Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

A Linear Estimator for Factor-Augmented Fixed-T Panels With Endogenous Regressors

Artūras Juodis & Vasilis Sarafidis

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2020, VOL. 00, NO. 0, 1–15
https://doi.org/10.1080/07350015.2020.1766469

A Linear Estimator for Factor-Augmented Fixed-T Panels With Endogenous Regressors

Artūras Juodis[a] and Vasilis Sarafidis[b,c]

[a]Faculty of Economics and Business, University of Groningen, Groningen, Netherlands; [b]Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria, Australia; [c]Department of Economics, BI Norwegian Business School, Oslo, Norway

ABSTRACT

A novel method-of-moments approach is proposed for the estimation of factor-augmented panel data models with endogenous regressors when T is fixed. The underlying methodology involves approximating the unobserved common factors using observed factor proxies. The resulting moment conditions are linear in the parameters. The proposed approach addresses several issues which arise with existing nonlinear estimators that are available in fixed T panels, such as local minima-related problems, a sensitivity to particular normalization schemes, and a potential lack of global identification. We apply our approach to a large panel of households and estimate the price elasticity of urban water demand. A simulation study confirms that our approach performs well in finite samples.

ARTICLE HISTORY
Received January 2019
Accepted April 2020

KEYWORDS
Common factors; Fixed T consistency; Moment conditions; Panel data; Urban water management

1. Introduction

The common factor approach has attracted considerable interest within panel data analysis because it offers a wide scope for controlling for unobservables, including situations where there is cross-sectional dependence (see, e.g., Sarafidis and Wansbeek 2012, 2020).

Holtz-Eakin, Newey, and Rosen (1988), Ahn, Lee, and Schmidt (2013), Robertson and Sarafidis (2015), Robertson, Sarafidis, and Westerlund (2018), and Juodis (2018), among others, have proposed various estimators for panels with endogenous covariates and "fixed T," where T denotes the number of time series observations. A common feature of these approaches is that identification relies on nonlinear moment conditions. For panels with "large T," popular (least-squares type) methods include those developed by Pesaran (2006) and Bai (2009), known in the literature as common correlated effects (CCE) and principal components (PC), respectively.

Given the computational simplicity inherent in CCE and PC, these approaches have been highly popular, both in terms of extending them to several additional theoretical settings (see, e.g., Kapetanios, Serlenga, and Shin 2020; Norkute et al. 2020 for recent developments), as well as in terms of applying these procedures to a large range of empirical areas. In contrast, it is fair to say that the literature on fixed-T panels remains largely unused by empirical researchers, despite its aforementioned volume. There are several factors that might help to explain this observation. First, there is usually no underlying theory to guide the selection of good starting values for a potentially large number of nuisance parameters, such as the unobserved common factors. This can potentially lead to local minima-related problems, frequently arising in estimation of factor models using iterative algorithms (see, e.g., Jiang et al. 2017). Second, identification of the parameters of interest typically requires imposing certain normalization restrictions. However, as was discussed in Kruiniger (2008) and Juodis and Sarafidis (2018), the choice of the normalization scheme can be crucial for the properties of estimators that rely on some form of quasi-differencing, depending on the underlying data generating process for the unknown factors. For example, the approach proposed by Holtz-Eakin, Newey, and Rosen (1988) requires that all factors take nonzero values in all time periods. Finally, as shown by Hayakawa (2016), the nonlinear moment conditions proposed in the literature do not always satisfy the global identification assumption, which is a necessary condition for consistency of GMM estimation.

The present article develops a novel GMM approach; the main idea is to replace the unobserved factors with proxies constructed from observables. We put forward two distinct methods for constructing factor proxies. The first one involves the use of a multiple weighting scheme applied on a single observable, which can be either an external variable or a regressor. The second method employs a single weighting scheme applied to multiple variables. We show that these two methods can also be combined. In both cases, the underlying assumption is that the variables employed to construct the factor proxies are driven by the common shocks that are relevant for the main variable of interest. The resulting method of moments estimator has a closed form solution, and avoids the aforementioned issues associated with nonlinear estimators. Under suitable regularity conditions (discussed in the article), the proposed estimator is consistent and asymptotically mixed-normal.

In response to issues associated with nonlinear method of moments estimators, Westerlund, Petrova, and Norkute (2019) recently advocated the use of pooled CCE in fixed-T panels. However, the computational simplicity of CCE comes with

CONTACT Arturas Juodis arturas.juodis@economists.lt Faculty of Economics and Business, University of Groningen, Nettelbosje 2, 9747 AE, Groningen, The Netherlands.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/UBES.

a price, as the method requires all regressors to be strictly exogenous and to exhibit a common factor structure. The latter restriction prohibits, for example, nonlinear partial effects, such as in regressions with quadratic terms. Recently, De Vos and Everaert (2019) extended CCE in fixed-T panels to the case of a lagged dependent variable, assuming that all other covariates are strictly exogenous. However, the proposed procedure is no longer linear and requires bias correction. In contrast, the approach presented in this article does not need bias correction of any sort, and accommodates regressors with different degrees of exogeneity, without imposing the requirement that *all* covariates (and/or instruments) have as many factors as the main variable on interest. At the same time, the resulting GMM estimator possesses the appealing linearity property of the CCE estimator.

