Instrumental Variable Estimation of Dynamic Linear Panel Data Models with Defactored Regressors under Cross-sectional Dependence

Sarafidis, Vasilis and Yamagata, Takashi

Univeristy of Sydney, University of York, United Kingdom

February 2010

Instrumental Variable Estimation of Dynamic Linear Panel Data Models with Defactored Regressors under Cross-sectional Dependence*

Vasilis Sarafidis[†] Takashi Yamagata[‡]
University of Sydney University of York

February 2010

Abstract

This paper develops an instrumental variable (IV) estimator for consistent estimation of dynamic panel data models with error cross-sectional dependence when both N and T, the cross-section and time series dimensions respectively, are large. Our approach asymptotically projects out the common factors from regressors using principal components analysis and then uses the defactored regressors as instruments to estimate the model in a standard way. Therefore, the proposed estimator is computationally very attractive. Furthermore, our procedure requires estimating only the common factors included in the regressors, leaving those that influence the dependent variable solely into the errors. Hence aside from computational simplicity the resulting approach allows parsimonious estimation of the model. The finite-sample performance of the IV estimator and the associated t-test is investigated using simulated data. The results show that the bias of the estimator is very small and the size of the t-test is correct even when (T, N) is as small as $(10, 50)$. The performance of an overidentifying restrictions test is also explored and the evidence suggests that it has good power when the key assumption is violated.

Key Words: method of moments; dynamic panel data; cross-sectional dependence
JEL Classification: C13, C15, C23.

*The usual disclaimer applies.

[†]Faculty of Economics and Business, University of Sydney, NSW 2006, Australia. Tel: +61-2-9036 9120; E-mail address: vasilis.sarafidis@sydney.edu.au.

[‡]Corresponding author. Department of Economics and Related Studies, University of York, York YO10 5DD, UK. Tel: +44-1904-43 3708; E-mail address: ty509@york.ac.uk.

1 Introduction

The rapid increase in the availability of panel data during the last few decades has invoked a large interest in developing ways to model and analyse them effectively. In particular, the issue of how to characterise 'between group' or cross-sectional dependence, and then creating consistent estimation methods and making asymptotically valid inferences, has proven both popular and challenging. The main complication arises because there is no natural ordering of the observations, in at least one dimension, contrary to pure time series data where a natural temporal ordering exists. To deal with this problem, the literature has adopted two different approaches, the spatial approach and the factor structure approach. The former requires that the sample correlations across individuals can be measured in relation to some index of spatial distance, in a geographic or economic sense, in which case one may formulate and estimate spatial models based on the method of maximum likelihood (e.g. Lee, 2004) and the generalised method of moments (see e.g. Kapoor, Kelejian, and Prucha, 2007; Kelejian and Prucha, 2010). The factor structure approach relaxes this requirement by assuming that there exists a common component, which is a linear combination of a finite number of time-varying common factors with individual-specific factor loadings. One can provide different interpretations of this process, depending on the application in mind. In macroeconomic panels the unobserved factors are frequently viewed as economy-wide shocks, affecting all individuals with different intensities; see e.g. Favero, Marcellino, and Neglia (2005). In microeconomic panels the factor error structure may be thought to 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, specific to each individual, while the factors may capture the vector of price of these skills, which changes intertemporally; see e.g. Carneiro, Hansen, and Heckman (2003) and Heckman, Stixrud, and Urzua (2006). Ahn, Lee, and Schmidt (2001) provide further examples.

Several methods have been proposed in the literature to estimate models with a multi-factor error structure allowing for possible correlations between the unobserved common components and the included regressors; Pesaran (2006) proposes augmenting standard panel data regression models with the cross section averages of the dependent and explanatory variables, which span asymptotically the unobserved factors. Bai (2009) proposes an iterative least squares estimator based on principal components analysis.[1] However, neither Pesaran (2006) or Bai (2009) consider dynamic panel data models or models with weakly exogenous regressors in general.

Ahn, Lee, and Schmidt (2006) put forward a GMM estimator that is based on a quasi-differencing transformation that eliminates the factor structure in the residuals. Robertson, Sarafidis, and Symons (2010) develop an instrumental-variable estimation procedure that introduces new parameters to represent the unobserved covariances between the covariates and the factor component of the residual, and they show that the resulting estimator is asymptotically more efficient than Ahn, Lee and Schmidt. Bai (2010) models the correlation between the common components and the included regressors as in Chamberlain (1982) and proposes estimating the model using a likelihood approach. Notice that contrary to Pesaran (2006) and Bai (2009) these methods allow for dynamic panel data models and weakly exogenous regressors. On the other hand, the

[1]For pure factor models, see Bai and Ng (2002), Bai (2003), Forni, Hallin, Lippi, and Reichlin (2000), among others.

associated estimation algorithms are rather complicated and they can be computationally expensive when T is moderately large.

In view of these, in this paper we propose a computationally attractive instrumental variable (IV) estimator for consistent estimation of dynamic linear panel data models under cross-sectional dependence when both N and T are large. Our approach at first stage asymptotically projects out the common component from regressors using principal components analysis and then uses the defactored regressors as instruments to estimate the structural parameters. The required assumption is that endogeneity of the covariates arises due to the non-zero correlation between these variables and the common components in the disturbance rather than the idiosyncratic component. Importantly, this assumption can be tested using an overidentifying restrictions test. Our procedure requires estimating solely the common factors included in the regressors, leaving those that influence only the dependent variable into the residuals. Hence aside from computational simplicity the resulting approach allows parsimonious estimation of the model. The finite sample performance of the proposed IV estimator and the associated t-test is investigated using simulated data. The results show that the bias of the estimator is very small and the size of the t-test is correct even when (T, N) is as small as $(10, 50)$. Furthermore, the overidentifying restrictions test appears to have good power when the key assumption is violated.

The paper is organised as follows. Section 2 discusses the model and the estimation methods. The finite sample performance of the proposed estimator is summarised in Section 3. Section 4 contains some concluding remarks.

2 Model and Estimation Method

Consider the following autoregressive distributed lag, ARDL(1,0), panel data model[2]:

$$y_{it} = \alpha + \lambda y_{i,t-1} + \beta' \mathbf{x}_{it} + u_{it}, i = 1, 2, ..., N, t = 0, 1, ..., T, \tag{1}$$

with multi-factor error structure

$$u_{it} = \alpha_i + \gamma_i' \mathbf{f}_t + \varepsilon_{it}, \tag{2}$$

where α_i is an individual-specific time-invariant effect; the common component, $\gamma_i' \mathbf{f}_t$, consists of $\mathbf{f}_t = (f_{1t}, f_{2t}, ..., f_{mt})'$, an $m \times 1$ vector of unobservable factors, and γ_i, an $m \times 1$ vector of factor loadings in the equation for y; ε_{it} is an idiosyncratic error. $\mathbf{x}_{it} = (x_{1it}, x_{2it}, ..., x_{kit})'$ is a $k \times 1$ vector of regressors, which obeys the following process:

$$\mathbf{x}_{it} = \mu_{ix} + \mathbf{\Gamma}_{ix}' \mathbf{f}_t + \mathbf{v}_{it}, \tag{3}$$

where μ_i is a vector of individual-specific effects, potentially correlated with α_i; $\mathbf{\Gamma}_{xi} = (\gamma_{x1i}, \gamma_{x2i}, ..., \gamma_{xki})$ is a $m \times k$ factor loading matrix; and $\mathbf{v}_{it}$ is an idiosyncratic error term.

We define $\Delta \equiv 1 - L$, where L is a lag operator such that $L^\ell y_t \equiv y_{t-\ell}$. Taking first-differences in (1) to eliminate the individual-specific time-invariant effects, α_i and μ_{ix}, and stacking the T observations for each i (making a $T \times 1$ vector) yields

$$\Delta \mathbf{y}_i = \lambda \Delta \mathbf{y}_{i,-1} + \Delta \mathbf{X}_i \beta_i + \Delta \mathbf{u}_i, \tag{4}$$

[2]The main results of this paper naturally extend to models with higher order lags, i.e. ARDL(p,q) for $p > 0$ and $q \geq 0$.

with

$$\Delta \mathbf{u}_i = \Delta \mathbf{F} \gamma_i + \Delta \varepsilon_i, \tag{5}$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{iT})'$, $\mathbf{y}_{i,-1} = L^1 \mathbf{y}_i = (y_{i0}, y_{i1}, ..., y_{iT-1})'$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})'$, $\mathbf{u}_i = (u_{i1}, u_{i2}, ..., u_{iT})'$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$ and $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})'$.[3] By letting $\Delta \mathbf{W}_i = (\Delta \mathbf{y}_{i,-1}, \Delta \mathbf{X}_i)$ and $\theta = (\lambda, \beta')'$, we can write (4) more concisely as follows:

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

Similarly, taking first-differences in (3) to eliminate the random effects, μ_{ix}, and stacking the T observations for each i yields

$$\Delta \mathbf{X}_i = \Delta \mathbf{F} \mathbf{\Gamma}_{xi} + \Delta \mathbf{V}_i. \tag{7}$$

$\Delta \mathbf{W}_i$ are heterogeneously cross sectionally correlated through the multifactor error structure. Also the composite error $\Delta \mathbf{u}_i$ is allowed to be serially correlated through the serial correlation in the factors, $\Delta \mathbf{f}_t$.

Our proposed approach involves asymptotically eliminating at first stage the common factors in $\Delta \mathbf{X}_i$ by projecting them out, and then using the defactored regressors as instruments to estimate the structural parameters of the model. To see the main idea, consider the following projection matrix:

$$\mathbf{M}_{\Delta F} = \mathbf{I}_T - \Delta \mathbf{F} \left(\Delta \mathbf{F}' \Delta \mathbf{F} \right)^{-1} \Delta \mathbf{F}'. \tag{8}$$

If $\Delta \mathbf{F}$ were observable, premultiplying $\Delta \mathbf{X}_i$ by $\mathbf{M}_{\Delta F}$ would yield $\mathbf{M}_{\Delta F} \Delta \mathbf{X}_i = \mathbf{M}_{\Delta F} \Delta \mathbf{V}_i$, which, under certain conditions to be specified shortly, satisfies $E(\Delta \mathbf{X}_i' \mathbf{M}_{\Delta F} \Delta \mathbf{u}_i) = E(\Delta \mathbf{V}_i' \mathbf{M}_{\Delta F} \Delta \varepsilon_i) = \mathbf{0}$. Now let

$$\mathbf{X}_{i,-j} = L^j \mathbf{X}_i. \tag{9}$$

Supposing that $\{y_{it}, \mathbf{x}_{it}'\}$, $t = -1, 0, 1, ..., T$ are observable, the $T \times k$ matrix $\Delta \mathbf{X}_{i,-1}$ is also observable (but not $\Delta \mathbf{X}_{i,-j}$ for $j > 1$). It is easily seen that $E(\Delta \mathbf{X}_{i,-1}' \mathbf{M}_{\Delta F} \Delta \mathbf{u}_i) = E(\Delta \mathbf{X}_{i,-1}' \mathbf{M}_{\Delta F} \Delta \varepsilon_i) = \mathbf{0}$. Now let

$$\mathbf{Z}_i = [\Delta \mathbf{X}_i, \Delta \mathbf{X}_{i,-1}] \;\; (T \times 2k). \tag{10}$$

Given model (6) it is clear that the defactored regressors satisfy instrument relevance, i.e. $E(\mathbf{Z}_i' \mathbf{M}_{\Delta F} \Delta \mathbf{W}_i) \neq \mathbf{0}$. Therefore, it is relatively straightforward to apply instrumental variable (IV) estimation using $\mathbf{M}_{\Delta F} \mathbf{Z}_i$ as an instrument vector for $\Delta \mathbf{W}_i$.[4]

In practice, the factor vector, $\Delta \mathbf{F}$, is not observable of course. As a result, we propose estimating $\Delta \mathbf{F}$ using a principal components approach, as advanced in Bai (2003) and Bai (2009).[5]

To obtain our results it is sufficient to make the following assumptions:

[3] In this paper we consider cross-sectional correlation that is due to the factor structure only, however, our results below would be asymptotically valid even when the idiosyncratic errors are weakly cross-sectionally correlated.

[4] More instruments become available when further histories of $\mathbf{x}_{it}$ are observable. In particular, given model (3), when $\{\mathbf{x}_{it}\}_{t=0-j}^{T}$ for $j \geq 0$ are observable, $(j+1)k$ instruments, $\{\Delta \mathbf{X}_{i,-(r-1)}\}_{r=1}^{j+1}$, become available.

[5] A popular alternative is the common correlated effects approach of Pesaran (2006). We consider this in the experimental section.

Assumption 1 (idiosyncratic errors): (i) ε_{it} is independently distributed across i and t, having mean zero, $E(\Delta\varepsilon_i\Delta\varepsilon_i') = \mathbf{\Sigma}_{\Delta\varepsilon i}$ which is positive definite. ε_{it} has finite fourth order moment; (ii) $\mathbf{v}_{it} = \mathbf{\Psi}_i(L)\mathbf{e}_{v,it}$, where $\mathbf{\Psi}_i(L)$ is absolutely summable and $\mathbf{e}_{v,it} \sim iid(\mathbf{0}, \mathbf{\Sigma}_{v,i})$ across i and t, where $\mathbf{\Sigma}_{v,i}$ is a positive definite matrix. $\mathbf{e}_{v,it}$ has finite fourth order moments and is group-wise independent from ε_{it}.

Assumption 2 (stationary factors): $\mathbf{f}_t = \mathbf{\Phi}_i(L)\mathbf{e}_{f,t}$, where $\mathbf{\Phi}(L)$ is absolutely summable, and $\mathbf{e}_{f,t} \sim iid(\mathbf{0}, \mathbf{\Sigma}_f)$, where $\mathbf{\Sigma}_f$ is a positive definite matrix. $\mathbf{e}_{f,t}$ has finite fourth order moments and is group-wise independent from $\mathbf{e}_{v,it}$ and ε_{it}.

