf{u}_i\\&=\frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\mathbf{\Gamma}_{xj}^{0}\mathbf{V}_j'\mathbf{M}_{F_x^0}\mathbf{u}_i\\&-\sqrt{\frac{T}{N}}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{n=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\mathbf{\Gamma}_{xn}^{0}\frac{\mathbf{V}_n'\mathbf{V}_j}{T}\mathbf{\Gamma}_{xj}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\left(\frac{\mathbf{F}_x^{0\prime}\mathbf{F}_x^0}{T}\right)^{-1}\frac{\mathbf{F}_x^{0\prime}\mathbf{u}_i}{T}\\&+o_p(1),\end{aligned}\tag{A.11}$$

and

$$\begin{aligned}&\frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\mathbf{\Gamma}_{xj}^{0}\mathbf{V}_{j,-1}'\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{u}_i\\&=\frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\mathbf{\Gamma}_{xj}^{0}\mathbf{V}_{j,-1}'\mathbf{M}_{F_{x,-1}^0}\mathbf{u}_i\\&-\sqrt{\frac{T}{N}}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{n=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\mathbf{\Gamma}_{xn}^{0}\frac{\mathbf{V}_{n,-1}'\mathbf{V}_{j,-1}}{T}\mathbf{\Gamma}_{xj}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\left(\frac{\mathbf{F}_{x,-1}^{0\prime}\mathbf{F}_{x,-1}^0}{T}\right)^{-1}\frac{\mathbf{F}_{x,-1}^{0\prime}\mathbf{u}_i}{T}\\&+o_p(1).\end{aligned}\tag{A.12}$$

Lemma 7 *Under Assumptions 1-5, as* $(N,T) \xrightarrow{j} \infty$ *such that* $N/T \to c$ *with* $0 < c < \infty$,

$$\frac{1}{\sqrt{N}T^{3/2}} \sum_{i=1}^{N} \mathbf{\Gamma}_{xi}^{0\prime} \left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1} \left(\frac{\hat{\mathbf{F}}_x^{\prime}\mathbf{F}_x^0}{T}\right)^{-1} \hat{\mathbf{F}}_x^{\prime} \bar{\mathbf{\Sigma}}_{kNT} \mathbf{M}_{\hat{F}_x} \mathbf{u}_i$$
$$= \frac{1}{\sqrt{N}T^{3/2}} \sum_{i=1}^{N} \mathbf{\Gamma}_{xi}^{0\prime} \left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1} \left(\frac{\mathbf{F}_x^{0\prime}\mathbf{F}_x^0}{T}\right)^{-1} \mathbf{F}_x^{0\prime} \bar{\mathbf{\Sigma}}_{kNT} \mathbf{M}_{F_x^0} \mathbf{u}_i + o_p(1), \tag{A.13}$$

and

$$\frac{1}{\sqrt{N}T^{3/2}} \sum_{i=1}^{N} \mathbf{\Gamma}_{xi}^{0\prime} \left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1} \left(\frac{\hat{\mathbf{F}}_{x,-1}^{\prime}\mathbf{F}_{x,-1}^0}{T}\right)^{-1} \hat{\mathbf{F}}_{x,-1}^{\prime} \bar{\mathbf{\Sigma}}_{kNT,-1} \mathbf{M}_{\hat{F}_{x,-1}} \mathbf{u}_i$$
$$= \frac{1}{\sqrt{N}T^{3/2}} \sum_{i=1}^{N} \mathbf{\Gamma}_{xi}^{0\prime} \left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1} \left(\frac{\mathbf{F}_{x,-1}^{0\prime}\mathbf{F}_{x,-1}^0}{T}\right)^{-1} \mathbf{F}_{x,-1}^{0\prime} \bar{\mathbf{\Sigma}}_{kNT,-1} \mathbf{M}_{F_{x,-1}^0} \mathbf{u}_i + o_p(1). \tag{A.14}$$

Lemma 8 *Under Assumptions 1-5, as* $(N,T) \xrightarrow{j} \infty$ *such that* $N/T \to c$ *with* $0 < c < \infty$,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{V}_i^{\prime} \mathbf{M}_{\hat{F}_x} \mathbf{u}_i$$
$$= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{V}_i^{\prime} \mathbf{M}_{F_x^0} \mathbf{u}_i$$
$$- \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\mathbf{V}_i^{\prime}\mathbf{V}_j}{T} \mathbf{\Gamma}_{xj}^{0\prime} \left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1} \left(\frac{\mathbf{F}_x^{0\prime}\mathbf{F}_x^0}{T}\right)^{-1} \frac{\mathbf{F}_x^{0\prime}\mathbf{u}_i}{T} + o_p(1), \tag{A.15}$$

and

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{V}_{i,-1}^{\prime} \mathbf{M}_{\hat{F}_{x,-1}} \mathbf{u}_i$$
$$= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{V}_{i,-1}^{\prime} \mathbf{M}_{F_{x,-1}^0} \mathbf{u}_i$$
$$- \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\mathbf{V}_{i,-1}^{\prime}\mathbf{V}_{j,-1}}{T} \mathbf{\Gamma}_{xj}^{0\prime} \left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1} \left(\frac{\mathbf{F}_{x,-1}^{0\prime}\mathbf{F}_{x,-1}^0}{T}\right)^{-1} \frac{\mathbf{F}_{x,-1}^{0\prime}\mathbf{u}_i}{T} + o_p(1). \tag{A.16}$$

Lemma 9 *Under Assumptions 1-5, as* $(N,T) \xrightarrow{j} \infty$ *such that* $N/T \to c$ *with* $0 < c < \infty$,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{V}_i^{\prime} \mathbf{M}_{\hat{F}_x} \mathbf{M}_{\hat{F}_y} \mathbf{u}_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{V}_i^{\prime} \mathbf{M}_{F_x^0} \mathbf{M}_{F_y^0} \boldsymbol{\varepsilon}_i + o_p(1), \tag{A.17}$$

and

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{V}_{i,-1}^{\prime} \mathbf{M}_{\hat{F}_{x,-1}} \mathbf{M}_{\hat{F}_y} \mathbf{u}_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{V}_{i,-1}^{\prime} \mathbf{M}_{F_{x,-1}^0} \mathbf{M}_{F_y^0} \boldsymbol{\varepsilon}_i + o_p(1). \tag{A.18}$$

Proof of Proposition 1. Consider

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \hat{\mathbf{Z}}_i^{\prime} \mathbf{u}_i, \tag{A.19}$$

where $\hat{\mathbf{Z}}_i = \left[\mathbf{M}_{\hat{F}_x}\mathbf{X}_i, \mathbf{M}_{\hat{F}_{x,-1}}\mathbf{X}_{i,-1}\right]$. We begin with the first component of $\hat{\mathbf{Z}}_i$, which is $\mathbf{M}_{\hat{F}_x}\mathbf{X}_i$. Firstly, note that

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{X}_i^{\prime} \mathbf{M}_{\hat{F}_x} \mathbf{u}_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{\Gamma}_{xi}^{0\prime} \mathbf{F}_x^{0\prime} \mathbf{M}_{\hat{F}_x} \mathbf{u}_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{V}_i^{\prime} \mathbf{M}_{\hat{F}_x} \mathbf{u}_i. \tag{A.20}$$

By using the results of Lemmas 3, 4, 6 and 7, the first term in (A.20) is given by

$$
\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\mathbf{F}_{x}^{0\prime}\mathbf{M}_{\hat{F}_x}\mathbf{u}_i \\
&= -\frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\mathbf{\Gamma}_{xj}^{0}\mathbf{V}_j^{\prime}\mathbf{M}_{F_x^0}\mathbf{u}_i \\
&+\sqrt{\frac{T}{N}}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{n=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\mathbf{\Gamma}_{xn}^{0}\frac{\mathbf{V}_n^{\prime}\mathbf{V}_j}{T}\mathbf{\Gamma}_{xj}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\left(\frac{\mathbf{F}_x^{0\prime}\mathbf{F}_x^{0}}{T}\right)^{-1}\frac{\mathbf{F}_x^{0\prime}\mathbf{u}_i}{T} \\
&-\frac{1}{\sqrt{N}T^{3/2}}\sum_{i=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\mathbf{F}_x^{0\prime}\bar{\mathbf{\Sigma}}_{kNT}\mathbf{M}_{F_x^0}\mathbf{u}_i+o_p\left(1\right).
\end{aligned}
\tag{A.21}
$$

By making use of Lemma 8, the second term in (A.20) is given by

$$
\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i^{\prime}\mathbf{M}_{\hat{F}_x}\mathbf{u}_i \\
&=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i^{\prime}\mathbf{M}_{F_x^0}\mathbf{u}_i \\
&-\sqrt{\frac{T}{N}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\mathbf{V}_i^{\prime}\mathbf{V}_j}{T}\mathbf{\Gamma}_{xj}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\left(\frac{\mathbf{F}_x^{0\prime}\mathbf{F}_x^{0}}{T}\right)^{-1}\frac{\mathbf{F}_x^{0\prime}\mathbf{u}_i}{T}+o_p\left(1\right).
\end{aligned}
\tag{A.22}
$$