We use our approach to estimate the price elasticity of residential water usage demand. This topic is of large interest, not only among economists but also across international environmental agencies, regulators, water utilities, and the general public. We find that urban water demand is more elastic in the long-run than in the short-run, which may be attributed to habit formation and technological constraints of water appliance efficiency. This result casts doubt on the potential effectiveness of scarcity pricing to balance demand and supply of water in periods of transitory droughts.

The remainder of this article is organized as follows. Section 2 introduces a linear panel model with common shocks. Section 3 develops the proposed approach. Section 4 reports a Monte Carlo study to assess the finite sample performance of the estimator. Section 5 presents the empirical application. Finally, Section 7 concludes. A supplementary appendix to this article discusses extensions to unbalanced panels and models with observed factors. In addition, it provides further finite sample evidence and contains proofs of our theoretical results.

1.1. Notation

The generic constants δ and M are used to denote a small and a large positive real number, respectively. For a generic matrix $\boldsymbol{A}$, vec$(\boldsymbol{A})$ denotes the vertical column stacking operator, and Col$(\boldsymbol{A})$ denotes the column space of $\boldsymbol{A}$. Moreover, $\otimes$ denotes the Kronecker product. $\boldsymbol{l}_{i,s:q}$, $s \leq q$, is defined as $\boldsymbol{l}_{i,s:q} = (l_{i,s}, \ldots, l_{i,q})'$. Finally, all random variables are defined on a common probability space $(\Omega, \mathcal{A}, P)$.

2. Model

We consider the following panel data model with a multifactor error structure:

$$y_{i,t} = \boldsymbol{x}_{i,t}'\boldsymbol{\beta} + \boldsymbol{\lambda}_i'\boldsymbol{f}_t + \varepsilon_{i,t}; \quad i = 1, \ldots, N; \quad t = 1, \ldots, T, \quad (1)$$

where $\boldsymbol{x}_{i,t} = (x_{i,t}^{(1)}, \ldots, x_{i,t}^{(K)})'$ denotes the vector of explanatory variables and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)'$ is the vector of the parameters of interest. $\boldsymbol{f}_t$ denotes the $[L \times 1]$ vector of unobserved common factors, $\boldsymbol{\lambda}_i$ denotes the associated factor loadings for individual i, and $\varepsilon_{i,t}$ is the remaining error term.

The multifactor error structure is appealing because it allows for multiple sources of *multiplicative* unobserved heterogeneity, as opposed to the one-way (or two-way) error components structure, which represents *additive* heterogeneity. For example, in a partial adjustment model of factor input prices, the factor component may capture common shocks that hit all producers, albeit with different intensities. In the estimation of production functions, the factor component may absorb different sources of technical inefficiency, which vary over time in an arbitrary way. In an empirical model of household water usage demand, the factor component may capture nonlinear effects of household size (typically unobserved) that depend on time-varying weather conditions.

Stacking the observations over time for each i, the model can be rewritten in vector form as

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i; \quad i = 1, \ldots, N, \quad (2)$$

where $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,T})'$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,T})'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,T})'$, while $\boldsymbol{F} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T)'$ is of dimension $[T \times L]$.

Denote by $\mathcal{F}$ the σ-field generated by all common shocks driving the individual-specific variables in the system. As such, $\mathcal{F}$ contains all factors $\boldsymbol{F}$, but we also allow variables $\{(\boldsymbol{x}_{i,t}, \boldsymbol{\lambda}_i)\}_{t=1}^T$ to be a function of other common shocks (not necessarily of linear factor structure), resulting in additional sources of dependence across cross-sectional units. For example, one can allow $x_{i,t} = b\left(\psi_i, g_t, \zeta_{i,t}\right)$, where $b(\cdot)$ is a linear/nonlinear function in all arguments.

Assumption 2.1. The DGP for all i and t satisfies the following restrictions:

1. $(\boldsymbol{X}_i, \boldsymbol{\varepsilon}_i, \boldsymbol{\lambda}_i)$ are identically distributed and independent across i, conditional on $\mathcal{F}$.
2. Each time-varying element $p_{i,t}^{(\cdot)}$ in $\boldsymbol{p}_{i,t} = \left(p_{i,t}^{(1)}, \ldots, p_{i,t}^{(K+1)}\right)' \equiv (\boldsymbol{x}_{i,t}', \varepsilon_{i,t})'$ satisfies $\mathrm{E}\left[\left|p_{i,t}^{(\cdot)}\right|^{4+\delta}\right] < \infty$ for all t.
3. Each time-invariant element $\lambda_i^{(\cdot)}$ in $\boldsymbol{\lambda}_i = (\lambda_i^{(1)}, \ldots, \lambda_i^{(L)})'$ satisfies $\mathrm{E}\left[\left|\lambda_i^{(\cdot)}\right|^{4+\delta}\right] < \infty$.
4. $\mathrm{E}_{\mathcal{F}}[\varepsilon_{i,t}|\boldsymbol{\lambda}_i, \boldsymbol{x}_{i,1:\tau_1(t)}^{(1)}, \ldots, \boldsymbol{x}_{i,1:\tau_K(t)}^{(K)}] = 0 \; \forall t$, for some positive integers $\tau_1(t), \ldots, \tau_K(t)$.