Assumption 3 (random effects): $\mu_i = (\alpha_i, \mu_{ix}')' \sim iid(\mathbf{0}, \mathbf{\Sigma}_\mu)$, $\mathbf{\Sigma}_\mu$ is positive semi-definite with each element having finite fourth order moments. μ_i is group-wise independent from ε_{it}, $\mathbf{e}_{v,it}$, and $\mathbf{e}_{f,t}$.

Assumption 4 (random factor loadings): (i) $\mathbf{\Gamma}_{xi} \sim iid(\mathbf{0}, \mathbf{\Sigma}_\Gamma)$ where $\mathbf{\Sigma}_{\Gamma x}$ is a $m \times m$ positive definite matrix, and each element of $\mathbf{\Gamma}_{xi}$ has finite fourth order moments. $\mathbf{\Gamma}_{xi}$ is an independent group from ε_{it}, $\mathbf{e}_{v,it}$, ξ_i, and $\mathbf{e}_{f,t}$; (ii) the eigenvalues of $\mathbf{\Sigma}_{\Gamma x}$ are different from those of $p\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathbf{f}_t\mathbf{f}_t'$.

Assumption 5 (identification of θ): (i) $E(\mathbf{Z}_i'\mathbf{M}_{\Delta F}\Delta\mathbf{W}_i) = \mathbf{A}_{i,T}$ is uniformly bounded and $\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{A}_{i,T} = \mathbf{A}$ is a fixed $2k\times(1+k)$ matrix with full column rank; (ii) $E(\mathbf{Z}_i'\mathbf{M}_{\Delta F}\mathbf{Z}_i) = \mathbf{B}_{i,T}$ is uniformly bounded and $\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{B}_{i,T} = \mathbf{B}$, which is a fixed positive definite square matrix of dimension $2k$; (iii) $E(\boldsymbol{Z}_i'\boldsymbol{M}_{\Delta F}\boldsymbol{\Delta u}_i\boldsymbol{\Delta u}_i'\boldsymbol{M}_{\Delta F}\boldsymbol{Z}_i) = \mathbf{\Omega}_{i,T}$ is uniformly bounded and $\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{\Omega}_{i,T} = \mathbf{\Omega}$, which is a fixed positive definite square matrix of dimension $2k$.

The assumptions above require some discussion. First of all, notice that Assumption 1(i) allows non-normality and cross-sectional heteroskedasticity in the idiosyncratic errors in the equation for y. Furthermore, Assumption 1 implies that the covariates are strongly exogenous with respect to the idiosyncratic error component (i.e. $E(\Delta\varepsilon_i|\Delta\mathbf{X}_i) = \mathbf{0}$). Dynamic panel data models with strongly exogenous regressors is a common framework in the economics literature; some examples include partial adjustment models for labour supply (e.g. Bover, 1991), household consumption models with habits (e.g. Becker, Grossman, and Murphy, 1994) and production functions with adjustment costs (e.g. Blundell and Bond, 2000). In these applications the autoregressive parameter captures consumption inertia due to habits, or costs of adjustment, so it has a structural significance. Notwithstanding strong exogeneity with respect to the idiosyncratic disturbance, it is reasonable to expect that the regressors may be correlated with the unobserved common factors and are, therefore, endogenous. For instance, in a production function the input decisions of the firm are likely to be correlated with its individual-specific unobserved characteristics, γ_i and α_i, that may or may not vary over time. Likewise, determinants of labour supply, such as the level of wage offered to an individual, are likely to be correlated with the common factors influencing supply itself. In fact, this is the standard fixed effects assumption employed in panel data models, extended to the factor structure. Notice that in this case, however, first-differencing does not remove endogeneity since the factor structure remains in the residuals. The strong exogeneity assumption of the covariates with respect to the purely idiosyncratic error component can be tested using an overidentifying restrictions test, as shown below.

Assumptions 1(ii) and 2 allow serially correlated but stationary idiosyncratic errors in the equation for x and the factors. These are slightly stronger than Bai (2003), but they 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 there exist higher order moments; see Assumptions D-F in Bai (2003)[6].

Assumption 3 is a random coefficient type assumption but permits non-zero correlation between the individual effects in the y and x equations. Assumption 4 is standard in the principal components literature; see e.g. Bai (2003) among others. Notice that the zero-mean restriction on the factor loadings is not binding because for large N one can always remove the non-zero mean by transforming the variables in terms of deviations from time-specific averages or by adding time dummies into the model (4). The resulting correlation between the factor loadings are clearly $O_p(1/N)$, thus the results we obtain below are not affected by this transformation; see Sarafidis, Yamagata, and Robertson (2009) for more details.

Finally, Assumption 5 is commonplace in overidentified instrumental variable (IV) estimation; for example see Wooldridge (2002, Ch5).

Remark 1 *Assumption 4(i) does not rule out possible non-zero correlation between the factor loadings in the y and x equations, i.e. it allows $E(\gamma_i \gamma'_{x\ell i}) \neq \mathbf{0}$ for all $\ell = 1, 2, ..., k$. Since the variables y_{it} and $\mathbf{x}_{it}$ of the same cross section unit i can be affected in a related manner by the same common shocks, allowing for this possibility is potentially important in practice.*

The first step of our approach is to consistently estimate the number of factors in $\Delta\mathbf{X}_i$ using, for example, the method proposed by Bai and Ng (2002), as T and N tend jointly to infinity. Since these estimators are consistent, our discussion below treats the number of factors, m, as known. Given m, the factors are extracted using principal components from $\{\Delta\mathbf{X}_i\}_{i=1}^N$. Define $\Delta\hat{\mathbf{F}}$ as $\sqrt{T}$ times eigenvectors corresponding to the m largest eigenvalues of the $T \times T$ matrix $\sum_{i=1}^N \Delta\mathbf{X}_i \Delta\mathbf{X}'_i$; see Bai (2003) for more details. Note that $\Delta\mathbf{F}$ and $\mathbf{\Gamma}_{xi}$ are estimated up to an invertible $m \times m$ matrix transformation. Since our aim is to marginal out the unobservable common components, we treat the principal component estimator $\Delta\hat{\mathbf{F}}$ consistent to $\Delta\mathbf{F}$ in the model, without loss of generality. This is allowed because the factors and factor loadings in the model always can be redefined as $\Delta\mathbf{F}\mathbf{H}$ and $\mathbf{H}^{-1}\mathbf{\Gamma}_{xi}$, respectively, for some invertible matrix $\mathbf{H}$.

The empirical counterpart of the projection matrix defined in (8) is given by

$$\mathbf{M}_{\Delta\hat{F}} = \mathbf{I}_T - \Delta\hat{\mathbf{F}} \left(\Delta\hat{\mathbf{F}}' \Delta\hat{\mathbf{F}} \right)^{-1} \Delta\hat{\mathbf{F}}'. \tag{11}$$

The associated transformed instrument matrix discussed above is

$$\tilde{\mathbf{Z}}_i = \mathbf{M}_{\Delta\hat{F}} \mathbf{Z}_i. \tag{12}$$

Remark 2 *Note that we do not estimate the common component $\Delta\mathbf{F}\gamma_i$ in the $\Delta\mathbf{y}_i$ equation using the information contained in $\Delta\mathbf{X}_i$. Instead, we orthogonalise $\Delta\mathbf{u}_i$ to the instruments $\mathbf{Z}_i$. To see the main difference, consider the case in which $\mathbf{F}$ could be partitioned as $(\mathbf{F}_1, \mathbf{F}_2)$. Suppose the error term in $\Delta\mathbf{y}_i$ is subject to the full set of the unobserved factors, namely $\Delta\mathbf{u}_i = \Delta\mathbf{F}_1\gamma_{1i} + \Delta\mathbf{F}_2\gamma_{2i} + \Delta\varepsilon_i$, while $\Delta\mathbf{X}_i$ contains only a subset of $\Delta\mathbf{F}$, i.e. $\Delta\mathbf{X}_i = \Delta\mathbf{F}_1\mathbf{\Gamma}_{x1i} + \Delta\mathbf{V}_{it}$. Assuming $cov(\gamma_{x1i}, \gamma_{2i}) = \mathbf{0}$ in addition to Assumption 4(ii), projecting out $\Delta\mathbf{F}_1$ only is required to make $\mathbf{Z}_i = (\Delta\mathbf{X}_i, \Delta\mathbf{X}_{i,-1})$ exogenous. This is because $E(\Delta\mathbf{X}'_i\mathbf{M}_{\Delta F_1}\Delta\mathbf{u}_i) = E[\Delta\mathbf{V}'_i\mathbf{M}_{\Delta F_1}(\Delta\mathbf{F}_2\gamma_{2i} + \Delta\varepsilon_i)] = \mathbf{0}$ with $\mathbf{M}_{\Delta F_1} = \mathbf{I}_T -$*

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

$\Delta\mathbf{F}_1\left(\Delta\mathbf{F}_1'\Delta\mathbf{F}_1\right)^{-1}\Delta\mathbf{F}_1'$, *and similarly* $E(\Delta\mathbf{X}_{i,-1}'\mathbf{M}_{\Delta F_1}\Delta\mathbf{u}_i) = E[(\Delta\mathbf{F}_{1,-1}\gamma_{x1i} + \Delta\mathbf{V}_{i,-1})'$ $\mathbf{M}_{\Delta F_1}\left(\Delta\mathbf{F}_2\gamma_{2i} + \Delta\varepsilon_i\right)] = \mathbf{0}$.[7]

We propose an instrumental variable (IV) estimator or two-stage least square estimator of θ,

$$\hat{\theta}_{IV} = \left(\mathbf{A}_{NT}'\mathbf{B}_{NT}^{-1}\mathbf{A}_{NT}\right)^{-1}\mathbf{A}_{NT}'\mathbf{B}_{NT}^{-1}\mathbf{g}_{NT}, \tag{13}$$

where

$$\mathbf{A}_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathbf{Z}}_i'\Delta\mathbf{W}_i,\ \mathbf{B}_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathbf{Z}}_i'\tilde{\mathbf{Z}}_i,\ \mathbf{g}_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathbf{Z}}_i'\Delta\mathbf{y}_i. \tag{14}$$

The natural variance estimator is

$$\hat{\mathbf{Q}}_{NT} = \frac{1}{NT}\left(\mathbf{A}_{NT}'\mathbf{B}_{NT}^{-1}\mathbf{A}_{NT}\right)^{-1}\mathbf{A}_{NT}'\mathbf{B}_{NT}^{-1}\hat{\mathbf{\Omega}}_{NT}\mathbf{B}_{NT}^{-1}\mathbf{A}_{NT}\left(\mathbf{A}_{NT}'\mathbf{B}_{NT}^{-1}\mathbf{A}_{NT}\right)^{-1}, \tag{15}$$

where

$$\hat{\mathbf{\Omega}}_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathbf{Z}}_i'\Delta\hat{\mathbf{u}}_i\Delta\hat{\mathbf{u}}_i'\tilde{\mathbf{Z}}_i \tag{16}$$

with $\mathbf{\Delta}\hat{\mathbf{u}}_i = \Delta\mathbf{y}_i - \Delta\mathbf{W}_i\hat{\theta}_{IV}$.[8]

Firstly let us discuss the consistency of this estimator. Initially, from (6) and (13) we obtain

$$\sqrt{NT}\left(\hat{\theta}_{IV} - \theta\right) = \left(\mathbf{A}_{NT}'\mathbf{B}_{NT}^{-1}\mathbf{A}_{NT}\right)^{-1}\mathbf{A}_{NT}'\mathbf{B}_{NT}^{-1}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\tilde{\mathbf{Z}}_i'\Delta\mathbf{u}_i\right). \tag{17}$$

The main results of Lemma 2 in the Appendix are

$$\frac{\mathbf{Z}_i'\boldsymbol{M}_{\Delta\hat{F}}\Delta\boldsymbol{W}_i}{T} - \frac{\mathbf{Z}_i'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{W}_i}{T} = O_p\left(\delta_{NT}^{-2}\right),\ \text{uniformly over } i, \tag{18}$$

$$\frac{\mathbf{Z}_i'\boldsymbol{M}_{\Delta\hat{F}}\Delta\boldsymbol{F}}{T} - \frac{\mathbf{Z}_i'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{F}}{T} = O_p\left(\delta_{NT}^{-2}\right),\ \text{uniformly over } i, \tag{19}$$

$$\frac{\mathbf{Z}_i'\boldsymbol{M}_{\Delta\hat{F}}\Delta\boldsymbol{\varepsilon}_i}{T} - \frac{\mathbf{Z}_i'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{\varepsilon}_i}{T} = O_p\left(\delta_{NT}^{-2}\right),\ \text{uniformly over } i, \tag{20}$$

with $\delta_{NT}^2 = \min\{N, T\}$.[9] Using Lemma 2 and a law of large numbers, it is easily seen that

$$\underset{N,T\to\infty}{\text{plim}}\ \mathbf{A}_{NT} = \mathbf{A},\ \underset{N,T\to\infty}{\text{plim}}\ \mathbf{B}_{NT} = \mathbf{B}, \tag{21}$$

[7] Alternatively, without assuming $cov(\gamma_{x1i}, \gamma_{2i}) = \mathbf{0}$, one could transform $\mathbf{Z}_i$ using a projection matrix $\mathbf{M}_{\mathbf{D}} = \mathbf{I}_T - \mathbf{D}\left(\mathbf{D}'\mathbf{D}\right)^{-1}\mathbf{D}'$, $\mathbf{D} = (\mathbf{\Delta F}_1\mathbf{\Delta F}_{1,-1})$ with $\mathbf{\Delta F}_{1,-1} = L\mathbf{\Delta F}_1$. Then, $E(\Delta\mathbf{X}_i'\mathbf{M}_{\mathbf{D}}\Delta\mathbf{u}_i) = \mathbf{0}$ and $E(\Delta\mathbf{X}_{i,-1}'\mathbf{M}_{\mathbf{D}}\Delta\mathbf{u}_i) = E[\Delta\mathbf{V}_{i,-1}'\mathbf{M}_{\Delta F_1}\left(\Delta\mathbf{F}_2\gamma_{2i} + \Delta\varepsilon_i\right)] = \mathbf{0}$.