So, by adding (A.21) and (A.22) together, rearranging the terms and using $\mathbf{M}_{F_x^0}\mathbf{F}_x^0=\mathbf{0}$, we get

$$
\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i^{\prime}\mathbf{M}_{\hat{F}_x}\mathbf{u}_i \\
&=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i^{\prime}\mathbf{M}_{F_x^0}\mathbf{u}_i \\
&-\frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\mathbf{\Gamma}_{xj}^{0}\mathbf{X}_j^{\prime}\mathbf{M}_{F_x^0}\mathbf{u}_i \\
&-\sqrt{\frac{T}{N}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\tilde{\mathbf{V}}_i^{\prime}\mathbf{V}_j}{T}\mathbf{\Gamma}_{xj}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\left(\frac{\mathbf{F}_x^{0\prime}\mathbf{F}_x^{0}}{T}\right)^{-1}\frac{\mathbf{F}_x^{0\prime}\mathbf{u}_i}{T} \\
&-\sqrt{\frac{N}{T}}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\left(\mathbf{\Upsilon}_{xkN}^{0}\right)^{-1}\left(\frac{\mathbf{F}_x^{0\prime}\mathbf{F}_x^{0}}{T}\right)^{-1}\mathbf{F}_x^{0\prime}\bar{\mathbf{\Sigma}}_{kNT}\mathbf{M}_{F_x^0}\mathbf{u}_i+o_p\left(1\right) \\
&=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\tilde{\mathbf{X}}_i^{\prime}\mathbf{M}_{F_x^0}\mathbf{u}_i+\sqrt{\frac{T}{N}}\mathbf{b}_{11NT}+\sqrt{\frac{N}{T}}\mathbf{b}_{21NT}+o_p\left(1\right),
\end{aligned}
\tag{A.23}
$$

where

$$
\tilde{\mathbf{X}}_i=\mathbf{X}_i-\frac{1}{N}\sum_{n=1}^{N}\mathbf{X}_n\mathbf{\Gamma}_{xn}^{0\prime}(\mathbf{\Upsilon}_{xkN}^{0})^{-1}\mathbf{\Gamma}_{xi}^{0},
$$

$$
\tilde{\mathbf{V}}_i=\mathbf{V}_i-\frac{1}{N}\sum_{n=1}^{N}\mathbf{V}_n\mathbf{\Gamma}_{xn}^{0\prime}(\mathbf{\Upsilon}_{xkN}^{0})^{-1}\mathbf{\Gamma}_{xi}^{0},
$$

$$
\mathbf{b}_{11NT}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\tilde{\mathbf{V}}_i^{\prime}\mathbf{V}_j}{T}\mathbf{\Gamma}_{xj}^{0\prime}(\mathbf{\Upsilon}_{xkN}^{0})^{-1}\left(\frac{\mathbf{F}_x^{0\prime}\mathbf{F}_x^{0}}{T}\right)^{-1}\frac{\mathbf{F}_x^{0\prime}\mathbf{u}_i}{T},
$$

$$
\mathbf{b}_{21NT}=-\frac{1}{NT}\sum_{i=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}(\mathbf{\Upsilon}_{xkN}^{0})^{-1}\left(\frac{\mathbf{F}_x^{0\prime}\mathbf{F}_x^{0}}{T}\right)^{-1}\mathbf{F}_x^{0\prime}\bar{\mathbf{\Sigma}}_{kNT}\mathbf{M}_{F_x^0}\mathbf{u}_i.
$$

As for the second component of $\hat{\mathbf{Z}}_i$, which is $\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{X}_{i,-1}$, by following the same steps as before and using again Lemmas 3, 4, 6, 7, 8, and using $\mathbf{M}_{F^0_{x,-1}}\mathbf{F}^0_{x,-1} = \mathbf{0}$, we obtain

$$\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}'_{i,-1}\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{u}_i \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}'_{i,-1}\mathbf{M}_{F^0_{x,-1}}\mathbf{u}_i \\
&- \frac{1}{\sqrt{NT}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\Gamma}^{0\prime}_{xi}\left(\boldsymbol{\Upsilon}^0_{xkN}\right)^{-1}\boldsymbol{\Gamma}^0_{xj}\mathbf{X}'_{j,-1}\mathbf{M}_{F^0_{x,-1}}\mathbf{u}_i \\
&- \sqrt{\frac{T}{N}}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\tilde{\mathbf{V}}'_{i,-1}\mathbf{V}_{j,-1}}{T}\boldsymbol{\Gamma}^{0\prime}_{xj}\left(\boldsymbol{\Upsilon}^0_{xkN}\right)^{-1}\left(\frac{\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}^0_{x,-1}}{T}\right)^{-1}\frac{\mathbf{F}^{0\prime}_{x,-1}\mathbf{u}_i}{T} \\
&- \frac{1}{\sqrt{N}T^{3/2}}\sum_{i=1}^{N}\boldsymbol{\Gamma}^{0\prime}_{xi}\left(\boldsymbol{\Upsilon}^0_{xkN}\right)^{-1}\left(\frac{\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}^0_{x,-1}}{T}\right)^{-1}\mathbf{F}^{0\prime}_{x,-1}\bar{\boldsymbol{\Sigma}}_{kNT,-1}\mathbf{M}_{F^0_{x,-1}}\mathbf{u}_i + o_p(1) \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\tilde{\mathbf{X}}'_{i,-1}\mathbf{M}_{F^0_{x,-1}}\mathbf{u}_i + \sqrt{\frac{T}{N}}\mathbf{b}_{12NT} + \sqrt{\frac{N}{T}}\mathbf{b}_{22NT} + o_p(1),
\end{aligned} \tag{A.24}$$

where

$$\begin{aligned}
\tilde{\mathbf{X}}_{i,-1} &= \mathbf{X}_{i,-1} - \frac{1}{N}\sum_{n=1}^{N}\mathbf{X}_{n,-1}\boldsymbol{\Gamma}^{0\prime}_{xn}(\boldsymbol{\Upsilon}^0_{xkN})^{-1}\boldsymbol{\Gamma}^0_{xi}, \\
\tilde{\mathbf{V}}_{i,-1} &= \mathbf{V}_{i,-1} - \frac{1}{N}\sum_{n=1}^{N}\mathbf{V}_{n,-1}\boldsymbol{\Gamma}^{0\prime}_{xn}(\boldsymbol{\Upsilon}^0_{xkN})^{-1}\boldsymbol{\Gamma}^0_{xi}, \\
\mathbf{b}_{12NT} &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\tilde{\mathbf{V}}'_{i,-1}\mathbf{V}_{j,-1}}{T}\boldsymbol{\Gamma}^{0\prime}_{xj}(\boldsymbol{\Upsilon}^0_{xkN})^{-1}\left(\frac{\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}^0_{x,-1}}{T}\right)^{-1}\frac{\mathbf{F}^{0\prime}_{x,-1}\mathbf{u}_i}{T}, \\
\mathbf{b}_{22NT} &= -\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Gamma}^{0\prime}_{xi}(\boldsymbol{\Upsilon}^0_{xkN})^{-1}\left(\frac{\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}^0_{x,-1}}{T}\right)^{-1}\mathbf{F}^{0\prime}_{x,-1}\bar{\boldsymbol{\Sigma}}_{kNT,-1}\mathbf{M}_{F^0_{x,-1}}\mathbf{u}_i.
\end{aligned}$$

Hence, we have

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\hat{\mathbf{Z}}'_i\mathbf{u}_i &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[\mathbf{M}_{\hat{F}_x}\mathbf{X}_i, \mathbf{M}_{\hat{F}_{x,-1}}\mathbf{X}_{i,-1}\right]'\mathbf{u}_i \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left[\mathbf{M}_{F^0_x}\tilde{\mathbf{X}}_i, \mathbf{M}_{F^0_{x,-1}}\tilde{\mathbf{X}}_{i,-1}\right]'\mathbf{u}_i + \sqrt{\frac{T}{N}}[\mathbf{b}'_{11NT}, \mathbf{b}'_{12NT}]' \\
&+ \sqrt{\frac{N}{T}}[\mathbf{b}'_{21NT}, \mathbf{b}'_{22NT}]' + o_p(1) \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\tilde{\mathbf{Z}}'_i\mathbf{u}_i + \sqrt{\frac{T}{N}}\mathbf{b}_{1NT} + \sqrt{\frac{N}{T}}\mathbf{b}_{2NT} + o_p(1),
\end{aligned}$$

where $\tilde{\mathbf{Z}}_i = \left[\mathbf{M}_{F^0_x}\tilde{\mathbf{X}}_i, \mathbf{M}_{F^0_{x,-1}}\tilde{\mathbf{X}}_{i,-1}\right]$, $\mathbf{b}_{1NT} = [\mathbf{b}'_{11NT}, \mathbf{b}'_{12NT}]'$ and $\mathbf{b}_{2NT} = [\mathbf{b}'_{21NT}, \mathbf{b}'_{22NT}]'$, which provides the expression given in Proposition 1. ■

Proof of Proposition 2. Now consider

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\hat{\mathbf{Z}}'_i\mathbf{M}_{\hat{F}_y}\mathbf{u}_i, \tag{A.25}$$

where $\hat{\mathbf{Z}}_i = \left[\mathbf{M}_{\hat{F}_x}\mathbf{X}_i, \mathbf{M}_{\hat{F}_{x,-1}}\mathbf{X}_{i,-1}\right]$. We start with the first component of $\hat{\mathbf{Z}}_i$, i.e. $\mathbf{M}_{\hat{F}_x}\mathbf{X}_i$, which can be written as

$$
\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_{\hat{F}_x}\mathbf{M}_{\hat{F}_y}\mathbf{u}_i \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{\Gamma}_{xi}^{0\prime}\mathbf{F}_x^{0\prime}\mathbf{M}_{\hat{F}_x}\mathbf{M}_{\hat{F}_y}\mathbf{u}_i + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{M}_{\hat{F}_x}\mathbf{M}_{\hat{F}_y}\mathbf{u}_i \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{M}_{\hat{F}_x}\mathbf{M}_{\hat{F}_y}\mathbf{u}_i + o_p(1) \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_i'\mathbf{M}_{F_x^0}\mathbf{M}_{F_y^0}\boldsymbol{\varepsilon}_i + o_p(1), \qquad \text{(A.26)}
\end{aligned}
$$

where the second and third equalities is due to Lemma 5 and 9, respectively.