Besides the fact that $\boldsymbol{F}$ is assumed to be random, the above assumptions are standard in the literature (see, e.g., Ahn, Lee, and Schmidt 2013, Assumption BA.1). Given the conditional independence assumption, all stochastic convergence modes in this article are conditional on $\mathcal{F}$. We will emphasize this technicality further in Section 3, when we discuss the asymptotic distribution of the proposed estimator. These assumptions are general enough to allow for conditional heteroscedasticity in both dimensions, for example, $\varepsilon_{i,t} = \sigma_i\xi_t\eta_{i,t}$, where $\eta_{i,t}$ is iid over i and t with unit variance, ξ_t is a sequence of constants, while σ_i is an iid sequence over i. Subject to some additional summability restrictions, the conditional iid restriction can be further relaxed to conditional independence with heterogeneous population moments, without affecting the consistency of the estimator. The supplementary appendix provides one such example. For instance, σ_i and $\boldsymbol{\lambda}_i$ could be treated as a sequence of fixed constants, as was advocated in Hsiao, Pesaran, and Tahmiscioglu (2002). However, such setup would require additional technical restrictions for the limits to be well defined (see, e.g., Gagliardini, Ossola, and Scaillet 2016, p. 996, for a related discussion).

Assumption 2.1.4 characterizes the exogeneity properties of the covariates. In particular, covariates that satisfy $\tau_k(t) = T$ ($\tau_k(t) = t$) are strictly (weakly) exogenous with respect to the idiosyncratic error component, and endogenous otherwise. The estimator proposed in this article allows for strictly/weakly exogenous regressors, such as lagged dependent variables and endogenous regressors.

Let $\boldsymbol{z}_i$ be a $[d \times 1]$ vector containing all internal instruments that are available by Assumption 2.1.4, as well as external instruments satisfying the corresponding assumption. Also, let $\boldsymbol{S} = \operatorname{diag}(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_T)$ denote a block-diagonal matrix with a typical block-diagonal entry equal to $\boldsymbol{S}_t$, where $\boldsymbol{S}_t$ is a $[\zeta_t \times d]$ selection matrix of zeros and ones that picks ζ_t valid instruments at time t from $\boldsymbol{z}_i$.

Under Assumption 2.1, the following set of $\zeta \equiv \sum_{t=1}^{T} \zeta_t$ population moment conditions is valid by construction:

$$\begin{aligned} &\mathrm{E}_{\mathcal{F}}\left[\boldsymbol{Z}_i'\left(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_0 - \boldsymbol{F}\boldsymbol{\lambda}_i\right)\right] \\ &\quad = \boldsymbol{S}\left(\operatorname{vec}\left(\mathrm{E}_{\mathcal{F}}\left[\boldsymbol{z}_i\left(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_0\right)'\right] - \boldsymbol{G}_{z,\lambda}\boldsymbol{F}'\right)\right) = \boldsymbol{0}_\zeta, \end{aligned} \tag{3}$$

where $\boldsymbol{Z}_i' \equiv \boldsymbol{S}(\boldsymbol{I}_T \otimes \boldsymbol{z}_i)$ and $\boldsymbol{G}_{z,\lambda} \equiv \mathrm{E}_{\mathcal{F}}\left(\boldsymbol{z}_i\boldsymbol{\lambda}_i'\right)$ is a $[d \times L]$ unknown population matrix that absorbs the unobserved covariances between instruments and factor loadings. The moment conditions in Equation (3) give rise to the estimators proposed by Robertson and Sarafidis (2015) (with $\boldsymbol{G}_{z,\lambda}$ estimated) and Ahn, Lee, and Schmidt (2013) (with $\boldsymbol{G}_{z,\lambda}$ quasi-differenced). Either way, the resulting moment conditions are nonlinear and, hence, potentially subject to the issues discussed in Section 1. In what follows, we put forward a strategy that circumvents the nonlinearity of the moment conditions in Equation (3). For future reference, notice that the last term in Equation (3) can be rewritten as $\boldsymbol{S}\left(\operatorname{vec}\left(\boldsymbol{G}_{z,\lambda}\boldsymbol{F}'\right)\right) = \boldsymbol{S}\left(\boldsymbol{F} \otimes \boldsymbol{I}_d\right)\boldsymbol{g}_{z,\lambda}$, where $\boldsymbol{g}_{z,\lambda} = \operatorname{vec}\left(\boldsymbol{G}_{z,\lambda}\right)$.

3. A New Approach for Dealing With Unobserved Factors in Fixed-T Panels

Let $\boldsymbol{F}_e$ denote a $[T \times L_e]$ dimensional matrix with $L_e \geq L$, such that $\boldsymbol{F} \in \operatorname{Col}(\boldsymbol{F}_e)$. Furthermore, let $\widehat{\boldsymbol{F}}_e$ be a consistent estimator of the column space of $\boldsymbol{F}_e$. In Section 3.1, we derive the asymptotic properties of the proposed GMM estimator based on $\widehat{\boldsymbol{F}}_e$, assuming that the model is identified from Equation (3). In Sections 3.2 and 3.3, we discuss different methods for constructing $\widehat{\boldsymbol{F}}_e$, depending on the model at hand. Finally, in Sections 3.4 and 3.5, we analyze identification, and we put forward two alternative procedures for implementing our approach in practice.