[8] Although the proposed IV estimator is based on first-differences, under Assumption 1(i), i.e. strong exogeneity of the regressors with respect to the idiosyncratic errors, our basic approach holds under alternative transformations, such as fixed effects or orthogonal deviations.

[9] See Appendix for the proof.

without any restrictions on N and T, where $\mathbf{A}$ and $\mathbf{B}$ are defined in Assumption 5. Also, by Lemma 2, it can be shown that

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\tilde{\mathbf{Z}}_i'\Delta\mathbf{u}_i &= \frac{1}{N}\sum_{i=1}^{N}\sqrt{NT}\frac{\mathbf{Z}_i'\boldsymbol{M}_{\Delta\hat{F}}\Delta\mathbf{u}_i}{T} \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{Z}_i'\boldsymbol{M}_{\Delta F}\Delta\varepsilon_i + O_p\left(\sqrt{NT}\delta_{NT}^{-2}\right). \qquad (22)
\end{aligned}$$

The first term of the right-hand side of (22) is $O_p(1)$, which tends to a standard normal variable with finite variance. In addition, the second term is $O_p\left(\min\left\{\sqrt{\frac{T}{N}},\sqrt{\frac{N}{T}}\right\}\right)$, which is $O_p(1)$ if $\frac{T}{N}$ tends to a finite positive constant c $(0<c<\infty)$ as N and $T\to\infty$ jointly. Therefore, in such a situation the IV estimator is $\sqrt{NT}$-consistent.

The above discussion is summarised in the following theorem:

Theorem 1 *Consider model (1)-(3) and suppose that Assumptions 1-5 hold true. Then,*

$$\hat{\theta}_{IV}-\theta\xrightarrow{p}\mathbf{0}$$

as N and $T\to\infty$ jointly in such a way that $T/N\to c$ with $0<c<\infty$, where $\hat{\theta}_{IV}$ is defined in (13).

Now we turn our attention to the asymptotic normality properties of the estimator. For this we require that the last term of (22), $O_p\left(\sqrt{NT}\delta_{NT}^{-2}\right)=O_p\left(\min\left\{\sqrt{\frac{T}{N}},\sqrt{\frac{N}{T}}\right\}\right)$, is asymptotically negligible. The condition that this term goes to zero asymptotically is $\min\left\{\frac{T}{N},\frac{N}{T}\right\}\to 0$ as $N\to\infty$ and $T\to\infty$ jointly. This is satisfied, for example, when $T/N=\min\left\{\frac{T}{N},\frac{N}{T}\right\}$, $T=bN^{1-\delta}$ for any finite positive constants b and δ. This is more stringent than the condition in Theorem 1, in that it does not allow T/N to converge to some positive finite constant, however, it permits many combinations of N and T. The results are summarised in the following theorem:

Theorem 2 *Suppose that Assumptions 1-5 hold true under model (1)-(3). Then,*
(i) $\min\left\{\frac{T}{N},\frac{N}{T}\right\}\to 0$ *as $N\to\infty$ and $T\to\infty$ jointly,*

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

where

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

is a positive definite matrix, with $\mathbf{A}$, $\mathbf{B}$, and $\boldsymbol{\Omega}$ defined in Assumption 5.
(ii) $\hat{\mathbf{Q}}_{NT}-\mathbf{Q}\xrightarrow{p}\mathbf{0}$ as N and T go to infinity when $\min\left\{\frac{T}{N},\frac{N}{T}\right\}\to 0$, *where $\hat{\mathbf{Q}}_{NT}$ is defined by (15).*

Proof. See Appendix. ■

Remark 3 *Theorem 2(ii), $\hat{\mathbf{Q}}_{NT}-\mathbf{Q}\xrightarrow{p}\mathbf{0}$, holds when $N\to\infty$ and $T\to\infty$ jointly in such a way that $N/T\to c$, $0<c<\infty$. See Appendix for a proof.*

Remark 4 *Similarly to the approach followed in Bai (2009), it is possible to derive the asymptotic bias of* $\sqrt{NT}(\hat{\theta}_{IV} - \theta)$ *assuming that* T/N *tends to a finite positive constant* c $(0 < c < \infty)$ *as* N *and* $T \to \infty$ *jointly, and then propose a bias-corrected estimator. Alternatively, bootstrap bias-correction of* $\hat{\theta}_{IV}$ *and bootstrapping the associated test statistic may largely cure the potential problem. However, as it will be shown later, the potential bias seems almost negligible in finite samples.*

Define the two-step estimator and the associated overidentifying restrictions test statistic as

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

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

where $\Delta\ddot{\mathbf{u}}_i = \Delta\mathbf{y}_i - \Delta\mathbf{W}_i\ddot{\theta}_{IV}$. Hansen (2007) shows that the t-test based on the variance estimator (15), in the context of a standard panel fixed effects estimation, is asymptotically valid even when T and N tend jointly to infinity. By a similar discussion, the asymptotic validity of the two-step estimator and the associated overidentifying restrictions test can be verified. The result is summarised in the following theorem:

Theorem 3 *Suppose that Assumptions 1-5 hold true under model (1)-(3). Then,* $\min\left\{\frac{T}{N}, \frac{N}{T}\right\} \to 0$ *as* $N \to \infty$ *and* $T \to \infty$ *jointly,*

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

for $k > 1$*, under the null hypothesis of strong exogeneity of the covariates, where* S_{NT} *is defined in (24).*

Proof. See Appendix. ■

The overidentifying restrictions test is particularly useful in our approach in order to test the assumption of strong exogeneity of the regressors with respect to the idiosyncratic errors in the equation for x, which is stated in Assumption 1(ii).

3 Monte Carlo Experiments

In this section we investigate the finite sample behaviour of the proposed estimator by means of Monte Carlo experiments. In particular, we study its bias and root mean square error (RMSE), as well as the size and power of the t-tests. Furthermore, we examine the finite sample performance of the overidentifying restrictions test. In the experiments we allow the case in which only m_x factors enter in the x equation, which are subset of m factors in the equation for y.

In order to investigate the relative performance of our estimator, four additional IV estimators are considered. All the estimators can be described in terms of the equation (13) of $\hat{\theta}_{IV}$, by redefining the matrix of instruments in (12). The associated variance estimator is also redefined in the same manner using (15). The first estimator is a variant of the popular Anderson-Hsiao estimator (Anderson and Hsiao, 1981, 1982), which is generally invalid asymptotically under a factor structure. The transformed matrix of instruments in (12) is redefined as

$$\tilde{\mathbf{Z}}_i = (\Delta\mathbf{y}_{i,-2}, \mathbf{\Delta X}_i, \mathbf{\Delta X}_{i,-1}), \tag{26}$$

where $\mathbf{y}_{i,-2} = L^2\mathbf{y}_i$.[10] The second estimator is the one proposed by Sarafidis, Yamagata, and Robertson (2009), which uses as instruments the untransformed regressors, assuming they are strictly exogenous,

$$\tilde{\mathbf{Z}}_i = \mathbf{Z}_i = (\mathbf{\Delta X}_i, \mathbf{\Delta X}_{i,-1}). \tag{27}$$

The third estimator defactors $\mathbf{Z}_i$ using cross-sectional averages instead of principal components, as proposed by Pesaran (2006),

$$\tilde{\mathbf{Z}}_i = \mathbf{M}_{\Delta\bar{X}}\mathbf{Z}_i \tag{28}$$

where $\mathbf{M}_{\Delta\bar{X}} = \mathbf{I}_T - \mathbf{\Delta\bar{X}}\left(\mathbf{\Delta\bar{X}}'\mathbf{\Delta\bar{X}}\right)^{-1}\mathbf{\Delta\bar{X}}'$ with $\mathbf{\Delta\bar{X}} = N^{-1}\sum_{i=1}^{N}\mathbf{\Delta X}_i$. The main advantage of this approach is that consistent estimation of the number of factors is not required. The disadvantage is that $\mathbf{M}_{\Delta\bar{X}}\mathbf{Z}_i$ is not a valid instrument when the rank condition on the factor loadings is not satisfied, namely $rank[E(\mathbf{\Gamma}_{xi})] \geq m_x$, where m_x is the number of factors in $\mathbf{X}_i$. Under the assumption that the factor loadings have zero mean (see Assumption 4), the rank condition is violated. As pointed out earlier, the cross section average of the factor loadings can be made zero by including time dummies in the estimating model.

The fourth estimator is our proposed one, which uses the instruments $\tilde{\mathbf{Z}}_i = \mathbf{M}_{\Delta\hat{F}}\mathbf{Z}_i$, where $\mathbf{M}_{\Delta\hat{F}}$ is the projection matrix of $\Delta\hat{\mathbf{F}}$, a $T \times \hat{m}_x$ matrix extracted from $\{\Delta\mathbf{X}_i\}_{i=1}^{N}$ using principal components, where $\hat{m}_x$ is the estimated number of factors. In the experiments the number of factors is estimated by the information criterion IC_1 proposed by Bai and Ng (2002). The maximum number of factors is set to $m_x + 1$.

Finally, an infeasible estimator is included as a benchmark, with instrument matrix equal to

$$\tilde{\mathbf{Z}}_i = \mathbf{M}_{\Delta F}\mathbf{Z}_i, \tag{29}$$

where $\mathbf{M}_{\Delta F}$ is the projection matrix of $\Delta\mathbf{F}$, a $T \times m_x$ matrix of factors in $\Delta\mathbf{X}_i$. Observe that our estimator is less efficient than the infeasible estimator, since the former estimates both the number of factors m_x as well as the factors themselves, $\Delta\mathbf{F}$.

3.1 Design

Consider a data generating process (DGP) with non-normal and time series and cross-sectionally heteroskedastic errors

$$y_{it} = \alpha_i + \lambda y_{it-1} + \sum_{\ell=1}^{k}\beta_\ell x_{\ell it} + \sum_{s=1}^{m}\gamma_{si}f_{st} + \varepsilon_{it},\ i = 1, 2, ..., N; t = -49, -48, ..., T, \tag{30}$$

where $\beta_\ell = (1-\lambda)/k$, $\alpha_i \sim iidN(0,1)$, $\gamma_{si} \sim iidN(0,1)$, $\varepsilon_{it} = \sigma_{it}(\epsilon_{it} - 1)/\sqrt{2}$, $\epsilon_{it} \sim iid\chi_1^2$, with $\sigma_{it}^2 = \eta_i\varphi_t$, $\eta_i \sim iid\chi_2^2/2$, and $\varphi_t = 1 - 0.01(T/2 + t)$ for $t = -1, 0, ..., T$ and unity otherwise. The covariates follow a multi-factor structure

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

[10] It might be reasonable to assume that $\Delta\mathbf{X}_{i,-2}$ is also available, though, for comparison purpose we did not include it as additional instruments.

$\ell = 1, 2, ..., k$ where $\mu_{\ell i} \sim iidN(0,1)$, and the factor loadings $\gamma_{\ell si}$ in the equation for $x_{\ell it}$ are correlated with those in the equation for y_{it}, such that

$$\gamma_{\ell si} = \rho_{\ell s}\gamma_{si} + (1-\rho_{\ell s}^2)^{1/2}\xi_{\ell si},\ \xi_{\ell si} \sim iidN(0,1),$$

$\ell = 1, 2$, $s = 1, ..., m$. The factors and the idiosyncratic errors of $x_{\ell it}$ are serially correlated such that

$$f_{st} = \rho_s f_{st-1} + (1-\rho_s^2)^{1/2}\zeta_{st},\ \zeta_{st} \sim iidN(0, 1/m),$$

so that $var(\sum_{s=1}^{m} f_{st}) = 1$ for any m, and

$$v_{\ell it} = \rho_\ell v_{\ell it-1} + (1-\rho_\ell^2)^{1/2}\varpi_{\ell it},\ \varpi_{\ell it} \sim iidN(0,1),$$

$\ell = 1, 2, ..., k$.

To consider the case in which only a subset of factors in the equation for y enters the x equation, we report the results for $m = 4$ and $m_x = 2$. In order to investigate the overidentified model ($2k > k + 1$), we report the results for $k = 2$. Also, we set $\rho_\ell = 0.8$, $\rho_s = 0.4$, $\rho_{\ell s} = 0.4$.[11] We examine $\lambda = 0.2, 0.5, 0.8$ with $\beta_\ell = (1-\lambda)/k$ for $\ell = 1, 2, ..., k$, so that $(1-\lambda)^{-1}\sum_{\ell=1}^{k}\beta_\ell = 1$. We consider several combinations of (T, N), in specific, $T \in \{10, 20, 50, 100, 200\}$ and $N \in \{10, 20, 50, 100, 200\}$. The results are obtained based on 2000 replications, and all tests are conducted at the 5% significance level.

3.2 Results

Tables 1 and 2 report the mean value of the estimated coefficient of the lagged dependent variable, $\hat{\lambda}$, and of x_{1it}, $\hat{\beta}_1$.[12] The IV estimator that makes use of $\mathbf{Z}_i = (\mathbf{\Delta X}_i, \mathbf{\Delta X}_{i,-1})$ only is highly biased. This is because the instruments are not orthogonal to the composite error term. Including the lagged dependent variable $\Delta\mathbf{y}_{i,-2}$ as an instrument appears to increase the bias of the estimator. Furthermore, projecting out $\Delta\bar{\mathbf{X}}$ from $\mathbf{Z}_i$ by premutiplying it by $\mathbf{M}_{\Delta\bar{X}}$ does not work. The reason for this result is that the rank condition is violated in our case because the mean value of the factor loadings is zero. In contrast, our proposed estimator, which uses $\mathbf{M}_{\Delta\hat{F}}\mathbf{Z}_i$ as instruments, has little bias for different values of λ and β_1 (including $\lambda = 0.8$). In fact, the bias appears to be very similar to that of the infeasible IV estimator, which makes use of the unobservable true factors. As a result, while the bias of our estimator is non-negligible when N is small, for $N \geq 50$ and $T \geq 10$ the bias lies within ± 0.001 in most of the cases.