As for the second component of $\hat{\mathbf{Z}}_i$, which is $\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{X}_{i,-1}$, by following the same steps as before and using again Lemmas 5 and 9, we yield

$$
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i,-1}'\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{M}_{\hat{F}_y}\mathbf{u}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{V}_{i,-1}'\mathbf{M}_{F_{x,-1}^0}\mathbf{M}_{F_y^0}\boldsymbol{\varepsilon}_i + o_p(1). \qquad \text{(A.27)}
$$

By combining the results above, we obtain the required expression. ■

Lemma 10 *Under Assumptions 1-5, as $(N,T)\xrightarrow{j}\infty$ such that $N/T\to c$ with $0<c<\infty$,*

$$
\begin{aligned}
T^{-1/2}\mathbf{X}_i'\left(\mathbf{M}_{\hat{F}_x}-\mathbf{M}_{F_x^0}\right)\mathbf{u}_i &= \sqrt{T}O_p\left(\delta_{NT}^{-2}\right), &\text{(A.28)}\\
T^{-1/2}\mathbf{X}_i'\mathbf{M}_{\hat{F}_{x,-1}}\left(\mathbf{M}_{\hat{F}_x}-\mathbf{M}_{F_x^0}\right)\mathbf{u}_i &= \sqrt{T}O_p\left(\delta_{NT}^{-2}\right), &\text{(A.29)}\\
T^{-1/2}\mathbf{X}_i'\left(\mathbf{M}_{\hat{F}_{x,-1}}-\mathbf{M}_{F_{x,-1}^0}\right)\mathbf{M}_{F_x^0}\mathbf{u}_i &= \sqrt{T}O_p\left(\delta_{NT}^{-2}\right). &\text{(A.30)}
\end{aligned}
$$

Proof of Proposition 3. Consider $T^{-1/2}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_x}\mathbf{u}_i$ where $\hat{\mathbf{Z}}_i=\left[\mathbf{M}_{\hat{F}_x}\mathbf{X}_i,\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{X}_{i,-1}\right]$. Let us start with the first component of $\mathbf{M}_{\hat{F}_x}\hat{\mathbf{Z}}_i$, i.e. $\mathbf{M}_{\hat{F}_x}\mathbf{X}_i$. By adding and subtracting we get

$$
\begin{aligned}
T^{-1/2}\mathbf{X}_i'\mathbf{M}_{\hat{F}_x}\mathbf{u}_i &= T^{-1/2}\mathbf{X}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i + T^{-1/2}\mathbf{X}_i'\left(\mathbf{M}_{\hat{F}_x}-\mathbf{M}_{F_x^0}\right)\mathbf{u}_i\\
&= T^{-1/2}\mathbf{X}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i + \sqrt{T}O_p\left(\delta_{NT}^{-2}\right) \qquad \text{(A.31)}
\end{aligned}
$$

where the second equality is due to result in (A.28) stated in Lemma 10.

Next is the second component of $\mathbf{M}_{\hat{F}_x}\hat{\mathbf{Z}}_i$, which is $\mathbf{M}_{\hat{F}_x}\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{X}_{i,-1}$. Again, by adding and subtracting and using Lemma 10, we get

$$
\begin{aligned}
&T^{-1/2}\mathbf{X}_{i,-1}'\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{M}_{\hat{F}_x}\mathbf{u}_i\\
&= T^{-1/2}\mathbf{X}_{i,-1}'\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{M}_{F_x^0}\mathbf{u}_i + T^{-1/2}\mathbf{X}_i'\mathbf{M}_{\hat{F}_{x,-1}}\left(\mathbf{M}_{\hat{F}_x}-\mathbf{M}_{F_x^0}\right)\mathbf{u}_i\\
&= T^{-1/2}\mathbf{X}_{i,-1}'\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{M}_{F_x^0}\mathbf{u}_i + \sqrt{T}O_p\left(\delta_{NT}^{-2}\right)\\
&= T^{-1/2}\mathbf{X}_{i,-1}'\mathbf{M}_{F_{x,-1}^0}\mathbf{M}_{F_x^0}\mathbf{u}_i + T^{-1/2}\mathbf{X}_{i,-1}'\left(\mathbf{M}_{\hat{F}_{x,-1}}-\mathbf{M}_{F_{x,-1}^0}\right)\mathbf{M}_{F_x^0}\mathbf{u}_i + \sqrt{T}O_p\left(\delta_{NT}^{-2}\right)\\
&= T^{-1/2}\mathbf{X}_{i,-1}'\mathbf{M}_{F_{x,-1}^0}\mathbf{M}_{F_x^0}\mathbf{u}_i + \sqrt{T}O_p\left(\delta_{NT}^{-2}\right). \qquad \text{(A.32)}
\end{aligned}
$$

Finally, by combining the results, we get

$$
T^{-1/2}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_x}\mathbf{u}_i = T^{-1/2}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i + \sqrt{T}O_p\left(\delta_{NT}^{-2}\right), \qquad \text{(A.33)}
$$

where $\mathbf{Z}_i=\left[\mathbf{M}_{F_x^0}\mathbf{X}_i,\mathbf{M}_{F_{x,-1}^0}\mathbf{X}_{i,-1}\right]$, which provides the expression given in Proposition 3. ■

Lemma 11 *Under Assumptions 1-5, as $(N,T)\xrightarrow{j}\infty$ such that $N/T\to c$ with $0<c<\infty$, $\frac{1}{NT}\sum_{i=1}^{N}\hat{\boldsymbol{\xi}}_{\hat{F}iT}\hat{\boldsymbol{\xi}}_{\hat{F}iT}' = \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\xi}_{\hat{F}iT}\boldsymbol{\xi}_{\hat{F}iT}' + o_p(1)$, where $\boldsymbol{\xi}_{\hat{F}iT}=\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y}\mathbf{u}_i$ and $\hat{\boldsymbol{\xi}}_{\hat{F}iT}=\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y}\hat{\mathbf{u}}_i$.*

Lemma 12 *Under Assumptions 1-5, as* $(N,T) \xrightarrow{j} \infty$ *such that* $N/T \to c$ *with* $0 < c < \infty$, $\frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{\xi}_{\hat{F}iT}\boldsymbol{\xi}_{\hat{F}iT}' - \boldsymbol{\Omega} = o_p(1)$, *where* $\boldsymbol{\Omega} = \text{plim}_{N,T\to\infty} \frac{1}{N}\sum_{i=1}^{N} E\left(T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_y^0}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{F_y^0}\mathbf{Z}_i\right)$.

Proposition 4 *Under Assumptions 1-5, as* $(N,T) \xrightarrow{j} \infty$ *such that* $N/T \to c$ *with* $0 < c < \infty$,

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \boldsymbol{\xi}_{\hat{F}iT} \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Omega}).$$

Proof. Proposition 2 and Lemma 12, together with Lemma **??**, yield the required result. ∎

Lemma 13 *Under Assumptions 1-5, as* $(N,T) \xrightarrow{j} \infty$ *such that* $N/T \to c$ *with* $0 < c < \infty$, $\hat{\mathbf{A}}_{NT} \xrightarrow{p} \mathbf{A}$, $\mathbf{B}_{NT} \xrightarrow{p} \mathbf{B}$, *where* $\hat{\mathbf{A}}_{NT} = \frac{1}{N}\sum_{i=1}^{N} T^{-1}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y}\mathbf{W}_i$, $\hat{\mathbf{B}}_{NT} = \frac{1}{N}\sum_{i=1}^{N} T^{-1}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y}\hat{\mathbf{Z}}_i$ *and* $\mathbf{A} = \lim_{N,T\to\infty} \frac{1}{N}\sum_{i=1}^{N} E(\mathbf{A}_{i,T})$, $\mathbf{B} = \lim_{N,T\to\infty} \frac{1}{N}\sum_{i=1}^{N} E(\mathbf{B}_{i,T})$, $\mathbf{A}_{i,T} = T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_y^0}\mathbf{W}_i$, $\mathbf{B}_{i,T} = T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_y^0}\mathbf{Z}_i$.