3.1. The Estimator

Assumption 3.1 ensures that $\widehat{\boldsymbol{F}}_e$ is an appropriate plug-in estimator of $\boldsymbol{F}_e$. Our setup is sufficiently general in that it allows for two important cases, namely: (i) the number of estimated factors is larger than the true number of factors that enter into the error term of the equation for y; (ii) the number of factor proxies in $\widehat{\boldsymbol{F}}_e$ that is required to identify $\boldsymbol{F}$ is larger than L. These cases are illustrated in Examples 1 (or 3), and 2, respectively.

Assumption 3.1. Factor proxies are asymptotically linear such that $\sqrt{N}\left(\widehat{\boldsymbol{F}}_e - \boldsymbol{F}_e\boldsymbol{A}_N\right) = \left(N^{-1/2}\sum_{i=1}^{N}\boldsymbol{\Psi}_i\right) + o_P(1)$ and $\boldsymbol{\Psi}_i = (\boldsymbol{I}_T \otimes \boldsymbol{\psi}_i')\boldsymbol{B}_N'$. Here $\boldsymbol{A}_N$ is an $[L_e \times L_e]$ rotation matrix, $\boldsymbol{\psi}_i$ is a $[q \times 1]$ vector, and $\boldsymbol{B}_N$ is a $[L_e \times Tq]$ selection matrix. Furthermore,

1. $\boldsymbol{\psi}_i$ is identically distributed and independent across i, conditional on $\mathcal{F}$. Moreover, $\psi_i^{(\cdot)}$ in $\boldsymbol{\psi}_i = (\psi_i^{(1)}, \ldots, \psi_i^{(q)})'$ satisfies $\mathrm{E}\left[\left|\psi_i^{(\cdot)}\right|^{4+\delta}\right] < \infty$, with $\mathrm{E}_{\mathcal{F}}[\boldsymbol{\psi}_i] = \boldsymbol{0}_q$.
2. $\boldsymbol{A}_N$ and $\boldsymbol{B}_N$ are such that $\boldsymbol{A}_N \xrightarrow{p} \boldsymbol{A}$ and $\boldsymbol{B}_N \xrightarrow{p} \boldsymbol{B}$. Here $\boldsymbol{A}, \boldsymbol{B}$ are $\mathcal{F}$-measurable.
3. For any value of N, including $N \to \infty$: $\operatorname{rk}(\boldsymbol{A}_N) = L_e$ a.s.
4. $\boldsymbol{F} \in \operatorname{Col}(\boldsymbol{F}_e)$ and $\operatorname{rk}(\boldsymbol{F}_e) = L_e$ a.s.

Assumption 3.1 is fairly intuitive. In particular, Assumptions 3.1.1 and Assumption 3.1.2 are employed so as to enable the application of a standard central limit theorem. Notice that, similarly to $\boldsymbol{Z}_i'$, $\boldsymbol{\Psi}_i'$ has a Kronecker product form. Thus, $\boldsymbol{B}_N$ is a selection matrix which, for each point t (i.e., for each $\widehat{\boldsymbol{f}}_{t,e}$), selects those elements of $\boldsymbol{\psi}_i$ that are of first-order ($\sqrt{N}$) importance. Therefore, the length of $\boldsymbol{\psi}_i$ need not be equal to TL_e, the number of elements in $\widehat{\boldsymbol{F}}_e$. Section 3.3 provides more details. Assumptions 3.1.3 and Assumption 3.1.4 ensure that the factor proxies in $\widehat{\boldsymbol{F}}_e$ asymptotically identify the L_e-dimensional column space of $\boldsymbol{F}_e$ a.s. Finally, if Assumption 3.1.4 is violated such that $\boldsymbol{F} \notin \operatorname{Col}(\boldsymbol{F}_e)$, then in the limit $\widehat{\boldsymbol{F}}_e$ will not approximate $\boldsymbol{F}$ in the model equation.

Let $\boldsymbol{\beta}_0$ denote the true value of $\boldsymbol{\beta}$, and $\boldsymbol{g}_0 \equiv \operatorname{vec}(\boldsymbol{G}_{z,\lambda_e}(\boldsymbol{A}_N^{-1})')$, which is of dimension $[dL_e \times 1]$. That is, we define $\boldsymbol{G}_{z,\lambda_e}$ with respect to the *extended* $[L_e \times 1]$ vector of factor loadings $\boldsymbol{\lambda}_{i,e}$, where $\boldsymbol{\lambda}_{i,e} \equiv \boldsymbol{R}\boldsymbol{\lambda}_i$, with $\boldsymbol{R}$ being the selection matrix of the form $\boldsymbol{F} = \boldsymbol{F}_e\boldsymbol{R}$ (the existence of $\boldsymbol{R}$ is guaranteed by part (d)). Moreover, denote by $\boldsymbol{\theta} = \left(\boldsymbol{\beta}', \boldsymbol{g}'\right)'$ the full parameter vector, and the corresponding true parameter vector by $\boldsymbol{\theta}_0$. It is worth mentioning that $\boldsymbol{G}_{z,\lambda_e}$ and $\boldsymbol{A}_N$ cannot be separately identified due to the usual rotation problem in factor models.