Tables 3 and 4 report the root mean square error (RMSEs) of $\hat{\lambda}$ and $\hat{\beta}_1$. Except when $T = 10$ or $N = 10$, the RMSE of the proposed estimator is very similar to the infeasible estimator and the difference between the two is mostly within ± 0.001. This may be remarkable considering that our estimator is subject to the extra uncertainty arising from the fact that both the number of factors and the factors themselves are unknown and estimated in the model.

Tables 5 and 6 provide the estimated size of the t-test. The size of t-test based on our approach is very close to the nominal level (5%), especially for $N \geq 50$ and $T \geq 10$. It is

[11] We considered other values of k, m, m_x (including $m = m_x$), ρ_ℓ, ρ_s, $\rho_{\ell s}$, as well as errors, ε_{it}, drawn from normal distribution with other heteroskedastic schemes (including homoskedasticity). These results, which are available upon request from the authors, confirm that the presented satisfactory results of our proposed estimator and associated tests are robust to varieties of experimental design.

[12] The results for the estimates of β_2 are very similar to those of β_1, which are not reported, but are available upon request from the authors.

also worth noting that the estimated size of the t-test based on our proposed estimator is similar to that based on the infeasible estimator. The reasonable power of the t-test based on our estimator is confirmed in Tables 7 and 8.

Table 9 reports the results of the overidentifying restrictions test. The size of our test is correct in most of the combinations of N and T. Importantly the test has good power when the idiosyncratic errors in x_{1it} are correlated with the idiosyncratic errors in y_{it}. Thus, this test can be a reliable statistical tool to check the key assumption of our approach.

4 Concluding Remarks

This paper has proposed a computationally attractive instrumental-variable procedure for consistent estimation of dynamic linear panel data models with error cross-sectional dependence when both N and T are large. Our approach involves projecting out the common factors in the regressors at first stage, and then using the defactored regressors as instruments for the endogenous variables. Aside from computational simplicity the method has the advantage that it does not require estimating possible distinct factors that enter directly only into the y process, thus leaving these factors in the residuals. Therefore, full specification of the model is not required. In practice, it is also possible that (a subset of) the factors that hit the covariates are orthogonal to the composite disturbance of the y process. In this case, full defactoring is not necessary for consistency of the IV etimator because instrument exogeneity merely requires projecting out the common components which are correlated with the factors that enter directly into the y process. Empirically, this issue can be addressed using a sequential testing method based on the overidentifying restrictions test that we have explored in this paper. In particular, one may start by testing whether the untransformed covariates are strongly exogenous with respect to the composite dusturbance. Notice that the null hypothesis will also be satisfied if the covariates do not have a factor structure at all. If the null is rejected, one may project out the factor corresponding to the largest eigenvalue of the $T \times T$ matrix $\sum_{i=1}^{N} \Delta \mathbf{X}_i \Delta \mathbf{X}_i'$ and test whether the defactored regressors yield valid instruments using the same statistic. If the null is rejected, one may project out two factors, in particular those associated with the two largest eigenvalues of $\sum_{i=1}^{N} \Delta \mathbf{X}_i \Delta \mathbf{X}_i'$, and so on. Naturally, the significance level used for this sequential method needs to be appropriately adjusted. The interested reader is recommended to refer to Ahn, Lee, and Schmidt (2006), Proposition 2.

Finally, notice that although the proofs of our results require N and T both large, under certain restrictions imposed in the covariates − in particular, asymptotic homoskedasticity and serial uncorrelatedness − it is possible to derive consistency and asymptotic normality of our estimator even for T fixed; see Bai (2003). On the other hand, the simulation evidence we have presented suggests that even if these conditions are not met in practice, the bias of the estimator can be practically negligible and the size of the t-test is correct for (T, N) as small as $(10, 50)$. Therefore, we hope that our approach provides a computationally attractive way to estimate dynamic panel data models with multi-factor residual structures, even in cases where T is moderately small.

Appendix: Mathematical Proofs

Lemma 1 *From Bai (2003) we have*

$$p\lim_{T,N\to\infty}\frac{\Delta\hat{\mathbf{F}}'\Delta\mathbf{F}}{T}=\mathbf{G} \tag{A.1}$$

By Lemmas B1, B2 and B3 of Bai (2003),

$$T^{-1}(\Delta\hat{\boldsymbol{F}}-\boldsymbol{\Delta F})'\Delta\boldsymbol{\varepsilon}_i=\boldsymbol{O}_p\,(\delta_{NT}^{-2}), \tag{A.2}$$

$$T^{-1}(\Delta\hat{\boldsymbol{F}}-\boldsymbol{\Delta F})'\Delta\boldsymbol{F}=\boldsymbol{O}_p(\delta_{NT}^{-2}) \tag{A.3}$$

and

$$T^{-1}(\Delta\hat{\boldsymbol{F}}-\boldsymbol{\Delta F})'\Delta\hat{\boldsymbol{F}}=\boldsymbol{O}_p(\delta_{NT}^{-2}), \tag{A.4}$$

where $\delta_{NT}=\min\{\sqrt{N},\sqrt{T}\}$.

Lemma 2 *Under Assumptions 1-5 we have*

$$\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta\hat{F}}\Delta\boldsymbol{W}_i}{T}-\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{W}_i}{T}=O_p\left(\delta_{NT}^{-2}\right),\ \textit{uniformly over } i, \tag{A.5}$$

$$\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta\hat{F}}\Delta\boldsymbol{F}}{T}-\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{F}}{T}=O_p\left(\delta_{NT}^{-2}\right),\ \textit{uniformly over } i \tag{A.6}$$

and

$$\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta\hat{F}}\Delta\boldsymbol{\varepsilon}_i}{T}-\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{\varepsilon}_i}{T}=O_p\left(\delta_{NT}^{-2}\right),\ \textit{uniformly over } i. \tag{A.7}$$

Proof. We start by proving (A.5). We need to determine the order of probability of $\left\|\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta\hat{F}}\Delta\boldsymbol{X}_i}{T}-\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{X}_i}{T}\right\|$. This is equal to

$$\left\|\frac{\Delta\boldsymbol{X}_i'\Delta\hat{\boldsymbol{F}}\left(\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}\right)^{-1}\Delta\hat{\boldsymbol{F}}'\boldsymbol{\Delta X}_i}{T}-\frac{\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\left(\Delta\boldsymbol{F}'\Delta\boldsymbol{F}\right)^{-1}\Delta\boldsymbol{F}'\Delta\boldsymbol{X}_i}{T}\right\|=$$

$$\frac{1}{T}\left\|\Delta\boldsymbol{X}_i'\Delta\hat{\boldsymbol{F}}\left(\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}\right)^{-1}\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i-\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\left(\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}\right)^{-1}\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i+\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\left(\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}\right)^{-1}\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i-\right.$$
$$\left.\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\left(\boldsymbol{\Delta F'\Delta F}\right)^{-1}\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i+\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\left(\boldsymbol{\Delta F'\Delta F}\right)^{-\mathbf{1}}\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i-\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\left(\Delta\boldsymbol{F}'\Delta\boldsymbol{F}\right)^{-1}\Delta\boldsymbol{F}'\Delta\boldsymbol{X}_i\right\|$$

$$=\frac{1}{T}\left\|\left(\Delta\boldsymbol{X}_i'\Delta\,\hat{\boldsymbol{F}}-\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\right)\left(\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}\right)^{-1}\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i+\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\left(\boldsymbol{\Delta F'\Delta F}\right)^{-\mathbf{1}}\left(\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i-\Delta\boldsymbol{F}'\Delta\boldsymbol{X}_i\right)+\right.$$
$$\left.\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\left(\left(\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}\right)^{-1}-\left(\boldsymbol{\Delta F'\Delta F}\right)^{-1}\right)\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i\right\|$$

$$\leq\left\|\frac{1}{T}\left(\Delta\boldsymbol{X}_i'\Delta\,\hat{\boldsymbol{F}}-\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\right)\left(\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}\right)^{-1}\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i\right\|+\left\|\frac{1}{T}\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\left(\left(\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}\right)^{-1}-\left(\boldsymbol{\Delta F'\Delta F}\right)^{-1}\right)\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i\right\|+$$
$$\left\|\frac{1}{T}\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\left(\boldsymbol{\Delta F'\Delta F}\right)^{-1}\left(\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i-\Delta\boldsymbol{F}'\Delta\boldsymbol{X}_i\right)\right\|. \tag{A.8}$$

We examine each of the above terms.

$$\begin{aligned}\left\|\frac{1}{T}\left(\Delta\boldsymbol{X}_i'\Delta\,\hat{\boldsymbol{F}}-\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\right)\left(\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}\right)^{-1}\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i\right\|&=\left\|\frac{1}{T}\left(\Delta\boldsymbol{X}_i'\Delta\,\hat{\boldsymbol{F}}-\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\right)\left(\frac{\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}}{T}\right)^{-1}\frac{\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i}{T}\right\|\\&\leq\left\|\frac{\Delta\boldsymbol{X}_i'\left(\Delta\hat{\boldsymbol{F}}-\boldsymbol{\Delta F}\right)}{T}\right\|\left\|\left(\frac{\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}}{T}\right)^{-1}\frac{\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i}{T}\right\|\\&=O_p\left(\delta_{NT}^{-2}\right),\quad\text{uniformly over } i,\end{aligned} \tag{A.9}$$

by (A.1) to (A.4). Next, we have

$$\left\|\frac{1}{T}\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\left(\left(\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}\right)^{-1}-(\boldsymbol{\Delta F'\Delta F})^{-1}\right)\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i\right\|$$
$$=\left\|\frac{\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}}{T}\left(\frac{\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}}{T}\right)^{-1}\frac{1}{T}\left((\boldsymbol{\Delta F'\Delta F})-\left(\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}\right)\right)\left(\frac{\Delta\boldsymbol{F}'\Delta\boldsymbol{F}}{T}\right)^{-1}\frac{\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i}{T}\right\|$$
$$\leq\left\|\frac{\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}}{T}-\frac{\Delta\boldsymbol{F}'\Delta\boldsymbol{F}}{T}\right\|\left\|\frac{\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}}{T}\left(\frac{\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}}{T}\right)^{-1}\right\|\left\|\left(\frac{\Delta\boldsymbol{F}'\Delta\boldsymbol{F}}{T}\right)^{-1}\frac{\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i}{T}\right\|.$$

Since

$$\begin{aligned}\frac{\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}}{T}&=\frac{1}{T}(\Delta\boldsymbol{F}+(\boldsymbol{\Delta}\hat{\boldsymbol{F}}-\boldsymbol{\Delta F}))'(\boldsymbol{\Delta F}+(\boldsymbol{\Delta}\hat{\boldsymbol{F}}-\boldsymbol{\Delta F})) \qquad (A.10)\\&=\frac{\Delta\boldsymbol{F}'\Delta\boldsymbol{F}}{T}+\frac{\Delta\boldsymbol{F}'(\boldsymbol{\Delta}\hat{\boldsymbol{F}}-\boldsymbol{\Delta F})}{T}+\frac{(\boldsymbol{\Delta}\hat{\boldsymbol{F}}-\boldsymbol{\Delta F})'\Delta\boldsymbol{F}}{T}+\frac{(\boldsymbol{\Delta}\hat{\boldsymbol{F}}-\boldsymbol{\Delta F})'(\boldsymbol{\Delta}\hat{\boldsymbol{F}}-\boldsymbol{\Delta F})}{T}\\&=\frac{\Delta\boldsymbol{F}'\Delta\boldsymbol{F}}{T}+O_p(\delta_{NT}^{-2})\end{aligned}$$

by (A.1) to (A.4), which leads to

$$\left\|\frac{1}{T}\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\left(\left(\Delta\hat{\boldsymbol{F}}'\Delta\hat{\boldsymbol{F}}\right)^{-1}-(\boldsymbol{\Delta F'\Delta F})^{-1}\right)\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i\right\|=O_p(\delta_{NT}^{-2}). \qquad (A.11)$$

Finally,

$$\left\|\frac{1}{T}\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}\left(\boldsymbol{\Delta F'\Delta F}\right)^{-1}\left(\Delta\hat{\boldsymbol{F}}'\Delta\boldsymbol{X}_i-\Delta\boldsymbol{F}'\Delta\boldsymbol{X}_i\right)\right\|\leq\left\|\frac{\Delta\boldsymbol{X}_i'\Delta\boldsymbol{F}}{T}\left(\frac{\Delta\boldsymbol{F}'\Delta\boldsymbol{F}}{T}\right)^{-1}\right\|\left\|\frac{\Delta\boldsymbol{X}_i'(\boldsymbol{\Delta}\hat{\boldsymbol{F}}-\boldsymbol{\Delta F})}{T}\right\|=O_p(\delta_{NT}^{-2}). \qquad (A.12)$$

by (A.1) to (A.4). Substituting (A.9), (A.11) and (A.12) into (A.8), we have

$$\left\|\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta\hat{F}}\Delta\boldsymbol{X}_i}{T}-\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{X}_i}{T}\right\|=O_p\left(\delta_{NT}^{-2}\right),\ \text{uniformly over } i,$$

In a similar manner, it is straightforward to prove that

$$\left\|\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta\hat{F}}\Delta\boldsymbol{y}_{i,-1}}{T}-\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{y}_{i,-1}}{T}\right\|=O_p\left(\delta_{NT}^{-2}\right),\ \text{uniformly over } i,$$

and thus

$$\left\|\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta\hat{F}}\Delta\boldsymbol{W}_i}{T}-\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{W}_i}{T}\right\|=O_p\left(\delta_{NT}^{-2}\right),\ \text{uniformly over } i,$$

as required.