Lemma 14 (Lemma 2.2.10 of Van der Vaart and Wellner (1996)) *Let* $x_1, \cdots, x_N$ *be arbitrary random variables that satisfy the tail bound:*

$$P(|x_i| > z) \le 2\text{exp}\left(-\frac{1}{2} \times \frac{z^2}{a + bz}\right)$$

for all z *(and all* i*) and fixed* $a, b > 0$*. Then,*

$$E\left|\sup_{1\le i\le N} x_i\right| \le \Delta\left(b \times \ln(N+1) + \sqrt{a \times \ln(N+1)}\right)$$

for some positive constant Δ.

Lemma 15 *Under Assumptions 2 to 4, and Assumption 7, we have*

$$(a)\quad N^{-1}T^{-1}\sum_{i=1}^{N} \|\mathbf{X}_i'\mathbf{M}_{\hat{F}_x}\mathbf{u}_i - \mathbf{X}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i\| = O_p(\delta_{NT}^{-2}).$$

$$(b)\quad N^{-1}T^{-1}\sum_{i=1}^{N} \|\mathbf{X}_{i,-1}'\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{M}_{\hat{F}_x}\mathbf{u}_i - \mathbf{X}_{i,-1}'\mathbf{M}_{F_{x,-1}^0}\mathbf{M}_{F_x^0}\mathbf{u}_i\| = O_p(\delta_{NT}^{-2}).$$

$$(c)\quad \sup_{1\le i,j\le N} \|T^{-1}\mathbf{X}_j'\mathbf{M}_{\hat{F}_x}\mathbf{X}_i - T^{-1}\mathbf{X}_j'\mathbf{M}_{F_x^0}\mathbf{X}_i\| = O_p(N^{1/2}\delta_{NT}^{-2}).$$

$$(d)\quad \sup_{1\le i,j\le N} \|T^{-1}\mathbf{X}_{j,-1}'\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{M}_{\hat{F}_x}\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{X}_{i,-1} - T^{-1}\mathbf{X}_{j,-1}'\mathbf{M}_{F_{x,-1}^0}\mathbf{M}_{F_x^0}\mathbf{M}_{F_{x,-1}^0}\mathbf{X}_{i,-1}\| = O_p(N^{1/2}\delta_{NT}^{-2}).$$

$$(e)\quad \sup_{1\le i,j\le N} \|T^{-1}\mathbf{X}_{j,-1}'\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{M}_{\hat{F}_x}\mathbf{X}_i - T^{-1}\mathbf{X}_{j,-1}'\mathbf{M}_{F_{x,-1}^0}\mathbf{M}_{F_x^0}\mathbf{X}_i\| = O_p(N^{1/2}\delta_{NT}^{-2}).$$

Lemma 16 *Under Assumptions 1 to 7, we have*

$$\begin{aligned}
(a)\quad & \sup_{1\le i,j\le N} \|T^{-1}\mathbf{X}_i'\mathbf{M}_{\hat{F}_x}\mathbf{y}_{i,-1} - T^{-1}\mathbf{X}_i'\mathbf{M}_{F_x^0}\mathbf{y}_{i,-1}\| \\
& = O_p(N^{1/2}\delta_{NT}^{-2}) + O_p(N^{3/4}T^{-1/2}\delta_{NT}^{-2}) + O_p(NT^{-1}\delta_{NT}^{-2}) + O_p(N^{1/4}T^{-1/2}). \\
(b)\quad & \sup_{1\le i,j\le N} \|T^{-1}\mathbf{X}_{i,-1}'\mathbf{M}_{\hat{F}_{x,-1}}\mathbf{M}_{\hat{F}_x}\mathbf{y}_{i,-1} - T^{-1}\mathbf{X}_{i,-1}'\mathbf{M}_{F_{x,-1}^0}\mathbf{M}_{F_x^0}\mathbf{y}_{i,-1}\| \\
& = O_p(N^{1/2}\delta_{NT}^{-2}) + O_p(N^{3/4}T^{-1/2}\delta_{NT}^{-2}) + O_p(NT^{-1}\delta_{NT}^{-2}) + O_p(N^{1/4}T^{-1/2}).
\end{aligned}$$

Lemma 17 *Under Assumptions 1 to 7, we have*

$$(a)\quad \sup_{1\le i\le N}\|T^{-1}\mathbf{X}_i'\mathbf{M}_{F_x^0}\mathbf{X}_i - T^{-1}E(\mathbf{V}_i'\mathbf{V}_i)\| = O_p(N^{1/4}T^{-1/2}) + O_p(N^{1/2}T^{-1}),$$

$$(b)\quad \sup_{1\le i\le N}\|T^{-1}\mathbf{X}_{i,-1}'\mathbf{M}_{F_{x,-1}^0}\mathbf{M}_{F_x^0}\mathbf{M}_{F_{x,-1}^0}\mathbf{X}_{i,-1} - T^{-1}E(\mathbf{V}_{i,-1}'\mathbf{V}_{i,-1})\| = O_p(N^{1/4}T^{-1/2}) + O_p(N^{1/2}T^{-1}),$$

$$(c)\quad \sup_{1\le i\le N}\|T^{-1}\mathbf{X}_{i,-1}'\mathbf{M}_{F_{x,-1}^0}\mathbf{M}_{F_x^0}\mathbf{X}_i - T^{-1}E(\mathbf{V}_{i,-1}'\mathbf{V}_i)\| = O_p(N^{1/4}T^{-1/2}) + O_p(N^{1/2}T^{-1}),$$

$$(d)\quad \sup_{1\le i\le N}\|T^{-1}\mathbf{X}_i'\mathbf{M}_{F_x^0}\mathbf{y}_{i,-1} - T^{-1}\sum_{s=1}^{\infty}E(\mathbf{V}_i'\mathbf{V}_{i,-s})\boldsymbol{\beta}_i\rho_i^{s-1}\| = O_p(N^{1/4}T^{-1/2}) + O_p(N^{1/2}T^{-1}).$$

$$(e)\quad \sup_{1\le i\le N}\|T^{-1}\mathbf{X}_{i,-1}'\mathbf{M}_{F_{x,-1}^0}\mathbf{M}_{F_x^0}\mathbf{y}_{i,-1} - T^{-1}\sum_{s=1}^{\infty}E(\mathbf{V}_{i,-1}'\mathbf{V}_{i,-s})\boldsymbol{\beta}_i\rho_i^{s-1}\| = O_p(N^{1/4}T^{-1/2}) + O_p(N^{1/2}T^{-1}).$$

Lemma 18 *Define*

$$\mathbf{A}_{i,T} = \begin{pmatrix} T^{-1}\sum_{s=1}^{\infty}\rho_i^{s-1}E(\mathbf{V}_i'\mathbf{V}_{i,-s})\boldsymbol{\beta}_i & T^{-1}E(\mathbf{V}_i'\mathbf{V}_i) \\ T^{-1}\sum_{s=1}^{\infty}\rho_i^{s-1}E(\mathbf{V}_{i,-1}'\mathbf{V}_{i,-s})\boldsymbol{\beta}_i & T^{-1}E(\mathbf{V}_{i,-1}'\mathbf{V}_i)\end{pmatrix}$$
$$\mathbf{B}_{i,T} = \begin{pmatrix} T^{-1}E(\mathbf{V}_i'\mathbf{V}_i) & T^{-1}E(\mathbf{V}_i'\mathbf{V}_{i,-1}) \\ T^{-1}E(\mathbf{V}_{i,-1}'\mathbf{V}_i) & T^{-1}E(\mathbf{V}_{i,-1}'\mathbf{V}_{i,-1})\end{pmatrix}$$

under Assumptions 1 to 7, we have

$$(a)\quad \sup_{1\le i\le N}\|(\hat{\tilde{\mathbf{A}}}_{i,T}'\hat{\tilde{\mathbf{B}}}_{i,T}^{-1}\hat{\tilde{\mathbf{A}}}_{i,T})^{-1}\hat{\tilde{\mathbf{A}}}_{i,T}'\hat{\tilde{\mathbf{B}}}_{i,T}^{-1} - (\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\|$$
$$= O_p(N^{1/4}T^{-1/2}\ln N) + O_p(N^{1/2}(\ln N)^5\delta_{NT}^{-2}),$$

$$(b)\quad \sup_{1\le i\le N}\|(\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\| = O_p((\ln N)^2),$$

$$(c)\quad \sup_{1\le i\le N}\|[(\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1} - (\mathbf{A}_{i,T}'\mathbf{B}_{i,T}^{-1}\mathbf{A}_{i,T})^{-1}\mathbf{A}_{i,T}'\mathbf{B}_{i,T}^{-1}]\|$$
$$= O_p(N^{1/4}T^{-1/2}(\ln N)^5) + O_p(N^{1/2}T^{-1}(\ln N)^5).$$

Proof of Theorem 1. By using the expression in (16), the result of Proposition 1 from which $\frac{1}{\sqrt{NT}}\sum_{i=1}^N\hat{\mathbf{Z}}_i'\mathbf{u}_i$ tends to a multivatiate random variable and is therefore $O_p(1)$, and $\sqrt{\frac{T}{N}}\mathbf{b}_{1NT}$ together with $\sqrt{\frac{N}{T}}\mathbf{b}_{2NT}$ are $O_p(1)$ as T/N tends to a finite positive constant c $(0<c<\infty)$ when N and $T\to\infty$ jointly. And so, $\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{IV}-\boldsymbol{\theta}\right) = O_p(1)$, which implies the required result. ■