Using the plug-in principle and replacing $\boldsymbol{F}$ by $\widehat{\boldsymbol{F}}_e$ in Equation (3), we define the following set of ζ estimating equations for $\boldsymbol{\theta}$:

$$\overline{\boldsymbol{\mu}}_N(\boldsymbol{\theta}) \equiv \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Z}_i'\left(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right) - \boldsymbol{S}\left(\widehat{\boldsymbol{F}}_e \otimes \boldsymbol{I}_d\right)\boldsymbol{g}. \tag{4}$$

The GMM estimator is defined as the minimizer of the following objection function:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \overline{\boldsymbol{\mu}}_N(\boldsymbol{\theta})'\boldsymbol{\Omega}_N\overline{\boldsymbol{\mu}}_N(\boldsymbol{\theta}), \tag{5}$$

where $\boldsymbol{\Omega}_N$ is some prespecified positive definite matrix. Notice that $\overline{\boldsymbol{\mu}}_N(\boldsymbol{\theta})$ is linear in $\boldsymbol{\theta}$, and thus the estimator $\widehat{\boldsymbol{\theta}}$ has a closed form solution.

The asymptotic distribution of the GMM estimator is determined primarily by the leading term in Equation (4). In particular, $\overline{\boldsymbol{\mu}}_N(\boldsymbol{\theta})$ can be expanded as follows:

$$\overline{\boldsymbol{\mu}}_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\mu}_i(\boldsymbol{\theta}) + o_P(N^{-1/2}), \tag{6}$$

where

$$\boldsymbol{\mu}_i(\boldsymbol{\theta}) = \boldsymbol{Z}_i'\left(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right) - \boldsymbol{S}\left((\boldsymbol{F}_e\boldsymbol{A}_N + \boldsymbol{\Psi}_i) \otimes \boldsymbol{I}_d\right)\boldsymbol{g}. \quad (7)$$

The following assumption imposes appropriate regularity conditions on $\overline{\boldsymbol{\mu}}_N(\boldsymbol{\theta})$.

Assumption 3.2.

1. $\boldsymbol{\Gamma}_{\boldsymbol{\beta}} \equiv \text{plim}_{N\to\infty} - \left[\partial\overline{\boldsymbol{\mu}}_N(\boldsymbol{\theta})/\partial\boldsymbol{\beta}'\right] = E_{\mathcal{F}}\left[\boldsymbol{Z}_i'\boldsymbol{X}_i\right]$ is $\mathcal{F}$-measurable and has full column rank a.s.
2. $\boldsymbol{\Gamma}_{\boldsymbol{g}} \equiv \text{plim}_{N\to\infty} - \left[\partial\overline{\boldsymbol{\mu}}_N(\boldsymbol{\theta})/\partial\boldsymbol{g}'\right] = \boldsymbol{S}(\boldsymbol{F}_e\boldsymbol{A} \otimes \boldsymbol{I}_d)$ is $\mathcal{F}$-measurable and has full column rank a.s.
3. The full-parameter Jacobian matrix $\boldsymbol{\Gamma} \equiv \left(\boldsymbol{\Gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Gamma}_{\boldsymbol{g}}\right)$ has full column rank a.s.
4. $\boldsymbol{\Delta} \equiv \text{plim}_{N\to\infty} N^{-1}\sum_{i=1}^N \boldsymbol{\mu}_i(\boldsymbol{\theta}_0)\boldsymbol{\mu}_i(\boldsymbol{\theta}_0)'$ is $\mathcal{F}$-measurable and has full column rank a.s.

It is worth noting that Assumptions 2.1 and 3.1 are sufficient to ensure convergence in probability of the matrices defined in Assumption 3.2. Thus, the only nontrivial restrictions imposed in Assumption 3.2 are the rank restrictions. Assumption 3.2.1 is a standard identification condition in IV estimation and requires the instruments to be correlated with the regressors. Assumption 3.2.2 requires that $\boldsymbol{F}_e\boldsymbol{A}_N$ has full column rank, which is already implied by Assumption 3.1.2. Violations of this restriction are examined in Section 3.4. Assumption 3.2.3 assumes that $\boldsymbol{\Gamma}_{\boldsymbol{\beta}}$ and $\boldsymbol{\Gamma}_{\boldsymbol{g}}$ are linearly independent. That is, $E_{\mathcal{F}}\left[\boldsymbol{Z}_i'\boldsymbol{X}_i\right]$ cannot lie in the column space of $\boldsymbol{F}_e$. As an example, Assumption 3.2.3 excludes situations where $\boldsymbol{Z}_i$ and/or $\boldsymbol{X}_i$ have degenerate idiosyncratic components with variance that is local-to-zero. Essentially, Assumption 3.2.3 is the GMM analogue of the generalized non-collinearity condition of least-squares based factor estimates, as per Bai (2009) and Moon and Weidner (2015). Lastly, Assumption 3.2.4 is also a standard condition and ensures that point-identified inference is asymptotically valid.