Likewise, it can be proved that

$$\left\|\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta\hat{F}}\Delta\boldsymbol{\varepsilon}_i}{T}-\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{\varepsilon}_i}{T}\right\|=O_p\left(\delta_{NT}^{-2}\right)\ \text{uniformly over } i$$

$$\left\|\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta\hat{F}}\Delta\boldsymbol{F}}{T}-\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{F}}{T}\right\|=\left\|\frac{\Delta\boldsymbol{X}_i'\boldsymbol{M}_{\Delta\hat{F}}\Delta\boldsymbol{F}}{T}\right\|=O_p\left(\delta_{NT}^{-2}\right)\ \text{uniformly over } i$$

Similar discussion holds for $\Delta\boldsymbol{X}_{i,-1}'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{W}_i/T$, $\Delta\boldsymbol{X}_{i,-1}'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{F}/T$, and $\Delta\boldsymbol{X}_{i,-1}'\boldsymbol{M}_{\Delta F}\Delta\boldsymbol{\varepsilon}_i/T$. Hence, we can obtain the results as required. ■

Lemma 3 $\hat{\boldsymbol{\Omega}}_{NT}-\boldsymbol{\Omega}\to\mathbf{0}$ *as* N *and* T *go to infinity in such a way that* $N/T\to c$, $0<c<\infty$.

Proof. Below, $o_p(1)$ term is asymptotically negligible when N and T go to infinity in such a way that $N/T \to c$, $0 < c < \infty$. Recall that

$$\hat{\mathbf{\Omega}}_{NT} = \frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathbf{Z}}_i'\Delta\hat{\mathbf{u}}_i\Delta\hat{\mathbf{u}}_i'\tilde{\mathbf{Z}}_i \tag{A.13}$$

with $\tilde{\mathbf{Z}}_i' = \mathbf{Z}_i'\mathbf{M}_{\Delta\hat{F}}$ and $\Delta\mathbf{u}_i = \Delta\mathbf{F}\gamma_i + \Delta\varepsilon_i$. Noting that $\mathbf{\Delta\hat{u}}_i = \Delta\mathbf{y}_i - \Delta\mathbf{W}_i\hat{\theta}_{IV} = \mathbf{\Delta u}_i - \Delta\mathbf{W}_i\left(\hat{\theta}_{IV} - \theta\right)$, we have

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\tilde{\mathbf{Z}}_i'\Delta\hat{\mathbf{u}}_i\Delta\hat{\mathbf{u}}_i'\tilde{\mathbf{Z}}_i &= \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\Delta\hat{F}}\Delta\mathbf{u}_i\Delta\mathbf{u}_i'\mathbf{M}_{\Delta\hat{F}}\mathbf{Z}_i \\
&\quad - \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\Delta\hat{F}}\Delta\mathbf{W}_i\left(\hat{\theta}_{IV} - \theta\right)\Delta\mathbf{u}_i'\mathbf{M}_{\Delta\hat{F}}\mathbf{Z}_i \\
&\quad - \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\Delta\hat{F}}\Delta\mathbf{u}_i\left(\hat{\theta}_{IV} - \theta\right)\Delta\mathbf{W}_i'\mathbf{M}_{\Delta\hat{F}}\mathbf{Z}_i \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\Delta\hat{F}}\Delta\mathbf{W}_i\left(\hat{\theta}_{IV} - \theta\right)\left(\hat{\theta}_{IV} - \theta\right)'\Delta\mathbf{W}_i'\mathbf{M}_{\Delta\hat{F}}\mathbf{Z}_i \\
&= I - II - III + IV.
\end{aligned}$$

First, look at II. This can be re-written as

$$\begin{aligned}
II &= \frac{T}{N}\sum_{i=1}^{N}\frac{\mathbf{Z}_i'\mathbf{M}_{\Delta\hat{F}}\Delta\mathbf{W}_i}{T}\left(\hat{\theta}_{IV} - \theta\right)\frac{\Delta\mathbf{u}_i'\mathbf{M}_{\Delta\hat{F}}\mathbf{Z}_i}{T} \\
&= \frac{T}{N}\sum_{i=1}^{N}\frac{\mathbf{Z}_i'\mathbf{M}_{\Delta F}\Delta\mathbf{W}_i}{T}\left(\hat{\theta}_{IV} - \theta\right)\frac{\Delta\varepsilon_i'\mathbf{M}_{\Delta F}\mathbf{Z}_i}{T} + \boldsymbol{O}_p\left(\sqrt{\frac{T}{N}}\right)\boldsymbol{O}_p(\delta_{NT}^{-2})
\end{aligned}$$

by (A.2), (A.4), and $\left(\hat{\theta}_{IV} - \theta\right) = \boldsymbol{O}_p\left(\sqrt{\frac{1}{TN}}\right)$. Since $\frac{\Delta\varepsilon_i'\mathbf{M}_{\Delta F}\mathbf{Z}_i}{T} = O_p\left(\frac{1}{\sqrt{T}}\right)$,

$$\begin{aligned}
II &= T\boldsymbol{O}_p(1)\,.\boldsymbol{O}_p\left(\sqrt{\frac{1}{TN}}\right).O_p\left(\frac{1}{\sqrt{T}}\right) + \boldsymbol{O}_p\left(\sqrt{\frac{T}{N}}\right)\boldsymbol{O}_p(\delta_{NT}^{-2}) \\
&= o_p(1).
\end{aligned}$$

A similar discussion yields $III = o_p(1)$. Furthermore,

$$\begin{aligned}
IV &= \frac{T}{N}\sum_{i=1}^{N}\frac{\mathbf{Z}_i'\mathbf{M}_{\Delta\hat{F}}\Delta\mathbf{W}_i}{T}\left(\hat{\theta}_{IV} - \theta\right)\frac{\Delta\mathbf{W}_i'\mathbf{M}_{\Delta\hat{F}}\mathbf{Z}_i}{T} \\
&= \frac{T}{N}\sum_{i=1}^{N}\frac{\mathbf{Z}_i'\mathbf{M}_{\Delta F}\Delta\mathbf{W}_i}{T}\left(\hat{\theta}_{IV} - \theta\right)\left(\hat{\theta}_{IV} - \theta\right)'\frac{\Delta\mathbf{W}_i'\mathbf{M}_{\Delta F}\mathbf{Z}_i}{T} + \boldsymbol{O}_p\left(\frac{1}{N}\right)\boldsymbol{O}_p(\delta_{NT}^{-2}) \\
&= \boldsymbol{o}_p(1).
\end{aligned}$$

Now,

$$\begin{aligned}
I &= \frac{T}{N}\sum_{i=1}^{N}\frac{\mathbf{Z}_i'\mathbf{M}_{\Delta\hat{F}}\Delta\mathbf{u}_i}{T}\frac{\Delta\mathbf{u}_i'\mathbf{M}_{\Delta\hat{F}}\mathbf{Z}_i}{T} \\
&= \frac{1}{TN}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\Delta F}\Delta\varepsilon_i\Delta\varepsilon_i'\mathbf{M}_{\Delta F}\mathbf{Z}_i + o_p(1)
\end{aligned}$$

by (A.2)-(A.4). By Assumption 5 (ii) and a law of large numbers,

$$\frac{1}{TN}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{M}_{\Delta F}\Delta\varepsilon_i\Delta\varepsilon_i'\mathbf{M}_{\Delta F}\mathbf{Z}_i - \Omega = o_p(1).$$

In sum, $\hat{\mathbf{\Omega}}_{NT} - \mathbf{\Omega} = o_p(1)$. As it is easily seen, the result holds when $\min\left\{\frac{T}{N}, \frac{N}{T}\right\} \to 0$ as $N \to \infty$ and $T \to \infty$ jointly, as required. ■

Proof of Theorem 2. (i) Using Lemma 2 and a LLN, $\text{plim}_{N,T\to\infty}\mathbf{A}_{NT} = \mathbf{A}$ and $\text{plim}_{N,T\to\infty}\mathbf{B}_{NT} = \mathbf{B}$. By Assumptions 1-5 and a LLN, $\text{plim}_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{Z}_i'\boldsymbol{M}_{\Delta F}\Delta\varepsilon_i\Delta\varepsilon_i\boldsymbol{M}_{\Delta F}\mathbf{Z}_i = \mathbf{\Omega}$. Applying Lemma 2 of Hansen (2007) yields $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{Z}_i'\boldsymbol{M}_{\Delta F}\Delta\varepsilon_i \xrightarrow{d} N\left(\mathbf{0}, \mathbf{\Omega}\right)$ (the proof of the lemma is provided in the Technical Appendix of Hansen (2007)), as required.

(ii) Under Assumptions 1-5 and using Lemma 2, Lemma 3 and a LLN, the result follows. ■

Proof of Theorem 3. Under Assumptions 1-5, together with the $\sqrt{NT}$ consistency of $\hat{\boldsymbol{\theta}}_{IV}$ of Theorem 1 and a LLN, by Lemma 3 we have $\hat{\mathbf{\Omega}}_{NT} - \mathbf{\Omega} \to \mathbf{0}$ as $N \to \infty$ and $T \to \infty$ jointly in such a way that T/N tends to a finite positive constant. The consistency of $\hat{\mathbf{\Omega}}_{NT}$ leads to the $\sqrt{NT}$ consistency of $\ddot{\theta}_{IV}$, and therefore under the null hypothesis a similar discussion for Theorem 2 yields $\frac{1}{\sqrt{NT}}\hat{\mathbf{\Omega}}^{-1/2}\sum_{i=1}^{N}\tilde{\mathbf{Z}}_i'\Delta\ddot{\mathbf{u}}_i \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{2k})$, when $\min\left\{\frac{T}{N}, \frac{N}{T}\right\} \to 0$ as $N \to \infty$ and $T \to \infty$ jointly. Finally, applying a standard proof for the asymptotic distribution of the overidentifying restrictions test under the null hypothesis, such as in Arellano (2003), will yield the desired result. ■

References

Ahn, S., Y. Lee, and P. Schmidt (2001): "GMM Estimation of Linear Panel Data Models with Time-Varying Individual Effects," *Journal of Econometrics*, 101, 219–255.

——— (2006): "Panel Data Models with Multiple Time Varying Effects," *mimeo, Arizona State University.*

Anderson, T., and C. Hsiao (1981): "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76, 598–606.

——— (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

Arellano, M. (2003): *Panel Data Econometrics.* Oxford University Press, Oxford.

Bai, J. (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–173.

——— (2009): "Panel Data Models with Interactive Fixed Effects," *Econometrica*, 77, 1229–1279.

——— (2010): "Likelihood Approach to Small T Dynamic Panel Data Models with Interactive Effects," *mimeo.*

Bai, J., and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 1235–1267.

Becker, G. S., M. Grossman, and K. Murphy (1994): "An Empirical Analysis of Cigarette Addiction," *The American Economic Review*, 84, 396–418.

Blundell, R., and S. Bond (2000): "GMM Estimation with Persistent Panel Data: An Application to Production Functions," *Econometric Reviews*, 19, 321–340.

Bover, O. (1991): "Relaxing Intertemporal Separability: A Rational Habits Model of Labor Supply Estimated from Panel Data," *Journal of Labour Economics*, 9, 85–100.

Carneiro, P., K. Hansen, and J. Heckman (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.*

Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5–46.

Favero, C., M. Marcellino, and F. Neglia (2005): "Principal Components at Work: The Empirical Analysis of Monetary Policy with Large Data Sets," *Journal of Applied Econometrics*, 20, 603–620.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000): "The Generalized Dynamic Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82, 540–554.

Hansen, C. (2007): "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data When T is Large," *Journal of Econometrics*, 141, 597–620.

Heckman, J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labour Economics*, 24, 411–482.

Kapoor, M., H. Kelejian, and I. Prucha (2007): "Panel Data Models with Spatially Correlated Error Components," *Journal of Econometrics*, 140, 97–130.

Kelejian, H., and I. Prucha (2010): "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances," *Journal of Econometrics*, forthcoming.

Lee, L. (2004): "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models," *Econometrica*, 72, 1899–1925.

Pesaran, M. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74, 339–390.

Robertson, D., V. Sarafidis, and J. Symons (2010): "IV Estimation of Panels with Factor Residuals," *mimeo*.

Sarafidis, V., T. Yamagata, and D. Robertson (2009): "A Test of Cross Section Dependence for a Linear Dynamic Panel Model with Regressors," *Journal of Econometrics*, 148, 149–161.

Wooldridge, J. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Table 1: Mean Value of the Estimated Coefficient of the Lagged Dependent Variable in ARDL(1,0) models. The Factor loadings in the X equation, which contains two factors, are correlated with those in the Y equation, where the number of factors is four. The mean value of the factor loadings is zero.