Proof of Theorem 2. (i) $\sqrt{NT}\left(\hat{\hat{\boldsymbol{\theta}}}_{IV}-\boldsymbol{\theta}\right) = \left(\hat{\mathbf{A}}_{NT}'\hat{\mathbf{B}}_{NT}^{-1}\hat{\mathbf{A}}_{NT}\right)^{-1}\hat{\mathbf{A}}_{NT}'\hat{\mathbf{B}}_{NT}^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^N\boldsymbol{\xi}_{\hat{F}iT}\right)$ $= \left(\mathbf{A}'\mathbf{B}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}'\mathbf{B}^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^N\boldsymbol{\xi}_{FiT}\right) + o_p(1)$, by the results of Proposition 2 and Lemma 13. Next, by the result of Proposition 4, we have $\sqrt{NT}\left(\hat{\hat{\boldsymbol{\theta}}}_{IV}-\boldsymbol{\theta}\right) \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Psi})$, as required. (ii) $\hat{\boldsymbol{\Psi}}-\boldsymbol{\Psi} = o_p(1)$ follows immediately from Lemmas 11, 12 and 13. ■

Proof of Theorem 3. Under Assumptions 1-5, noting $\hat{\mathbf{u}}_i = \mathbf{u}_i - \mathbf{W}_i\left(\hat{\hat{\boldsymbol{\theta}}}_{IV2}-\boldsymbol{\theta}\right)$ we have $\frac{1}{\sqrt{NT}}\sum_{i=1}^N\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y}\hat{\mathbf{u}}_i =$ $\frac{1}{\sqrt{NT}}\sum_{i=1}^N\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y}\mathbf{u}_i - \hat{\mathbf{A}}_{NT}\sqrt{NT}\left(\hat{\hat{\boldsymbol{\theta}}}_{IV2}-\boldsymbol{\theta}\right)$. Since $\sqrt{NT}\left(\hat{\hat{\boldsymbol{\theta}}}_{IV2}-\boldsymbol{\theta}\right) = \left(\mathbf{A}'\boldsymbol{\Omega}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}'\boldsymbol{\Omega}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\mathbf{Z}_i'\mathbf{M}_{F_y^0}\boldsymbol{\varepsilon}_i$ $+\, o_p(1)$ by Corollary 1 and defining $\mathbf{L} = \boldsymbol{\Omega}^{-1/2}\mathbf{A}$ of rank$(k+1)$ we have $\hat{\boldsymbol{\Omega}}_{NT}^{-1/2}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y}\hat{\mathbf{u}}_i =$ $\mathbf{M}_L\boldsymbol{\Omega}^{-1/2}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\mathbf{Z}_i'\mathbf{M}_{F_y^0}\boldsymbol{\varepsilon}_i + o_p(1)$ with $\mathbf{M}_L = \mathbf{I}_{2k} - \mathbf{L}\left(\mathbf{L}'\mathbf{L}\right)^{-1}\mathbf{L}'$ whose rank is $k-1$, which yields $\frac{1}{NT}\sum_{i=1}^N\hat{\mathbf{u}}_i'\mathbf{M}_{\hat{F}_y}\hat{\mathbf{Z}}_i\hat{\boldsymbol{\Omega}}_{NT}^{-1}\sum_{i=1}^N\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_y}\hat{\mathbf{u}}_i \xrightarrow{d} \chi_{k-1}^2$ as required. ■

Proof of Theorem 4. By Proposition 3 $T^{-1/2}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_x}\mathbf{u}_i = T^{-1/2}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i + o_p(1)$ as $(N,T)\xrightarrow{j}\infty$ as $N/T\to c$ for $0<c<\infty$. It is immediate that, under Assumptions 1-8, for each i, $T^{-1/2}\mathbf{Z}_i'\mathbf{u}_i \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Sigma}_i)$. A similar line of the argument in the proof of Lemma 7 ensures that $\hat{\tilde{\mathbf{A}}}_{i,T}-\tilde{\mathbf{A}}_{i,T}\xrightarrow{p}\mathbf{0}$ and $\hat{\tilde{\mathbf{B}}}_{i,T}-\tilde{\mathbf{B}}_{i,T}\xrightarrow{p}\mathbf{0}$ as $T\to\infty$, and togther with Assumption 8 we see that $p\lim_{T\to\infty}\hat{\tilde{\mathbf{A}}}_{i,T} = \mathbf{A}_i$ and $p\lim_{T\to\infty}\hat{\tilde{\mathbf{B}}}_{i,T} = \mathbf{B}_i$, thus the required result follows. ■

Proof of Theorem 5. Note that the instrumental variable (IV) or two-stage least squares estimator of $\boldsymbol{\theta}_i$ is $\hat{\boldsymbol{\theta}}_{IV,i} = (\hat{\tilde{\mathbf{A}}}'_{i,T}\hat{\tilde{\mathbf{B}}}^{-1}_{i,T}\hat{\tilde{\mathbf{A}}}_{i,T})^{-1}\hat{\tilde{\mathbf{A}}}'_{i,T}\hat{\tilde{\mathbf{B}}}^{-1}_{i,T}\hat{\mathbf{g}}_{i,T}$, then we have

$$\widehat{\boldsymbol{\theta}}_{IVMG} - \boldsymbol{\theta} = N^{-1}\sum_{i=1}^{N}(\widehat{\boldsymbol{\theta}}_{IV,i} - \boldsymbol{\theta}) = N^{-1}\sum_{i=1}^{N}(\widehat{\boldsymbol{\theta}}_{IV,i} - \boldsymbol{\theta}_i) + N^{-1}\sum_{i=1}^{N}\boldsymbol{\eta}_i$$

where the first term is

$$\begin{aligned}
N^{-1}\sum_{i=1}^{N}(\widehat{\boldsymbol{\theta}}_{IV,i} - \boldsymbol{\theta}_i) =& N^{-1}\sum_{i=1}^{N}(\hat{\tilde{\mathbf{A}}}'_{i,T}\hat{\tilde{\mathbf{B}}}^{-1}_{i,T}\hat{\tilde{\mathbf{A}}}_{i,T})^{-1}\hat{\tilde{\mathbf{A}}}'_{i,T}\hat{\tilde{\mathbf{B}}}^{-1}_{i,T}(T^{-1}\hat{\mathbf{Z}}'_i\mathbf{M}_{\hat{F}_x}\mathbf{u}_i)\\
=&N^{-1}\sum_{i=1}^{N}(\tilde{\mathbf{A}}'_{i,T}\tilde{\mathbf{B}}^{-1}_{i,T}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}'_{i,T}\tilde{\mathbf{B}}^{-1}_{i,T}(T^{-1}\mathbf{Z}'_i\mathbf{M}_{F^0_x}\mathbf{u}_i)\\
&+ N^{-1}\sum_{i=1}^{N}\Big[(\hat{\tilde{\mathbf{A}}}'_{i,T}\hat{\tilde{\mathbf{B}}}^{-1}_{i,T}\hat{\tilde{\mathbf{A}}}_{i,T})^{-1}\hat{\tilde{\mathbf{A}}}'_{i,T}\hat{\tilde{\mathbf{B}}}^{-1}_{i,T} - (\tilde{\mathbf{A}}'_{i,T}\tilde{\mathbf{B}}^{-1}_{i,T}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}'_{i,T}\tilde{\mathbf{B}}^{-1}_{i,T}\Big](T^{-1}\mathbf{Z}'_i\mathbf{M}_{F^0_x}\mathbf{u}_i)\\
&+ N^{-1}\sum_{i=1}^{N}\Big[(\hat{\tilde{\mathbf{A}}}'_{i,T}\hat{\tilde{\mathbf{B}}}^{-1}_{i,T}\hat{\tilde{\mathbf{A}}}_{i,T})^{-1}\hat{\tilde{\mathbf{A}}}'_{i,T}\hat{\tilde{\mathbf{B}}}^{-1}_{i,T} - (\tilde{\mathbf{A}}'_{i,T}\tilde{\mathbf{B}}^{-1}_{i,T}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}'_{i,T}\tilde{\mathbf{B}}^{-1}_{i,T}\Big](T^{-1}\hat{\mathbf{Z}}'_i\mathbf{M}_{\hat{F}_x}\mathbf{u}_i - T^{-1}\mathbf{Z}'_i\mathbf{M}_{F^0_x}\mathbf{u}_i)\\
&+ N^{-1}\sum_{i=1}^{N}(\tilde{\mathbf{A}}'_{i,T}\tilde{\mathbf{B}}^{-1}_{i,T}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}'_{i,T}\tilde{\mathbf{B}}^{-1}_{i,T}(T^{-1}\hat{\mathbf{Z}}'_i\mathbf{M}_{\hat{F}_x}\mathbf{u}_i - T^{-1}\mathbf{Z}'_i\mathbf{M}_{F^0_x}\mathbf{u}_i)\\
=&\mathbb{G}_1 + \mathbb{G}_2 + \mathbb{G}_3 + \mathbb{G}_4
\end{aligned}$$