Remark 1. The Jacobian matrix $\boldsymbol{\Gamma}$ has ζ rows and $K + dL_e$ columns. Therefore, Assumption 3.2 requires that $\zeta \geq K + dL_e$. To illustrate the meaning of this requirement, suppose that all elements of $\boldsymbol{z}_i$ are strictly exogenous, such that the largest possible set of (internal) instruments is given by $\zeta = dT$. A necessary condition for identification is that $d(T - L_e) \geq K$, which means that the number of factor proxies, L_e, should be strictly smaller than the number of time periods, T. Similar conclusions apply when some of the elements in $\boldsymbol{z}_i$ are weakly exogenous or endogenous, except in this case $\boldsymbol{g}$ is not identifiable without additional normalizations, see Section 3.3 in Juodis and Sarafidis (2018) for details.

The following theorem summarizes the properties of the proposed estimator.

Theorem 1. Suppose that Assumptions 2.1, 3.1, and 3.2 hold true. Then for $N \to \infty$,

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \left[\left(\boldsymbol{\Gamma}'\boldsymbol{\Omega}\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Omega}\right]\boldsymbol{\Delta}^{1/2}\boldsymbol{\pi} \quad (\mathcal{F}\text{-stably}), \quad (8)$$

where $\boldsymbol{\Omega}$ is some $\mathcal{F}$-measurable matrix such that $\text{plim}_{N\to\infty} \boldsymbol{\Omega}_N = \boldsymbol{\Omega}$, while $\boldsymbol{\Gamma}$, $\boldsymbol{\Omega}$, and $\boldsymbol{\Delta}$ are independent of $\boldsymbol{\pi} \sim N(\boldsymbol{0}_\zeta, \boldsymbol{I}_\zeta)$.

Proof. See the supplementary appendix. □

Theorem 1 adopts the notion of $\mathcal{C}$-stable convergence, introduced by Kuersteiner and Prucha (2013), and characterizes convergence as $\mathcal{F}$-stable. Hence, the GMM estimator $\widehat{\boldsymbol{\theta}}$ is consistent, and asymptotically mixed-normal. While this is an important distinction between the properties of the proposed estimator and those of Robertson and Sarafidis (2015) (who treated factors as fixed), it plays no role for inference procedures based on standardized statistics, as long as $\boldsymbol{\Gamma}$ and $\boldsymbol{\Delta}$ can be consistently estimated from their sample analogues. The result of Theorem 1 is general enough to establish general stable convergence, but we present the result as $\mathcal{F}$-stable because we wish to emphasize measurability of all random matrices with respect to $\mathcal{F}$.

As our focus lies on asymptotically linear estimators of $\boldsymbol{F}_e$, consistent estimation of $\boldsymbol{\Delta}$ requires a plug-in estimator of $\boldsymbol{\Psi}_i$. In particular, if such estimator $\widehat{\boldsymbol{\Psi}}_i$ is available, $\boldsymbol{\Delta}$ can be estimated consistently using the conventional formula

$$\widehat{\boldsymbol{\Delta}} = \frac{1}{N}\sum_{i=1}^N \widehat{\boldsymbol{\mu}}_i\left(\widehat{\boldsymbol{\theta}}\right)\widehat{\boldsymbol{\mu}}_i\left(\widehat{\boldsymbol{\theta}}\right)', \quad (9)$$

where $\widehat{\boldsymbol{\mu}}_i(\boldsymbol{\theta})$ is the feasible plug-in estimate of $\boldsymbol{\mu}_i(\boldsymbol{\theta})$, that is,

$$\widehat{\boldsymbol{\mu}}_i(\boldsymbol{\theta}) = \boldsymbol{Z}_i'\left(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right) - \boldsymbol{S}\left((\widehat{\boldsymbol{F}}_e + \widehat{\boldsymbol{\Psi}}_i) \otimes \boldsymbol{I}_d\right)\boldsymbol{g}. \quad (10)$$

Finally, the usual two-step GMM estimator can be obtained by setting $\boldsymbol{\Omega}_N = \widehat{\boldsymbol{\Delta}}^{-1}$.

3.2. Construction of Factor Proxies, $\widehat{\boldsymbol{F}}_e$

This section puts forward two specific methods for constructing $\widehat{\boldsymbol{F}}_e$. These methods are motivated by our empirical application as well as common practice in the large-T panel data literature.

3.2.1. Method I: One Variable and Multiple Weights

Suppose there exists a single variable $\mathbf{v}_i = \left(\mathrm{v}_{i,1}, \ldots, \mathrm{v}_{i,T}\right)'$ driven by $\boldsymbol{F}$, as well as possibly additional factors. That is, $\mathbf{v}_i$ satisfies

$$\mathbf{v}_i = \boldsymbol{F}_e\boldsymbol{\gamma}_i + \boldsymbol{u}_i, \quad (11)$$

where $\boldsymbol{F}_e$ is $[T \times L_e]$ with $L_e \geq L$, such that $\boldsymbol{F} \in \text{Col}(\boldsymbol{F}_e)$. $\mathbf{v}_i$ can be either internal, that is, one of the regressors, or external, in the spirit of Pesaran, Smith, and Yamagata (2013). The existence of such variable is quite plausible in panel data models (see, e.g., Hansen and Liao 2019; Karabıyık, Urbain, and Westerlund 2019) because economic agents inhabit common economic environments and therefore many variables are often subject to common shocks, such as changes in technology and productivity, changes in preferences and tastes, and so on.