Instruments	$\lambda = 0.2$					$\lambda = 0.5$					$\lambda = 0.8$				
$\Delta \mathbf{y}_{i,-2}, \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.226	0.235	0.234	0.237	0.241	0.519	0.536	0.537	0.543	0.550	0.683	0.757	0.788	0.804	0.818
20	0.231	0.236	0.240	0.240	0.239	0.532	0.541	0.548	0.549	0.547	0.760	0.797	0.815	0.820	0.816
50	0.236	0.237	0.240	0.240	0.239	0.542	0.546	0.550	0.551	0.550	0.803	0.818	0.827	0.827	0.824
100	0.239	0.240	0.241	0.240	0.239	0.549	0.550	0.552	0.550	0.549	0.822	0.828	0.829	0.826	0.823
200	0.240	0.240	0.240	0.240	0.239	0.551	0.551	0.551	0.551	0.550	0.831	0.829	0.828	0.827	0.825
$\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.170	0.171	0.169	0.170	0.173	0.440	0.453	0.450	0.453	0.457	0.586	0.669	0.697	0.705	0.712
20	0.166	0.171	0.172	0.171	0.170	0.445	0.454	0.455	0.455	0.454	0.633	0.695	0.711	0.710	0.708
50	0.169	0.170	0.172	0.172	0.170	0.452	0.454	0.456	0.457	0.454	0.699	0.709	0.713	0.716	0.709
100	0.170	0.172	0.172	0.170	0.170	0.453	0.456	0.457	0.454	0.454	0.704	0.713	0.715	0.710	0.709
200	0.172	0.171	0.171	0.171	0.170	0.456	0.455	0.455	0.455	0.454	0.713	0.710	0.712	0.712	0.710
$\mathbf{M}_{\Delta \bar{X}} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.177	0.178	0.173	0.176	0.179	0.448	0.461	0.457	0.463	0.467	0.587	0.673	0.705	0.722	0.736
20	0.170	0.177	0.179	0.179	0.177	0.454	0.464	0.467	0.467	0.465	0.664	0.698	0.734	0.735	0.731
50	0.175	0.177	0.179	0.178	0.177	0.461	0.464	0.467	0.466	0.464	0.725	0.728	0.734	0.733	0.729
100	0.177	0.179	0.179	0.177	0.177	0.464	0.468	0.467	0.464	0.464	0.726	0.736	0.735	0.729	0.729
200	0.179	0.177	0.177	0.177	0.177	0.468	0.465	0.465	0.465	0.464	0.735	0.730	0.731	0.731	0.730
$\mathbf{M}_{\Delta \hat{F}} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.190	0.200	0.196	0.198	0.203	0.473	0.489	0.493	0.497	0.504	0.611	0.668	0.780	0.789	0.806
20	0.188	0.199	0.201	0.200	0.200	0.480	0.498	0.501	0.501	0.500	0.688	0.785	0.804	0.801	0.800
50	0.193	0.196	0.200	0.200	0.199	0.488	0.494	0.499	0.500	0.499	0.767	0.787	0.799	0.800	0.798
100	0.195	0.199	0.200	0.199	0.199	0.493	0.498	0.500	0.499	0.499	0.784	0.797	0.801	0.797	0.798
200	0.197	0.198	0.199	0.200	0.199	0.495	0.496	0.499	0.499	0.499	0.790	0.792	0.797	0.799	0.798
$\mathbf{M}_{\Delta F} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.196	0.203	0.195	0.197	0.201	0.482	0.503	0.493	0.496	0.501	0.586	0.628	0.779	0.788	0.801
20	0.194	0.201	0.201	0.200	0.199	0.490	0.501	0.502	0.500	0.499	0.737	0.792	0.806	0.801	0.798
50	0.198	0.198	0.200	0.200	0.199	0.497	0.498	0.501	0.500	0.499	0.784	0.793	0.801	0.801	0.798
100	0.201	0.201	0.201	0.200	0.199	0.501	0.502	0.502	0.499	0.499	0.802	0.805	0.804	0.799	0.798
200	0.202	0.200	0.200	0.200	0.200	0.503	0.500	0.500	0.500	0.499	0.806	0.800	0.800	0.800	0.799

Notes: The DGP follows $y_{it} = \alpha_i + \lambda y_{it-1} + \sum_{\ell=1}^{k} \beta_\ell x_{\ell it} + \sum_{s=1}^{m} \gamma_{si} f_{st} + \varepsilon_{it}$ with $k = 2$ and $m = 4$, where $\beta_\ell = (1-\lambda)/k$, $\alpha_i \sim iidN(0,1), \gamma_{si} \sim iidN(0,1), \varepsilon_{it} = \sigma_{it}(\epsilon_{it} - 1)/2^{1/2}, \epsilon_{it} \sim iid\chi_1^2$, with $\sigma_{it}^2 = \eta_i \varphi_t$, $\eta_i \sim iid\chi_2^2/2$, $\varphi_t = 1 - 0.01(T/2 + t)$ for $t = -1, 0, ..., T$, otherwise unity, $x_{\ell it} = \mu_{\ell i} + \sum_{s=1}^{m_x} \gamma_{\ell si} f_{st} + v_{\ell it}$ with $m_x = 2$, $\ell = 1, 2, ..., k$ where $\mu_{\ell i} \sim iidN(0,1)$, and the factor loadings, $\gamma_{\ell si}$, in $x_{\ell it}$ are correlated with those in y_{it} such that $\gamma_{\ell si} = \rho_{\ell s}\gamma_{si} + (1-\rho_{\ell s}^2)^{1/2}\xi_{si}$, $\xi_{si} \sim iidN(0,1)$ with $\rho_{\ell s} = 0.4$, $\ell = 1, 2$, $s = 1, ..., 4$, $f_{st} = \rho_s f_{st-1} + (1-\rho_s^2)^{1/2}\zeta_{st}$, $\zeta_{st} \sim iidN(0, 1/m)$ with $\rho_s = 0.4$.so that $var(\sum_{s=1}^{m} f_{st}) = 1$ for any m, $v_{\ell it} = \rho_\ell v_{\ell it-1} + (1-\rho_\ell^2)^{1/2}\varpi_{\ell it}, \varpi_{\ell it} \sim iidN(0,1)$ with $\rho_\ell = 0.8$, $\ell = 1, 2, ..., k$. The first column describes the instruments used for the IV estimation of the model; $\mathbf{Z}_i = (\mathbf{\Delta X}_i, \mathbf{\Delta X}_{i,-1})$ a $T \times 2k$ matrix, where $\mathbf{\Delta X}_i = (\Delta \mathbf{x}_{i1}, ..., \Delta \mathbf{x}_{iT})'$ with $\Delta = 1 - L$ and $\Delta \mathbf{x}_{it} = (\Delta x_{1it}, ..., \Delta x_{kit})'$, L is the lag operator and $\mathbf{\Delta X}_{i,-1} = (\Delta \mathbf{x}_{i0}, ..., \Delta \mathbf{x}_{iT-1})'$. Also, $\mathbf{\Delta y}_{i,-2} = (\Delta y_{i,-1}, ..., \Delta y_{iT-2})'$, $\mathbf{M}_{\Delta \bar{X}} = \mathbf{I}_T - \mathbf{\Delta \bar{X}}(\mathbf{\Delta \bar{X}}'\mathbf{\Delta \bar{X}})^{-1}\mathbf{\Delta \bar{X}}'$ with $\mathbf{\Delta \bar{X}} = N^{-1}\sum_{i=1}^{N} \mathbf{\Delta X}_i$, $\mathbf{M}_{\Delta \hat{F}} = \mathbf{I}_T - \Delta\hat{\mathbf{F}}(\Delta\hat{\mathbf{F}}'\Delta\hat{\mathbf{F}})^{-1}\Delta\hat{\mathbf{F}}'$, where $\Delta\hat{\mathbf{F}}$ is $T \times \hat{m}_x$ matrix computed from the principal component estimator extracted from $\{\Delta \mathbf{X}_i\}_{i=1}^{N}$, $\hat{m}_x$ is estimated by IC_1 in Bai and Ng (2002) with maximum number equal to three, $\mathbf{M}_{\Delta F} = \mathbf{I}_T - \Delta\mathbf{F}(\Delta\mathbf{F}'\Delta\mathbf{F})^{-1}\Delta\mathbf{F}'$, where $\Delta\mathbf{F}$ is a $T \times m_x$ matrix of true factors. All the experiments are based on 2000 replications.

Table 2: Mean Value of the Estimated Coefficient of the Regressor in ARDL(1,0) models. The Factor loadings in the X equation, which contains two factors, are correlated with those in the Y equation, where the number of factors is four. The mean value of the factor loadings is zero.

Instruments	$\beta_1 = 0.4$ ($\beta_1 = \frac{1-\lambda}{k}, \lambda = 0.2$)					$\beta_1 = 0.25$ ($\beta_1 = \frac{1-\lambda}{k}, \lambda = 0.5$)					$\beta_1 = 0.10$ ($\beta_1 = \frac{1-\lambda}{k}, \lambda = 0.8$)				
$\Delta\mathbf{y}_{i,-2}, \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.617	0.616	0.617	0.614	0.615	0.467	0.468	0.468	0.465	0.467	0.302	0.311	0.313	0.310	0.313
20	0.616	0.617	0.622	0.620	0.619	0.466	0.468	0.473	0.471	0.470	0.309	0.312	0.319	0.317	0.316
50	0.620	0.619	0.622	0.623	0.623	0.470	0.470	0.474	0.475	0.474	0.315	0.316	0.320	0.321	0.320
100	0.621	0.622	0.624	0.623	0.624	0.472	0.473	0.475	0.474	0.475	0.318	0.319	0.321	0.320	0.321
200	0.616	0.620	0.625	0.625	0.625	0.467	0.471	0.476	0.476	0.477	0.313	0.317	0.322	0.322	0.322
$\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.611	0.609	0.608	0.604	0.605	0.457	0.458	0.456	0.452	0.454	0.294	0.301	0.301	0.298	0.299
20	0.607	0.607	0.611	0.609	0.608	0.456	0.455	0.460	0.457	0.456	0.300	0.300	0.306	0.303	0.302
50	0.610	0.608	0.611	0.612	0.612	0.458	0.457	0.459	0.460	0.460	0.304	0.302	0.305	0.306	0.305
100	0.610	0.611	0.612	0.612	0.612	0.459	0.459	0.460	0.460	0.460	0.304	0.304	0.306	0.305	0.305
200	0.605	0.609	0.613	0.613	0.614	0.453	0.457	0.461	0.461	0.462	0.299	0.302	0.307	0.306	0.307
$\mathbf{M}_{\Delta\bar{X}}\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.552	0.562	0.556	0.556	0.555	0.399	0.413	0.405	0.405	0.404	0.225	0.253	0.251	0.251	0.252
20	0.562	0.563	0.563	0.562	0.562	0.411	0.412	0.412	0.411	0.411	0.258	0.258	0.259	0.258	0.258
50	0.563	0.559	0.566	0.567	0.564	0.412	0.408	0.415	0.416	0.412	0.261	0.256	0.262	0.263	0.259
100	0.564	0.565	0.565	0.566	0.565	0.413	0.414	0.414	0.414	0.414	0.260	0.261	0.261	0.261	0.261
200	0.556	0.561	0.566	0.567	0.567	0.405	0.410	0.415	0.416	0.416	0.253	0.257	0.262	0.263	0.262
$\mathbf{M}_{\Delta\hat{F}}\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.406	0.416	0.402	0.402	0.402	0.255	0.268	0.252	0.252	0.252	0.084	0.123	0.101	0.102	0.102
20	0.409	0.405	0.402	0.401	0.399	0.259	0.255	0.252	0.251	0.249	0.107	0.104	0.103	0.102	0.099
50	0.403	0.398	0.401	0.400	0.399	0.253	0.248	0.251	0.250	0.249	0.103	0.098	0.101	0.100	0.099
100	0.405	0.400	0.399	0.400	0.400	0.255	0.250	0.249	0.250	0.250	0.105	0.100	0.099	0.100	0.100
200	0.397	0.399	0.400	0.400	0.400	0.247	0.249	0.250	0.250	0.250	0.098	0.099	0.100	0.100	0.100
$\mathbf{M}_{\Delta F}\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.400	0.407	0.402	0.400	0.400	0.244	0.259	0.252	0.250	0.250	0.088	0.113	0.101	0.100	0.100
20	0.401	0.402	0.402	0.401	0.399	0.251	0.252	0.252	0.251	0.249	0.104	0.101	0.102	0.101	0.099
50	0.403	0.398	0.401	0.400	0.399	0.253	0.248	0.251	0.250	0.249	0.103	0.098	0.101	0.100	0.099
100	0.402	0.399	0.399	0.400	0.400	0.252	0.250	0.249	0.250	0.250	0.103	0.100	0.099	0.100	0.100
200	0.398	0.399	0.400	0.400	0.400	0.248	0.249	0.250	0.250	0.250	0.098	0.099	0.100	0.100	0.100

Notes: See notes to Table 1. The reported estimates are of β_1 which is coefficient on x_{1it}. The results for the estimates of β_2 are very similar and not reported (available upon request from the authors).

Table 3: Root Mean Square Error of the Estimated Coefficient of the Lagged Dependent Variable in ARDL(1,0) models. The Factor loadings in the X equation, which contains two factors, are correlated with those in the Y equation, where the number of factors is four. The mean value of the factor loadings is zero.

Instruments	$\lambda = 0.2$					$\lambda = 0.5$					$\lambda = 0.8$				
$\Delta \mathbf{y}_{i,-2}, \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.200	0.144	0.105	0.087	0.080	0.293	0.210	0.149	0.120	0.109	0.555	0.424	0.274	0.206	0.171
20	0.135	0.103	0.076	0.067	0.060	0.193	0.144	0.106	0.091	0.079	0.363	0.280	0.185	0.143	0.114
50	0.090	0.072	0.057	0.052	0.048	0.127	0.101	0.077	0.069	0.063	0.220	0.172	0.114	0.090	0.071
100	0.072	0.059	0.051	0.046	0.044	0.101	0.082	0.068	0.061	0.056	0.167	0.124	0.086	0.067	0.053
200	0.061	0.052	0.046	0.043	0.042	0.084	0.070	0.061	0.057	0.054	0.127	0.094	0.067	0.052	0.043
$\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.287	0.174	0.115	0.084	0.067	0.524	0.279	0.179	0.130	0.105	1.468	0.710	0.395	0.262	0.215
20	0.171	0.121	0.079	0.061	0.051	0.272	0.191	0.123	0.095	0.078	0.796	0.472	0.252	0.189	0.156
50	0.107	0.078	0.053	0.044	0.039	0.168	0.120	0.081	0.067	0.060	0.402	0.250	0.159	0.132	0.119
100	0.078	0.057	0.042	0.038	0.035	0.123	0.089	0.065	0.058	0.054	0.259	0.178	0.127	0.115	0.106
200	0.058	0.046	0.037	0.033	0.032	0.090	0.070	0.057	0.051	0.050	0.193	0.139	0.111	0.100	0.098
$\mathbf{M}_{\Delta \bar{X}} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.332	0.198	0.128	0.092	0.075	0.633	0.330	0.196	0.141	0.116	1.352	1.876	0.487	0.300	0.244
20	0.176	0.126	0.080	0.061	0.049	0.299	0.199	0.124	0.094	0.076	0.833	0.738	0.259	0.189	0.150
50	0.106	0.075	0.050	0.041	0.035	0.165	0.116	0.076	0.063	0.054	0.514	0.242	0.151	0.124	0.106
100	0.075	0.054	0.038	0.033	0.030	0.117	0.083	0.059	0.051	0.046	0.251	0.168	0.117	0.101	0.090
200	0.054	0.041	0.032	0.028	0.027	0.084	0.064	0.049	0.043	0.041	0.182	0.127	0.097	0.085	0.081
$\mathbf{M}_{\Delta \hat{F}} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.334	0.203	0.123	0.081	0.056	0.619	0.363	0.196	0.123	0.086	1.786	1.419	0.499	0.264	0.179
20	0.171	0.118	0.072	0.052	0.035	0.285	0.186	0.113	0.080	0.054	0.786	0.463	0.253	0.164	0.110
50	0.100	0.068	0.042	0.031	0.021	0.157	0.106	0.065	0.047	0.032	0.348	0.226	0.133	0.096	0.065
100	0.069	0.047	0.029	0.021	0.015	0.109	0.074	0.045	0.033	0.023	0.248	0.156	0.093	0.066	0.047
200	0.049	0.032	0.021	0.014	0.010	0.076	0.050	0.032	0.022	0.016	0.185	0.105	0.066	0.043	0.032
$\mathbf{M}_{\Delta F} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.309	0.184	0.116	0.080	0.055	0.550	0.304	0.184	0.122	0.085	1.283	3.370	0.461	0.257	0.176
20	0.165	0.117	0.072	0.051	0.035	0.298	0.185	0.113	0.079	0.054	0.762	0.536	0.255	0.163	0.109
50	0.098	0.068	0.042	0.030	0.021	0.155	0.105	0.065	0.047	0.032	0.345	0.225	0.132	0.095	0.065
100	0.069	0.047	0.029	0.021	0.015	0.109	0.074	0.045	0.032	0.023	0.259	0.155	0.093	0.066	0.047
200	0.049	0.032	0.021	0.014	0.010	0.076	0.050	0.032	0.021	0.016	0.172	0.104	0.066	0.043	0.032

Notes: See notes to Table 1.