We first consider the terms $\mathbb{G}_2$, $\mathbb{G}_3$, and $\mathbb{G}_4$. With Lemma 15 (a)-(b), we have

$$N^{-1}T^{-1}\sum_{i=1}^{N}\|T^{-1}\hat{\mathbf{Z}}'_i\mathbf{M}_{\hat{F}_x}\mathbf{u}_i - T^{-1}\mathbf{Z}'_i\mathbf{M}_{F^0_x}\mathbf{u}_i\| = O_p(\delta_{NT}^{-2}) \tag{A.34}$$

With the above facts, $\mathbb{G}_2$ is bounded in norm by

$$\begin{aligned}
&N^{-1}\sum_{i=1}^{N}\|T^{-1}\mathbf{Z}'_i\mathbf{M}_{F^0_x}\mathbf{u}_i\| \cdot \sup_{1\le i\le N}\|(\hat{\tilde{\mathbf{A}}}'_{i,T}\hat{\tilde{\mathbf{B}}}^{-1}_{i,T}\hat{\tilde{\mathbf{A}}}_{i,T})^{-1}\hat{\tilde{\mathbf{A}}}'_{i,T}\hat{\tilde{\mathbf{B}}}^{-1}_{i,T} - (\tilde{\mathbf{A}}'_{i,T}\tilde{\mathbf{B}}^{-1}_{i,T}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}'_{i,T}\tilde{\mathbf{B}}^{-1}_{i,T}\|\\
=&O_p(N^{1/4}T^{-1}\ln N) + O_p(N^{1/2}T^{-1/2}(\ln N)^5\delta_{NT}^{-2})
\end{aligned}$$

Analogously, with (A.34), we can show that $\mathbb{G}_3 = O_p(N^{1/4}T^{-1/2}\ln N\delta_{NT}^{-2}) + O_p(N^{1/2}(\ln N)^5\delta_{NT}^{-4})$ and $\mathbb{G}_4 = O_p((\ln N)^2\delta_{NT}^{-2})$. Consider $\mathbb{G}_1$. Define

$$\begin{aligned}
\mathbb{H}_{1i} &= \begin{pmatrix}\mathbf{V}'_i\mathbf{F}_y\boldsymbol{\gamma}_{yi}\\ \mathbf{V}'_{i,-1}\mathbf{F}_y\boldsymbol{\gamma}_{yi}\end{pmatrix}, \mathbb{H}_{2i} = \begin{pmatrix}-\mathbf{V}'_i\mathbf{F}^0_x(\mathbf{F}^{0\prime}_x\mathbf{F}^0_x)^{-1}\mathbf{F}^{0\prime}_x\mathbf{F}_y\boldsymbol{\gamma}_{yi}\\ -\mathbf{V}'_{i,-1}\mathbf{F}^0_{x,-1}(\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}^0_{x,-1})^{-1}\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}_y\boldsymbol{\gamma}_{yi}\end{pmatrix},\\
\mathbb{H}_{3i} &= \begin{pmatrix}\mathbf{V}'_i\boldsymbol{\varepsilon}_i\\ \mathbf{V}'_{i,-1}\boldsymbol{\varepsilon}_i\end{pmatrix}, \mathbb{H}_{4i} = \begin{pmatrix}-\mathbf{V}'_i\mathbf{F}^0_x(\mathbf{F}^{0\prime}_x\mathbf{F}^0_x)^{-1}\mathbf{F}^{0\prime}_x\boldsymbol{\varepsilon}_i\\ -\mathbf{V}'_{i,-1}\mathbf{F}^0_{x,-1}(\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}^0_{x,-1})^{-1}\mathbf{F}^{0\prime}_{x,-1}\boldsymbol{\varepsilon}_i\end{pmatrix},\\
\mathbb{H}_{5i} &= \begin{pmatrix}\mathbf{0}\\ \mathbf{V}'_{i,-1}\mathbf{P}_{\mathbf{F}^0_x}\mathbf{F}_y\boldsymbol{\gamma}_{yi}\end{pmatrix}, \mathbb{H}_{6i} = \begin{pmatrix}\mathbf{0}\\ -\mathbf{V}'_{i,-1}\mathbf{F}^0_{x,-1}(\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}^0_{x,-1})^{-1}\mathbf{F}^{0\prime}_{x,-1}\mathbf{P}_{\mathbf{F}^0_x}\mathbf{F}_y\boldsymbol{\gamma}_{yi}\end{pmatrix},\\
\mathbb{H}_{7i} &= \begin{pmatrix}\mathbf{0}\\ \mathbf{V}'_{i,-1}\mathbf{P}_{\mathbf{F}^0_x}\boldsymbol{\varepsilon}_i\end{pmatrix}, \mathbb{H}_{8i} = \begin{pmatrix}\mathbf{0}\\ -\mathbf{V}'_{i,-1}\mathbf{F}^0_{x,-1}(\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}^0_{x,-1})^{-1}\mathbf{F}^{0\prime}_{x,-1}\mathbf{P}_{\mathbf{F}^0_x}\boldsymbol{\varepsilon}_i\end{pmatrix},
\end{aligned}$$

with $\mathbf{u}_i = \mathbf{F}_y\boldsymbol{\gamma}_{yi} + \boldsymbol{\varepsilon}_i$, we have

$$\begin{aligned}
\mathbb{G}_1 =&N^{-1}\sum_{i=1}^{N}[\left(\tilde{\mathbf{A}}'_{i,T}\tilde{\mathbf{B}}^{-1}_{i,T}\tilde{\mathbf{A}}_{i,T}\right)^{-1}\tilde{\mathbf{A}}'_{i,T}\tilde{\mathbf{B}}^{-1}_{i,T} - \left(\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\right)^{-1}\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}](T^{-1}\mathbf{Z}'_i\mathbf{M}_{F^0_x}\mathbf{u}_i)\\
&+\sum_{\ell=1}^{8}N^{-1}\sum_{i=1}^{N}\left(\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\right)^{-1}\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}T^{-1}\mathbb{H}_{\ell i}.
\end{aligned}$$

With (18), we can show that the first term is $O_p(N^{1/4}(\ln N)^5T^{-1}) + O_p(N^{1/2}(\ln N)^5T^{-3/2})$. It's easy to show that the fifth term, the eighth term and the ninth term both are $O_p((\ln N)^2T^{-1})$. For the second term, we have

$$\begin{aligned}
&E\|N^{-1}\sum_{i=1}^{N}\left(\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\right)^{-1}\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}T^{-1}\mathbb{H}_{1i}\|^2\\
=&\mathrm{tr}\Big(N^{-2}T^{-2}\sum_{i\neq j}E\Big[\left(\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\right)^{-1}\mathbf{A}'_{i,T}\Big]\mathbf{B}^{-1}_{i,T}E\Big[\mathbb{H}_{1i}\mathbb{H}'_{1j}\Big]\mathbf{B}^{-1}_{j,T}E\Big[\mathbf{A}_{j,T}\left(\mathbf{A}'_{j,T}\mathbf{B}^{-1}_{j,T}\mathbf{A}_{j,T}\right)^{-1}\Big]\Big)\\
&+\mathrm{tr}\Big(N^{-2}T^{-2}\sum_{i=1}^{N}E\Big[\left(\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\right)^{-1}\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}E\Big[\mathbb{H}_{1i}\mathbb{H}'_{1i}\Big]\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\left(\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\right)^{-1}\Big]\Big)\\
\leq&\Delta\Big\|N^{-2}T^{-2}\sum_{i=1}^{N}E\mathrm{vec}\Big[\left(\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\right)^{-1}\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}E\Big[\mathbb{H}_{1i}\mathbb{H}'_{1i}\Big]\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\left(\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\right)^{-1}\Big]\Big\|\\
=&\Delta\Big\|N^{-2}T^{-2}\sum_{i=1}^{N}E\Big(\Big[\left(\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\right)^{-1}\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\Big]\otimes\Big[\left(\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\right)^{-1}\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\Big]\Big)\mathbf{vec}(E\Big[\mathbb{H}_{1i}\mathbb{H}'_{1i}\Big])\Big\|\\
\leq&\Delta N^{-2}T^{-2}\sum_{i=1}^{N}\Big\|E\Big(\Big[\left(\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\right)^{-1}\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\Big]\otimes\Big[\left(\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\right)^{-1}\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\Big]\Big)\Big\|\Big\|E(\mathbb{H}_{1i}\mathbb{H}'_{1i})\Big\|\\
\leq&\Delta N^{-2}T^{-2}\sum_{i=1}^{N}E\|(\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T})^{-1}\mathbf{A}'_{i,T}\|^2\|\mathbf{B}^{-1}_{i,T}\|\|E(\mathbb{H}_{1i}\mathbb{H}'_{1i})\|\leq\Delta N^{-2}T^{-2}\sum_{i=1}^{N}\|E(\mathbb{H}_{1i}\mathbb{H}'_{1i})\|\\
\leq&\Delta N^{-2}T^{-2}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}(\|E(\mathbf{v}_{is}\mathbf{v}'_{it})\|+2\|E(\mathbf{v}_{is}\mathbf{v}'_{i,t-1})\|+\|E(\mathbf{v}_{i,s-1}\mathbf{v}'_{i,t-1})\|)\leq\Delta N^{-1}T^{-1}
\end{aligned}$$