Let $\boldsymbol{w}_i$ denote an $[L_e \times 1]$ vector of individual-specific weights, such that $E_{\mathcal{F}}\left[\boldsymbol{u}_i\boldsymbol{w}_i'\right] = \mathbf{O}_{T\times L_e}$ and $E_{\mathcal{F}}\left[\boldsymbol{\gamma}_i\boldsymbol{w}_i'\right] = \boldsymbol{G}_{\gamma,w}$ a.s. In terms of the general notation employed in Assumption 3.1, this setup corresponds to setting $\boldsymbol{A} = \boldsymbol{G}_{\gamma,w}$ and $\boldsymbol{\Psi}_i = \boldsymbol{u}_i\boldsymbol{w}_i' + \boldsymbol{F}_e(\boldsymbol{\gamma}_i\boldsymbol{w}_i' - \boldsymbol{G}_{\gamma,w})$. Thus, Assumption 3.1 translates into the requirement that $\text{rk}\left(\boldsymbol{G}_{\gamma,w}\right) = L_e$. In this case,

$$\widehat{\boldsymbol{F}}_e = \frac{1}{N}\sum_{i=1}^N \mathbf{v}_i\boldsymbol{w}_i', \quad (12)$$

is a suitable estimator of $\boldsymbol{F}_e$. Furthermore, the corresponding plug-in estimator of $\widehat{\boldsymbol{\Psi}}_i$ is simply given by

$$\widehat{\boldsymbol{\Psi}}_i = \mathbf{v}_i\boldsymbol{w}_i' - \widehat{\boldsymbol{F}}_e. \tag{13}$$

It is apparent that Method I permits cases where the L_e factor proxies in $\widehat{\boldsymbol{F}}_e$ identify more factors than those entering into the error term of the equation for y. Such generality is appealing because $\mathbf{v}_i$ may contain more factors than those that already drive $y_{i,t}$.

Example 1. Suppose that $L = 1$, such that the factor component in Equation (2) reduces to $\boldsymbol{f}^{(1)}\lambda_i^{(1)}$, but $\mathbf{v}_i$ is driven by two factors, that is, $\boldsymbol{F}_e = (\boldsymbol{f}^{(1)},\boldsymbol{f}^{(2)})$ and $\boldsymbol{w}_i$ is a $[2 \times 1]$ vector. Then,

$$\widehat{\boldsymbol{F}}_e = \frac{1}{N}\sum_{i=1}^{N}\mathbf{v}_i\boldsymbol{w}_i' \xrightarrow{p} \boldsymbol{F}_e\boldsymbol{A}, \tag{14}$$

where $\boldsymbol{A} = \mathrm{E}_{\mathcal{F}}[\boldsymbol{\gamma}_i\boldsymbol{w}_i']$. Consistent estimability of $\boldsymbol{\beta}$ requires that $\mathrm{rk}(\boldsymbol{A}) = 2$. Notice that if one makes use of one weight only, w_i, $\boldsymbol{A} = \mathrm{E}_{\mathcal{F}}[\boldsymbol{\gamma}_i w_i]$ has at most rank 1. Thus, in this case $\widehat{\boldsymbol{F}}_e$ cannot estimate $\boldsymbol{f}^{(1)}$ consistently, unless $\mathrm{E}_{\mathcal{F}}[\gamma_i^{(2)}w_i] = 0$, where $\gamma_i^{(2)}$ denotes the bottom entry in $\boldsymbol{\gamma}_i$. Section 3.4 discusses in detail the situation where the full rank restriction on $\boldsymbol{A}$ is violated.

The aforedescribed method requires two distinct ingredients: $\mathbf{v}_i$ and $\boldsymbol{w}_i$. There are several potential choices for $\boldsymbol{w}_i$. Consider initially the case $L_e = L = 1$ and suppose that the factor component in Equation (11) reduces to $\boldsymbol{f}^{(1)}\gamma_i^{(1)}$. One simple choice is to set the value of the weight across all individuals equal to a fixed constant, that is, $w_i = 1$. In this case, Equation (12) becomes the cross-sectional average of $\mathbf{v}_i$, whereas $\boldsymbol{G}_{\gamma,w}$ reduces to $\mu_\gamma \equiv \mathrm{E}_{\mathcal{F}}\left[\gamma_i^{(1)}\right]$. Thus, Assumption 3.1 implies that $\mu_\gamma \neq 0$ and $\mathrm{E}_{\mathcal{F}}[\boldsymbol{u}_i] = \boldsymbol{0}_T$. An alternative choice, and one which is especially appealing in autoregressive models, is to use random weights that are independent across i conditional on $\mathcal{F}$, such as lagged values of the observed data. For instance, in ARDL(p,q) models a natural choice is to set $w_i = y_{i,0}$ (or $w_i = x_{i,0}^{(k)}$), which requires $\mathrm{E}_{\mathcal{F}}[\gamma_i^{(1)}y_{i,0}] \neq 0$ (or $\mathrm{E}_{\mathcal{F}}[\gamma_i^{(1)}x_{i,0}^{(k)}] \neq 0$), as in the "correlated random effects" framework.