Table 4: Root Mean Square Error of the Estimated Coefficient of the Regressor in ARDL(1,0) models. The Factor loadings in the X equation, which contains two factors, are correlated with those in the Y equation, where the number of factors is four. The mean value of the factor loadings is zero.

Instruments	$\beta_1 = 0.4$ ($\beta_1 = \frac{1-\lambda}{k}, \lambda = 0.2$)					$\beta_1 = 0.25$ ($\beta_1 = \frac{1-\lambda}{k}, \lambda = 0.5$)					$\beta_1 = 0.10$ ($\beta_1 = \frac{1-\lambda}{k}, \lambda = 0.8$)				
$\Delta \mathbf{y}_{i,-2}, \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.320	0.274	0.242	0.228	0.224	0.322	0.276	0.243	0.229	0.225	0.321	0.274	0.239	0.225	0.221
20	0.282	0.252	0.238	0.228	0.224	0.283	0.253	0.239	0.230	0.225	0.279	0.250	0.235	0.226	0.220
50	0.264	0.243	0.232	0.228	0.226	0.265	0.244	0.233	0.230	0.227	0.260	0.239	0.229	0.225	0.223
100	0.257	0.241	0.231	0.228	0.226	0.257	0.242	0.232	0.229	0.227	0.253	0.238	0.228	0.224	0.223
200	0.250	0.237	0.232	0.229	0.227	0.251	0.238	0.233	0.230	0.228	0.247	0.234	0.229	0.225	0.224
$\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.319	0.266	0.233	0.218	0.213	0.331	0.267	0.231	0.216	0.212	0.414	0.281	0.230	0.213	0.208
20	0.274	0.242	0.227	0.218	0.213	0.273	0.241	0.226	0.216	0.211	0.315	0.242	0.222	0.212	0.206
50	0.255	0.232	0.221	0.217	0.214	0.253	0.230	0.219	0.215	0.212	0.251	0.227	0.214	0.211	0.208
100	0.246	0.230	0.220	0.216	0.214	0.244	0.228	0.218	0.214	0.212	0.239	0.224	0.213	0.209	0.208
200	0.240	0.226	0.221	0.217	0.215	0.238	0.224	0.219	0.215	0.213	0.233	0.219	0.214	0.210	0.208
$\mathbf{M}_{\Delta \bar{X}} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.334	0.259	0.201	0.181	0.172	0.359	0.263	0.200	0.180	0.171	0.395	0.363	0.204	0.178	0.169
20	0.256	0.212	0.186	0.175	0.170	0.257	0.211	0.185	0.174	0.169	0.284	0.217	0.184	0.172	0.166
50	0.220	0.191	0.179	0.175	0.169	0.219	0.190	0.178	0.174	0.167	0.227	0.188	0.175	0.171	0.164
100	0.207	0.187	0.176	0.172	0.169	0.205	0.186	0.174	0.171	0.168	0.202	0.183	0.171	0.167	0.165
200	0.194	0.182	0.175	0.173	0.170	0.193	0.181	0.174	0.171	0.169	0.190	0.177	0.171	0.168	0.166
$\mathbf{M}_{\Delta \hat{F}} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.426	0.270	0.150	0.101	0.069	0.480	0.274	0.151	0.101	0.069	0.634	0.409	0.159	0.103	0.070
20	0.240	0.146	0.094	0.065	0.045	0.242	0.147	0.094	0.065	0.045	0.271	0.156	0.095	0.065	0.045
50	0.132	0.089	0.057	0.039	0.027	0.133	0.089	0.057	0.039	0.027	0.135	0.090	0.057	0.039	0.027
100	0.093	0.063	0.038	0.027	0.019	0.093	0.063	0.038	0.027	0.019	0.094	0.063	0.038	0.027	0.019
200	0.064	0.044	0.027	0.019	0.013	0.064	0.044	0.027	0.019	0.013	0.064	0.044	0.027	0.019	0.013
$\mathbf{M}_{\Delta F} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.323	0.218	0.138	0.096	0.068	0.360	0.223	0.138	0.096	0.068	0.426	0.459	0.145	0.098	0.069
20	0.207	0.141	0.092	0.065	0.045	0.212	0.142	0.092	0.065	0.045	0.237	0.158	0.094	0.065	0.045
50	0.123	0.087	0.056	0.039	0.027	0.124	0.087	0.056	0.039	0.027	0.126	0.088	0.056	0.039	0.027
100	0.087	0.061	0.037	0.027	0.019	0.088	0.061	0.037	0.027	0.019	0.089	0.061	0.037	0.027	0.019
200	0.060	0.043	0.026	0.019	0.013	0.060	0.043	0.026	0.019	0.013	0.060	0.043	0.026	0.019	0.013

Notes: See notes to Table 1.

Table 5: Estimated Size of the t-test for the Estimated Coefficient of the Lagged Dependent Variable in ARDL(1,0) models. The Factor loadings in the X equation, which contains two factors, are correlated with those in the Y equation, where the number of factors is four. The mean value of the factor loadings is zero.

Instruments	$H_0 : \lambda = 0.2$ (DGP: $\lambda = 0.2$)					$H_0 : \lambda = 0.5$ (DGP: $\lambda = 0.5$)					$H_0 : \lambda = 0.8$ (DGP: $\lambda = 0.8$)				
$\Delta \mathbf{y}_{i,-2}, \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.146	0.113	0.145	0.206	0.321	0.140	0.109	0.137	0.180	0.277	0.154	0.116	0.107	0.118	0.172
20	0.149	0.137	0.166	0.270	0.377	0.138	0.125	0.137	0.231	0.325	0.138	0.107	0.098	0.123	0.167
50	0.170	0.171	0.264	0.418	0.601	0.167	0.159	0.223	0.359	0.529	0.139	0.120	0.098	0.145	0.172
100	0.200	0.232	0.418	0.599	0.771	0.179	0.207	0.362	0.516	0.697	0.142	0.112	0.120	0.154	0.206
200	0.262	0.334	0.593	0.811	0.937	0.246	0.303	0.513	0.711	0.893	0.158	0.132	0.156	0.189	0.268
$\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.128	0.085	0.101	0.110	0.142	0.135	0.093	0.111	0.115	0.149	0.149	0.123	0.134	0.132	0.171
20	0.114	0.095	0.097	0.131	0.188	0.115	0.106	0.105	0.140	0.192	0.140	0.129	0.131	0.161	0.219
50	0.139	0.122	0.133	0.195	0.322	0.149	0.129	0.138	0.201	0.333	0.171	0.156	0.161	0.222	0.361
100	0.145	0.141	0.196	0.313	0.517	0.159	0.150	0.204	0.322	0.521	0.174	0.174	0.225	0.345	0.541
200	0.174	0.195	0.304	0.490	0.763	0.184	0.206	0.314	0.501	0.766	0.203	0.230	0.338	0.529	0.773
$\mathbf{M}_{\Delta \bar{X}} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.130	0.082	0.102	0.102	0.143	0.136	0.089	0.110	0.106	0.148	0.160	0.118	0.124	0.117	0.163
20	0.119	0.093	0.092	0.113	0.147	0.129	0.100	0.096	0.122	0.156	0.145	0.131	0.112	0.143	0.169
50	0.127	0.110	0.103	0.160	0.250	0.135	0.114	0.111	0.165	0.257	0.159	0.130	0.130	0.179	0.275
100	0.132	0.114	0.155	0.247	0.393	0.136	0.118	0.161	0.255	0.399	0.145	0.134	0.177	0.271	0.417
200	0.150	0.156	0.231	0.370	0.582	0.154	0.162	0.236	0.377	0.589	0.162	0.181	0.258	0.398	0.603
$\mathbf{M}_{\Delta \hat{F}} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.116	0.075	0.062	0.055	0.044	0.121	0.077	0.064	0.055	0.043	0.133	0.103	0.073	0.060	0.040
20	0.091	0.077	0.055	0.056	0.056	0.094	0.078	0.055	0.058	0.058	0.117	0.086	0.059	0.059	0.055
50	0.106	0.070	0.057	0.061	0.048	0.111	0.072	0.054	0.063	0.049	0.122	0.079	0.047	0.060	0.044
100	0.107	0.070	0.045	0.064	0.059	0.100	0.066	0.044	0.065	0.056	0.092	0.064	0.044	0.060	0.055
200	0.095	0.076	0.059	0.043	0.059	0.092	0.072	0.059	0.043	0.061	0.090	0.074	0.056	0.046	0.059
$\mathbf{M}_{\Delta F} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.097	0.066	0.061	0.049	0.044	0.092	0.067	0.061	0.053	0.041	0.115	0.085	0.071	0.063	0.040
20	0.091	0.069	0.053	0.057	0.053	0.089	0.070	0.053	0.055	0.054	0.105	0.075	0.057	0.057	0.055
50	0.098	0.069	0.053	0.058	0.050	0.095	0.069	0.051	0.057	0.049	0.104	0.070	0.045	0.053	0.047
100	0.095	0.069	0.050	0.059	0.060	0.087	0.065	0.047	0.059	0.059	0.080	0.060	0.045	0.057	0.059
200	0.095	0.070	0.052	0.043	0.059	0.092	0.068	0.052	0.044	0.060	0.084	0.065	0.050	0.044	0.058

Notes: See notes to Table 1. The t test is based on the variance estimator defined by (15) but replacing instruments with associated estimators accordingly (see Section 3 for more details). The tests are conducted at the 5% significance level.

Table 6: Estimated Size of the t-test for the Regressor in ARDL(1,0) models. The Factor loadings in the X equation, which contains two factors, are correlated with those in the Y equation, where the number of factors is four. The mean value of the factor loadings is zero.

Instruments															
$\Delta\mathbf{y}_{i,-2}, \mathbf{Z}_i$	$H_0 : \beta_1 = 0.4$ (DGP: $\beta_1 = 0.4$)					$H_0 : \beta_1 = 0.25$ (DGP: $\beta_1 = 0.25$)					$H_0 : \beta_1 = 0.1$ (DGP: $\beta_1 = 0.1$)				
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.277	0.349	0.593	0.800	0.962	0.271	0.344	0.595	0.804	0.963	0.230	0.312	0.566	0.790	0.960
20	0.351	0.473	0.782	0.954	0.999	0.349	0.476	0.783	0.957	0.999	0.329	0.461	0.773	0.953	0.999
50	0.445	0.629	0.924	0.998	1.000	0.446	0.631	0.923	0.998	1.000	0.440	0.622	0.918	0.997	1.000
100	0.495	0.706	0.966	1.000	1.000	0.493	0.709	0.965	1.000	1.000	0.490	0.701	0.962	1.000	1.000
200	0.527	0.732	0.978	1.000	1.000	0.529	0.734	0.979	1.000	1.000	0.529	0.728	0.977	1.000	1.000
$\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.267	0.333	0.576	0.789	0.957	0.252	0.321	0.563	0.785	0.956	0.208	0.271	0.519	0.767	0.950
20	0.335	0.455	0.766	0.948	0.999	0.333	0.451	0.760	0.945	0.999	0.294	0.409	0.740	0.940	0.998
50	0.429	0.607	0.907	0.996	1.000	0.430	0.610	0.908	0.996	1.000	0.414	0.592	0.906	0.995	1.000
100	0.469	0.680	0.954	0.999	1.000	0.469	0.678	0.953	0.999	1.000	0.458	0.671	0.953	0.999	1.000
200	0.508	0.703	0.970	1.000	1.000	0.506	0.702	0.970	1.000	1.000	0.502	0.697	0.971	1.000	1.000
$\mathbf{M}_{\Delta\bar{X}}\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.168	0.198	0.310	0.483	0.714	0.147	0.191	0.299	0.482	0.710	0.123	0.156	0.266	0.444	0.692
20	0.234	0.292	0.515	0.751	0.939	0.221	0.286	0.511	0.751	0.938	0.201	0.252	0.490	0.740	0.934
50	0.318	0.432	0.760	0.953	0.996	0.319	0.431	0.759	0.952	0.996	0.294	0.421	0.755	0.950	0.996
100	0.363	0.552	0.849	0.985	1.000	0.363	0.552	0.849	0.985	1.000	0.356	0.545	0.848	0.983	1.000
200	0.389	0.584	0.921	0.996	1.000	0.389	0.581	0.922	0.996	1.000	0.388	0.576	0.919	0.996	1.000
$\mathbf{M}_{\Delta\hat{F}}\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.092	0.060	0.056	0.047	0.051	0.086	0.058	0.052	0.047	0.052	0.068	0.047	0.042	0.042	0.049
20	0.095	0.060	0.066	0.051	0.050	0.098	0.058	0.066	0.052	0.049	0.081	0.049	0.060	0.049	0.046
50	0.089	0.065	0.063	0.058	0.048	0.087	0.062	0.061	0.058	0.048	0.078	0.064	0.061	0.056	0.046
100	0.089	0.065	0.048	0.051	0.049	0.088	0.066	0.048	0.051	0.051	0.085	0.063	0.047	0.053	0.052
200	0.088	0.069	0.055	0.051	0.056	0.088	0.069	0.054	0.050	0.056	0.086	0.068	0.056	0.051	0.056
$\mathbf{M}_{\Delta F}\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.090	0.063	0.055	0.043	0.052	0.082	0.058	0.049	0.043	0.052	0.071	0.050	0.044	0.039	0.050
20	0.098	0.065	0.059	0.052	0.049	0.096	0.061	0.060	0.052	0.050	0.082	0.050	0.056	0.047	0.049
50	0.098	0.064	0.059	0.055	0.049	0.096	0.063	0.059	0.055	0.048	0.080	0.060	0.059	0.053	0.047
100	0.094	0.076	0.050	0.051	0.050	0.092	0.073	0.050	0.051	0.050	0.088	0.071	0.050	0.051	0.051
200	0.088	0.073	0.051	0.049	0.056	0.089	0.072	0.051	0.050	0.056	0.089	0.068	0.051	0.050	0.055

Notes: See notes to Tables 5 and 1.