because

$$|E(\mathbf{f}_{y,s}\gamma_{yi}\gamma'_{yj}\mathbf{f}'_{y,t})|\leq\sqrt{E\|\mathbf{f}_{y,s}\|^4E\|\gamma_{yi}\|^4}\leq\Delta$$

and

$$\begin{aligned}
&E(\mathbb{H}_{1i}\mathbb{H}'_{1j})\\
=&\begin{pmatrix}\sum_{s=1}^T\sum_{t=1}^T E(\mathbf{v}_{is}\mathbf{v}'_{jt})E(\mathbf{f}_{y,s}\gamma_{yi}\gamma'_{yj}\mathbf{f}'_{y,t}) & \sum_{s=1}^T\sum_{t=1}^T E(\mathbf{v}_{is}\mathbf{v}'_{j,t-1})E(\mathbf{f}_{y,s}\gamma_{yi}\gamma'_{yj}\mathbf{f}'_{y,t})\\ \sum_{s=1}^T\sum_{t=1}^T E(\mathbf{v}_{i,s-1}\mathbf{v}'_{j,t})E(\mathbf{f}_{y,s}\gamma_{yi}\gamma'_{yj}\mathbf{f}'_{y,t}) & \sum_{s=1}^T\sum_{t=1}^T E(\mathbf{v}_{i,s-1}\mathbf{v}'_{j,t-1})E(\mathbf{f}_{y,s}\gamma_{yi}\gamma'_{yj}\mathbf{f}'_{y,t})\end{pmatrix},
\end{aligned}$$

which indicates that $E(\mathbb{H}_{1i}\mathbb{H}'_{1j})=0$ for $i\neq j$, then the second term is $O_p(N^{-1/2}T^{-1/2})$. Consider the third term. Note that

$$\begin{aligned}
\mathbb{H}_{2i}=\mathrm{vec}(\mathbb{H}_{2i})&=\begin{pmatrix}\gamma'_{yi}\otimes(\mathbf{V}'_i\mathbf{F}^0_x)\mathrm{vec}\big[(\mathbf{F}^{0\prime}_x\mathbf{F}^0_x)^{-1}\mathbf{F}^{0\prime}_x\mathbf{F}_y\big]\\ \gamma'_{yi}\otimes(\mathbf{V}'_{i,-1}\mathbf{F}^0_{x,-1})\mathrm{vec}\big[(\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}^0_{x,-1})^{-1}\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}_y\big]\end{pmatrix}\\
&=\begin{pmatrix}\gamma'_{yi}\otimes(\mathbf{V}'_i\mathbf{F}^0_x) & \mathbf{0}\\ \mathbf{0} & \gamma'_{yi}\otimes(\mathbf{V}'_{i,-1}\mathbf{F}^0_{x,-1})\end{pmatrix}\begin{pmatrix}\mathrm{vec}\big[(\mathbf{F}^{0\prime}_x\mathbf{F}^0_x)^{-1}\mathbf{F}^{0\prime}_x\mathbf{F}_y\big]\\ \mathrm{vec}\big[(\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}^0_{x,-1})^{-1}\mathbf{F}^{0\prime}_{x,-1}\mathbf{F}_y\big]\end{pmatrix}\\
&=\mathbb{H}_{2ia}\times\mathbb{H}_{2ib}.
\end{aligned}$$

It's easy to prove that $\mathbb{H}_{2ib}=O_p(1)$. Following the argument in the proof of the second term, we can prove that

$$-N^{-1}\sum_{i=1}^{N}\left(\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}\mathbf{A}_{i,T}\right)^{-1}\mathbf{A}'_{i,T}\mathbf{B}^{-1}_{i,T}T^{-1}\mathbb{H}_{2ia}=O_p(N^{-1/2}T^{-1/2}).$$

Then the third term is $O_p(N^{-1/2}T^{-1/2})$. Analogously, the forth term, the sixth term and the seventh term can be proved to be $O_p(N^{-1/2}T^{-1/2})$. Thus, $\mathbb{G}_1=O_p(N^{1/4}(\ln N)^5T^{-1})+O_p(N^{1/2}(\ln N)^5T^{-3/2})+O_p(N^{-1/2}T^{-1/2})$.

Combining the above terms, we can show that

$$N^{-1}\sum_{i=1}^{N}(\widehat{\boldsymbol{\theta}}_{IV,i}-\boldsymbol{\theta}_i)=O_p(N^{1/4}(\ln N)^5T^{-1})+O_p(N^{1/2}(\ln N)^5T^{-3/2})+O_p((\ln N)^2\delta^{-2}_{NT}).$$

Note that $N^{-1}\sum_{i=1}^{N}\boldsymbol{\eta}_i=O_p(N^{-1/2})$, if $N^{3+\delta}/T^4\to 0$ for any $\delta>0$, we have

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{IVMG}-\boldsymbol{\theta})=N^{-1/2}\sum_{i=1}^{N}\boldsymbol{\eta}_i+o_p(1)$$

and

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_{IVMG}-\boldsymbol{\theta})\xrightarrow{d}N(\mathbf{0},\boldsymbol{\Sigma}_\eta).$$

Next, we consider the consistency of $\widehat{\mathbf{\Sigma}}_\eta$. By decomposition, we have

$$\begin{aligned}&\sum_{i=1}^N(\hat{\boldsymbol\theta}_{IV,i}-\boldsymbol\theta+\boldsymbol\theta-\hat{\boldsymbol\theta}_{IVMG})(\hat{\boldsymbol\theta}_{IV,i}-\boldsymbol\theta+\boldsymbol\theta-\hat{\boldsymbol\theta}_{IVMG})'\\&=\sum_{i=1}^N\boldsymbol\eta_i\boldsymbol\eta_i'+\sum_{i=1}^N(\hat{\boldsymbol\theta}_{IV,i}-\boldsymbol\theta_i)(\hat{\boldsymbol\theta}_{IV,i}-\boldsymbol\theta_i)'+\sum_{i=1}^N(\hat{\boldsymbol\theta}_{IV,i}-\boldsymbol\theta_i)\boldsymbol\eta_i'+\sum_{i=1}^N\boldsymbol\eta_i(\hat{\boldsymbol\theta}_{IV,i}-\boldsymbol\theta_i)'\\&\quad-N(\boldsymbol\theta-\hat{\boldsymbol\theta}_{IVMG})'(\boldsymbol\theta-\hat{\boldsymbol\theta}_{IVMG}).\end{aligned}$$

Then

$$\begin{aligned}&\widehat{\mathbf{\Sigma}}_\eta-\mathbf{\Sigma}_\eta\\=&\frac{1}{N-1}\sum_{i=1}^N(\boldsymbol\eta_i\boldsymbol\eta_i'-\mathbf{\Sigma}_\eta)+\frac{1}{N-1}\sum_{i=1}^N\left(\hat{\boldsymbol\theta}_{IV,i}-\boldsymbol\theta_i\right)\left(\hat{\boldsymbol\theta}_{IV,i}-\boldsymbol\theta_i\right)'+\frac{1}{N-1}\sum_{i=1}^N\left(\hat{\boldsymbol\theta}_{IV,i}-\boldsymbol\theta_i\right)\boldsymbol\eta_i'\\&+\frac{1}{N-1}\sum_{i=1}^N\boldsymbol\eta_i\left(\hat{\boldsymbol\theta}_{IV,i}-\boldsymbol\theta_i\right)'-\frac{N}{N-1}(\boldsymbol\theta-\hat{\boldsymbol\theta}_{IVMG})'(\boldsymbol\theta-\hat{\boldsymbol\theta}_{IVMG})\\=&\mathbb{J}_1+\cdots+\mathbb{J}_5.\end{aligned}$$

Easily, we can derive that $\mathbb{J}_1=O_p(N^{-1/2})$, $\mathbb{J}_5=O_p(N^{-1})$. Consider $\mathbb{J}_3$, which is

$$\begin{aligned}&\frac{1}{N-1}\sum_{i=1}^N(\widehat{\boldsymbol\theta}_{IV,i}-\boldsymbol\theta_i)\boldsymbol\eta_i'\\=&\frac{1}{N-1}\sum_{i=1}^N\left(\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\tilde{\mathbf{A}}_{i,T}\right)^{-1}\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}(T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i)\boldsymbol\eta_i'\\&+\frac{1}{N-1}\sum_{i=1}^N\left[(\hat{\tilde{\mathbf{A}}}_{i,T}'\hat{\tilde{\mathbf{B}}}_{i,T}^{-1}\hat{\tilde{\mathbf{A}}}_{i,T})^{-1}\hat{\tilde{\mathbf{A}}}_{i,T}'\hat{\tilde{\mathbf{B}}}_{i,T}^{-1}-(\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\right](T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i)\boldsymbol\eta_i'\\&+\frac{1}{N-1}\sum_{i=1}^N\left[(\hat{\tilde{\mathbf{A}}}_{i,T}'\hat{\tilde{\mathbf{B}}}_{i,T}^{-1}\hat{\tilde{\mathbf{A}}}_{i,T})^{-1}\hat{\tilde{\mathbf{A}}}_{i,T}'\hat{\tilde{\mathbf{B}}}_{i,T}^{-1}-(\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\right](T^{-1}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat F_x}\mathbf{u}_i-T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i)\boldsymbol\eta_i'\\&+\frac{1}{N-1}\sum_{i=1}^N(\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}(T^{-1}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat F_x}\mathbf{u}_i-T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i)\boldsymbol\eta_i'.\end{aligned}$$

With $\sup_{1\le i\le N}\|\boldsymbol\eta_i\|=O_p(N^{1/4})$, we can follow the argument in the proof of the terms $\mathbb{G}_2$ to $\mathbb{G}_4$, to prove that the second term is $O_p(N^{3/4}(\ln N)^5T^{-1/2}\delta_{NT}^{-2})$, the third term is $O_p(N^{3/4}(\ln N)^5\delta_{NT}^{-4})$, the forth is $O_p(N^{1/4}(\ln N)^2\delta_{NT}^{-2})$. By Lemma 18 (b), the first term is bounded in norm by

$$(N-1)^{-1}\sum_{i=1}^N\|T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i\|\cdot\sup_{1\le i\le N}\|(\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\|\cdot\sup_{1\le i\le N}\|\boldsymbol\eta_i\|=O_p(N^{1/4}(\ln N)^2T^{-1/2}).$$

Then $\mathbb{J}_3=O_p(N^{3/4}(\ln N)^5T^{-1/2}\delta_{NT}^{-2})+O_p(N^{3/4}(\ln N)^5\delta_{NT}^{-4})+O_p(N^{1/4}(\ln N)^2\delta_{NT}^{-2})+O_p(N^{1/4}(\ln N)^2T^{-1/2})$. $\mathbb{J}_4$ is the same order of $\mathbb{J}_3$ since it is transpose of $\mathbb{J}_3$.

Consider $\mathbb{J}_2$, with $(a+b+c+d)^2\le 4(a^2+b^2+c^2+d^2)$, it is bounded in norm by

$$\begin{aligned}&\frac{1}{N-1}\sum_{i=1}^N\|\widehat{\boldsymbol\theta}_{IV,i}-\boldsymbol\theta_i\|^2\\\le&\frac{4}{N-1}\sum_{i=1}^N\|\left(\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\tilde{\mathbf{A}}_{i,T}\right)^{-1}\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}(T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i)\|^2\\&+\frac{4}{N-1}\sum_{i=1}^N\left\|\left[(\hat{\tilde{\mathbf{A}}}_{i,T}'\hat{\tilde{\mathbf{B}}}_{i,T}^{-1}\hat{\tilde{\mathbf{A}}}_{i,T})^{-1}\hat{\tilde{\mathbf{A}}}_{i,T}'\hat{\tilde{\mathbf{B}}}_{i,T}^{-1}-(\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\right](T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i)\right\|^2\\&+\frac{4}{N-1}\sum_{i=1}^N\left\|\left[(\hat{\tilde{\mathbf{A}}}_{i,T}'\hat{\tilde{\mathbf{B}}}_{i,T}^{-1}\hat{\tilde{\mathbf{A}}}_{i,T})^{-1}\hat{\tilde{\mathbf{A}}}_{i,T}'\hat{\tilde{\mathbf{B}}}_{i,T}^{-1}-(\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\right](T^{-1}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat F_x}\mathbf{u}_i-T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i)\right\|^2\\&+\frac{4}{N-1}\sum_{i=1}^N\|(\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}(T^{-1}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat F_x}\mathbf{u}_i-T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i)\|^2.\end{aligned}$$

By Lemma 18 (b), the first term is bounded in norm by

$$4(N-1)^{-1}\sum_{i=1}^{N}\|T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i\|^2\cdot(\sup_{1\le i\le N}\|(\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\tilde{\mathbf{A}}_{i,T})^{-1}\tilde{\mathbf{A}}_{i,T}'\tilde{\mathbf{B}}_{i,T}^{-1}\|)^2=O_p((\ln N)^4T^{-1})$$

similarly, we can show that the second term is $O_p(N(\ln N)^{10}T^{-1}\delta_{NT}^{-4})$. Following the argument in the proof of Lemma 15 (a) and (b), we can show that $N^{-1}\sum_{i=1}^{N}\|T^{-1}\hat{\mathbf{Z}}_i'\mathbf{M}_{\hat{F}_x}\mathbf{u}_i-T^{-1}\mathbf{Z}_i'\mathbf{M}_{F_x^0}\mathbf{u}_i\|^2=O_p(\delta_{NT}^{-4})$. Then similar to the argument in the proof of the first term, we can prove that the third term is $O_p(N(\ln N)^{10}\delta_{NT}^{-8})$ and the forth term is $O_p((\ln N)^4\delta_{NT}^{-4})$. Then $\mathbb{J}_2=O_p((\ln N)^4T^{-1})+O_p(N(\ln N)^{10}T^{-1}\delta_{NT}^{-4})+O_p(N(\ln N)^{10}\delta_{NT}^{-8})+O_p((\ln N)^4\delta_{NT}^{-4})$.

Combining the terms $\mathbb{J}_1$ to $\mathbb{J}_5$, we can derive that $\widehat{\mathbf{\Sigma}}_\eta-\mathbf{\Sigma}_\eta=o_p(1)$. Thus, we complete the proof. ■

References

Ahn, S., Horenstein, A., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81, 1203–1227.

Ahn, S., Lee, Y., Schmidt, P., 2013. Panel data models with multiple time varying effects. Journal of Econometrics 174, 1–14.

Alvarez, J., Arellano, M., 2003. The time series and cross-section asymptotics of dynamic panel data estimators. Econometrica 71, 1121–1159.

Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–173.

Bai, J., 2009a. Panel data models with interactive fixed effects. Econometrica 77, 1229–1279.

Bai, J., 2009b. Supplement to "panel data models with interactive fixed effects": technical details and proofs. Econometrica Supplemental Material .

Bai, J., Li, K., 2014. Theory and methods of panel data models with interactive effects. Annals of Statistics 42, 142–170.

Bai, J., Ng, S., 2010. Instrumental variable estimation in a data rich environment. Econometric Theory 26, 1577–1606. doi:10.1017/S0266466609990727.

Bai, J., Ng, S., 2013. Principal components estimation and identification of static factors. Journal of Econometrics 176, 18–29.

Bai, J., Wang, P., 2016. Econometric analysis of large factor models. Annual Review of Economics 8, 53–80.

Bun, M., Kiviet, J., 2006. The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. Journal of Econometrics 132, 409–444.

Carneiro, P., Hansen, K., Heckman, J., 2003. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college. NBER working paper No. w9546 .

Chudik, A., Pesaran, M.H., 2015a. Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. Journal of Econometrics 188, 393–420.

Chudik, A., Pesaran, M.H., 2015b. Large panel data models with cross-sectional dependence: A survey, in: Baltagi, B.H. (Ed.), The Oxford Handbook of Panel Data. Oxford University Press, pp. 3–45.

Favero, C., Marcellino, M., Neglia, F., 2005. Principal components at work: The empirical analysis of monetary policy with large data sets. Journal of Applied Econometrics 20, 603–620.

Harding, M., Lamarche, C., 2011. Least squares estimation of a panel data model with multifactor error structure and endogenous covariates. Economics Letters 111, 197–199.

Heckman, J., Stixrud, J., Urzua, S., 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. Journal of Labour Economics 24, 411–482.

Juodis, A., Sarafidis, V., 2018. Fixed t dynamic panel data estimators with multifactor errors. Econometric Reviews 36, 893 – 929.

Juodis, A., Sarafidis, V., 2020. A linear estimator for factor-augmented fixed-t panels with endogenous regressors. mimeo URL: `https://www.monash.edu/business/ebs/research/publications/ebs/wp05-2020.pdf`.

Kapetanios, G., Marcellino, M., 2010. Factor-GMM estimation with large sets of possibly weak instruments. Computational Statistics & Data Analysis 54, 2655 – 2675. URL: `http://www.sciencedirect.com/science/article/pii/S0167947310001532`, doi:https://doi.org/10.1016/j.csda.2010.04.008. the Fifth Special Issue on Computational Econometrics.

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

Moon, H.R., Weidner, M., 2017. Dynamic linear panel regression models with interactive fixed effects. Econometric Theory 33, 158–195.

Moon, H.R., Weidner, M., 2019. Nuclear norm regularized estimation of panel regression models. cemmap working paper 14/19.

Newey, W., 1985. Generalized method of moments specification testing. Journals of Econometrics 29, 229–256.

Okui, R., 2009. The optimal choice of moments in dynamic panel data models. Journal of Econometrics 151, 1–16.

Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967–1012.

Robertson, D., Sarafidis, V., 2015. IV estimation of panels with factor residuals. Journal of Econometrics 185, 526–541.

Sarafidis, V., Wansbeek, T., 2012. Cross-sectional dependence in panel data analysis. Econometric Reviews 31, 483–531.

Sarafidis, V., Yamagata, T., Robertson, D., 2009. A test of cross section dependence for a linear dynamic panel model with regressors. Journal of Econometrics 148, 149–161.

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

Wooldridge, J., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press.