Table 7: Estimated Power of the t-test for the Estimated Coefficient of the Lagged Dependent Variable in ARDL(1,0) models. The Factor loadings in the X equation, which contains two factors, are correlated with those in the Y equation, where the number of factors is four. The mean value of the factor loadings is zero.

Instruments	$H_0 : \lambda = 0.1$ (DGP: $\lambda = 0.2$)					$H_0 : \lambda = 0.4$ (DGP: $\lambda = 0.5$)					$H_0 : \lambda = 0.7$ (DGP: $\lambda = 0.8$)				
$\Delta \mathbf{y}_{i,-2}, \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.222	0.178	0.240	0.308	0.417	0.202	0.153	0.180	0.228	0.294	0.201	0.155	0.167	0.178	0.231
20	0.229	0.228	0.281	0.418	0.587	0.190	0.174	0.183	0.245	0.380	0.189	0.162	0.166	0.205	0.285
50	0.276	0.327	0.472	0.656	0.833	0.215	0.222	0.259	0.363	0.534	0.209	0.196	0.207	0.274	0.409
100	0.349	0.433	0.654	0.849	0.962	0.241	0.254	0.359	0.525	0.723	0.221	0.219	0.269	0.399	0.587
200	0.464	0.611	0.842	0.967	0.998	0.289	0.352	0.500	0.695	0.872	0.256	0.284	0.382	0.554	0.772
$\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.229	0.226	0.337	0.492	0.675	0.202	0.182	0.258	0.344	0.484	0.209	0.181	0.220	0.256	0.324
20	0.281	0.325	0.513	0.731	0.920	0.236	0.249	0.367	0.509	0.752	0.224	0.209	0.262	0.330	0.492
50	0.424	0.547	0.822	0.961	0.999	0.332	0.397	0.612	0.838	0.973	0.267	0.287	0.400	0.567	0.803
100	0.601	0.770	0.969	1.000	1.000	0.455	0.570	0.840	0.978	1.000	0.320	0.393	0.579	0.800	0.957
200	0.771	0.929	0.999	1.000	1.000	0.602	0.789	0.977	1.000	1.000	0.433	0.562	0.811	0.968	0.999
$\mathbf{M}_{\Delta \bar{X}} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.215	0.209	0.307	0.406	0.566	0.202	0.179	0.232	0.292	0.411	0.207	0.174	0.207	0.220	0.281
20	0.276	0.297	0.482	0.658	0.875	0.232	0.232	0.309	0.447	0.672	0.219	0.199	0.223	0.284	0.412
50	0.409	0.513	0.792	0.946	0.998	0.308	0.363	0.577	0.791	0.955	0.248	0.263	0.343	0.502	0.724
100	0.576	0.733	0.961	0.999	1.000	0.414	0.531	0.801	0.959	0.999	0.291	0.347	0.502	0.734	0.915
200	0.740	0.922	0.999	1.000	1.000	0.559	0.753	0.962	1.000	1.000	0.380	0.503	0.747	0.931	0.995
$\mathbf{M}_{\Delta \hat{F}} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.194	0.167	0.236	0.330	0.481	0.176	0.147	0.172	0.212	0.283	0.180	0.145	0.134	0.137	0.144
20	0.238	0.250	0.367	0.564	0.806	0.198	0.177	0.230	0.325	0.501	0.177	0.142	0.145	0.170	0.222
50	0.357	0.432	0.697	0.901	0.993	0.264	0.278	0.420	0.623	0.872	0.198	0.179	0.188	0.263	0.394
100	0.502	0.640	0.909	0.992	1.000	0.330	0.399	0.620	0.863	0.986	0.200	0.207	0.273	0.416	0.646
200	0.671	0.862	0.994	1.000	1.000	0.461	0.620	0.868	0.988	1.000	0.261	0.283	0.429	0.641	0.876
$\mathbf{M}_{\Delta F} \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.187	0.185	0.247	0.344	0.491	0.171	0.148	0.182	0.220	0.297	0.173	0.138	0.137	0.144	0.150
20	0.220	0.241	0.373	0.565	0.810	0.172	0.172	0.230	0.328	0.508	0.161	0.145	0.139	0.173	0.223
50	0.346	0.414	0.690	0.898	0.993	0.240	0.273	0.407	0.618	0.871	0.177	0.172	0.180	0.266	0.396
100	0.479	0.622	0.905	0.991	1.000	0.293	0.381	0.610	0.859	0.983	0.170	0.187	0.256	0.407	0.649
200	0.647	0.853	0.993	1.000	1.000	0.430	0.592	0.859	0.988	1.000	0.234	0.246	0.412	0.637	0.875

Notes: See notes to 5 and 1.

Table 8: Estimated Power of the t-test for the Regressor in ARDL(1,0) models. The Factor loadings in the X equation, which contains two factors, are correlated with those in the Y equation, where the number of factors is four. The mean value of the factor loadings is zero.

Instruments	$H_0 : \beta_1 = 0.3$ (DGP: $\beta_1 = 0.4$)					$H_0 : \beta_1 = 0.15$ (DGP: $\beta_1 = 0.25$)					$H_0 : \beta_1 = 0$ (DGP: $\beta_1 = 0.1$)				
$\Delta\mathbf{y}_{i,-2}, \mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.170	0.168	0.258	0.381	0.602	0.163	0.167	0.259	0.386	0.606	0.133	0.145	0.231	0.359	0.577
20	0.192	0.201	0.356	0.559	0.811	0.193	0.202	0.356	0.567	0.818	0.175	0.191	0.339	0.539	0.795
50	0.230	0.268	0.492	0.772	0.953	0.233	0.272	0.500	0.780	0.958	0.223	0.263	0.478	0.757	0.946
100	0.238	0.323	0.578	0.835	0.985	0.239	0.327	0.584	0.842	0.985	0.240	0.316	0.562	0.819	0.980
200	0.261	0.340	0.625	0.878	0.995	0.264	0.344	0.634	0.881	0.995	0.262	0.334	0.613	0.867	0.994
$\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.155	0.154	0.237	0.343	0.560	0.147	0.150	0.227	0.339	0.551	0.122	0.126	0.188	0.303	0.515
20	0.183	0.184	0.326	0.509	0.763	0.180	0.175	0.315	0.496	0.759	0.160	0.158	0.281	0.461	0.722
50	0.214	0.250	0.444	0.708	0.915	0.210	0.241	0.436	0.700	0.910	0.201	0.222	0.413	0.666	0.893
100	0.227	0.290	0.522	0.774	0.964	0.223	0.287	0.512	0.772	0.960	0.214	0.272	0.490	0.753	0.951
200	0.243	0.309	0.560	0.827	0.987	0.242	0.307	0.556	0.817	0.987	0.235	0.298	0.538	0.797	0.980
$\mathbf{M}_{\Delta\bar{X}}\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.107	0.090	0.118	0.147	0.233	0.101	0.084	0.111	0.141	0.228	0.082	0.068	0.093	0.127	0.205
20	0.133	0.109	0.151	0.234	0.368	0.129	0.100	0.149	0.229	0.360	0.109	0.089	0.135	0.209	0.334
50	0.149	0.147	0.230	0.363	0.552	0.145	0.145	0.223	0.358	0.543	0.141	0.130	0.207	0.331	0.512
100	0.156	0.156	0.270	0.437	0.645	0.154	0.155	0.265	0.428	0.636	0.146	0.149	0.256	0.408	0.617
200	0.165	0.177	0.290	0.472	0.728	0.163	0.173	0.283	0.467	0.720	0.158	0.167	0.270	0.449	0.700
$\mathbf{M}_{\Delta\hat{F}}\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.103	0.091	0.110	0.175	0.305	0.094	0.085	0.107	0.173	0.305	0.077	0.068	0.088	0.167	0.293
20	0.120	0.121	0.209	0.357	0.601	0.113	0.115	0.204	0.356	0.605	0.102	0.103	0.195	0.354	0.601
50	0.181	0.259	0.458	0.735	0.946	0.178	0.255	0.461	0.738	0.948	0.172	0.252	0.458	0.740	0.951
100	0.281	0.423	0.763	0.948	1.000	0.279	0.418	0.765	0.949	1.000	0.272	0.409	0.764	0.949	1.000
200	0.474	0.680	0.958	0.999	1.000	0.478	0.682	0.960	0.999	1.000	0.478	0.678	0.961	1.000	1.000
$\mathbf{M}_{\Delta F}\mathbf{Z}_i$															
T,N	10	20	50	100	200	10	20	50	100	200	10	20	50	100	200
10	0.110	0.084	0.119	0.196	0.315	0.100	0.076	0.114	0.192	0.313	0.087	0.068	0.108	0.180	0.309
20	0.128	0.133	0.221	0.358	0.610	0.124	0.125	0.215	0.360	0.610	0.116	0.113	0.204	0.353	0.608
50	0.201	0.281	0.467	0.739	0.951	0.196	0.280	0.466	0.739	0.951	0.193	0.265	0.466	0.742	0.955
100	0.315	0.452	0.777	0.952	1.000	0.314	0.449	0.778	0.952	1.000	0.298	0.437	0.779	0.950	1.000
200	0.492	0.712	0.962	1.000	1.000	0.487	0.708	0.961	1.000	1.000	0.491	0.704	0.961	1.000	1.000

Notes: See notes to Tables 5 and 1.

Table 9: Estimated Size and Power of the Overidentifying Restrictions Test in ARDL(1,0) models, for $\lambda = 0.5$, $\beta_1 = 0.25$.

Instruments	**Size**					**Power:** ε_{it} and x_{1ix} are correlated				
$\mathbf{M}_{\Delta\hat{F}}\mathbf{Z}_i$										
T,N	10	20	50	100	200	10	20	50	100	200
10	0.038	0.041	0.044	0.048	0.049	0.051	0.094	0.195	0.370	0.672
20	0.032	0.051	0.044	0.041	0.053	0.084	0.184	0.462	0.753	0.943
50	0.032	0.043	0.045	0.046	0.048	0.179	0.435	0.847	0.988	1.000
100	0.035	0.045	0.050	0.042	0.038	0.321	0.694	0.977	0.999	1.000
200	0.035	0.040	0.047	0.048	0.049	0.497	0.872	0.999	1.000	1.000
$\mathbf{M}_{\Delta F}\mathbf{Z}_i$										
T,N	10	20	50	100	200	10	20	50	100	200
10	0.039	0.047	0.048	0.048	0.050	0.053	0.101	0.208	0.386	0.668
20	0.030	0.051	0.044	0.045	0.050	0.085	0.191	0.463	0.748	0.945
50	0.033	0.044	0.044	0.046	0.047	0.190	0.435	0.856	0.986	1.000
100	0.035	0.048	0.047	0.042	0.038	0.337	0.701	0.979	1.000	1.000
200	0.035	0.038	0.045	0.049	0.049	0.511	0.874	1.000	1.000	1.000

Notes: The DGP follows $y_{it} = \alpha_i + \lambda y_{it-1} + \sum_{\ell=1}^{k} \beta_\ell x_{\ell it} + \sum_{s=1}^{m} \gamma_{si} f_{st} + \varepsilon_{it}$, $\varepsilon_{it} = \sigma_{it}(\eta_{it} - 1)/2^{1/2}$, $\eta_{it} \sim iid\chi_1^2$, with $k = 2$ and $m = 4$, where $\beta_\ell = (1-\lambda)/k$, $x_{\ell it} = \mu_{\ell i} + \sum_{s=1}^{m_x} \gamma_{\ell si} f_{st} + v_{\ell it}$ with $m_x = 2$, $v_{\ell it} = \rho_\ell v_{\ell it-1} + (1-\rho_\ell^2)^{1/2}\varpi_{\ell it}$, $\varpi_{\ell it} = \pi_\ell \varepsilon_{it} + (1-\pi_\ell^2)^{1/2}\varrho_{\ell it}$ with $\varrho_{\ell it} \sim iidN(0,1)$, $\ell = 1, 2, ..., k$. To estimate the size of the test the identical DGP to that in Table 1 is used, namely, $\pi_\ell = 0$ for $\ell = 1, 2 (= k)$. To estimate the power of the test, $\pi_1 = 0.4$ and $\pi_2 = 0$ so that the idiosyncratic error of x_{1it} is correlated with ε_{it}. The rest of the specifications of the DGP is identical to those in Table 1. The overidentifying restrictions test statistic is defined by (24), and the 5% critical value from χ_1^2 distribution is used for the test. All the experiments are based on 2000 replications.