When $L_e > 1$, the use of a single weight would violate Assumption 3.1, as was pointed out in Example 1. Thus, a possible "automated" strategy for constructing the $[L_e \times 1]$ dimensional vector $\boldsymbol{w}_i$ is to pick the first $L_e - 1$ observations of a single regressor, such that $\boldsymbol{w}_i = \left(1, \boldsymbol{x}_{i,1:L-1}^{(k)}\right)$ for some k. Alternatively, one can rely upon the initial condition of dependent and independent variables, that is, $\boldsymbol{w}_i = \left(1, y_{i,0}, x_{i,0}^{(1)}, \ldots, x_{i,0}^{(L_e-2)}\right)'$. Lastly, one can use powers of $y_{i,0}$ or $x_{i,0}^{(k)}$, such as $y_{i,0}^2$ and so on. This option does not require the distribution of $y_{i,t}$ to be symmetric because noncentral moments are used. To illustrate the aforementioned strategies, let $K = 2$, $L = 3$, $x_{i,t}^{(1)} \equiv y_{i,t-1}$, and let $x_{i,t}^{(2)} \equiv x_{i,t}$ be treated as weakly exogenous. For this specification, one could set $\boldsymbol{w}_i = (1, y_{i,0}, y_{i,1})'$, $\boldsymbol{w}_i = (1, y_{i,0}, x_{i,0})'$, or $\boldsymbol{w}_i = (1, y_{i,0}, y_{i,0}^2)'$.

Since the parameters of interest can be estimated based on different weighting schemes, the approach proposed in this article provides a more flexible way of dealing with unobserved factors compared to other methods that are available in the literature. The practical question of how to actually select among different values of $\boldsymbol{w}_i$ (and/or different v's) is discussed in Section 3.5.

Remark 2. Independently from this research, Gagliardini and Gouriéroux (2017) and Fan and Liao (2019) have recently advocated a similar construction of factor proxies, which involves prespecified (potentially arbitrary) weights $\boldsymbol{w}_i$. Unlike our study, the prime focus of those studies lies on the asymptotic properties of factor estimates when both N and T are large.

3.2.2. Method II: Multiple Variables, Single Weight

The second approach involves the construction of factor proxies based on a prespecified *single* weighting scheme, w_i, and L_e distinct time-varying variables. As before, these variables can be either internal or external. In particular, let $\boldsymbol{V}_i = (\mathbf{v}_i^{(1)}, \ldots, \mathbf{v}_i^{(L_e)})$ be a $[T \times L_e]$ matrix, such that

$$\boldsymbol{V}_i = \boldsymbol{F}_e\boldsymbol{\Upsilon}_i + \boldsymbol{U}_i, \tag{15}$$

where $\boldsymbol{\Upsilon}_i = \left(\boldsymbol{\gamma}_i^{(1)}, \ldots, \boldsymbol{\gamma}_i^{(L_e)}\right)$ and $\boldsymbol{U}_i = \left(\boldsymbol{u}_i^{(1)}, \ldots, \boldsymbol{u}_i^{(L_e)}\right)$. In this case, $\widehat{\boldsymbol{F}}_e$ is defined as

$$\widehat{\boldsymbol{F}}_e = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{V}_i w_i = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{F}_e\boldsymbol{\Upsilon}_i w_i + \boldsymbol{U}_i w_i). \tag{16}$$

Notice that the use of a single weight requires that such weight is valid for each column of $\boldsymbol{V}_i$, that is, $\mathrm{E}_{\mathcal{F}}\left[\boldsymbol{U}_i w_i\right] = \boldsymbol{0}_{T\times L_e}$. However, the above formulation does not imply that each column in $\boldsymbol{V}_i$ must be driven by the full set of columns in $\boldsymbol{F}_e$. For instance, let $L = L_e = 2$, $\boldsymbol{F} = \boldsymbol{F}_e = (\boldsymbol{f}^{(1)},\boldsymbol{f}^{(2)})$ with $\mathbf{v}_i^{(1)} = \boldsymbol{f}^{(1)}\boldsymbol{\gamma}_i^{(1)} + \boldsymbol{u}_i^{(1)}$ and $\mathbf{v}_i^{(2)} = \boldsymbol{f}^{(2)}\boldsymbol{\gamma}_i^{(2)} + \boldsymbol{u}_i^{(2)}$. This structure can be represented by Equation (15) using a diagonal $\boldsymbol{\Upsilon}_i$ matrix.

We note that the setup in Method II is sufficiently general in that it accommodates cases where the number of factor proxies required to approximate $\boldsymbol{F}$ is larger than L itself. The following example illustrates this point.

Example 2. Consider an autoregressive, AR(1), model with $L = 1$:

$$y_{i,t} = \alpha y_{i,t-1} + \lambda_i f_t + \varepsilon_{i,t}; \quad t = 1, \ldots, T, \tag{17}$$

or, in vector form,

$$\boldsymbol{y}_i = \alpha\boldsymbol{y}_{i,-1} + \boldsymbol{f}\lambda_i + \boldsymbol{\varepsilon}_i, \tag{18}$$

where $\boldsymbol{y}_{i,-1} = (y_{i,0}, \ldots, y_{i,T-1})'$, $\boldsymbol{f} = (f_1, \ldots, f_T)'$ and $\boldsymbol{y}_i$ is defined in Equation (2). As it is shown in Everaert and De Groote (2016) and Juodis, Karabıyık, and Westerlund (2020), the CCE-style factor proxies, $\widehat{\boldsymbol{F}}_e = (\overline{\boldsymbol{y}}, \overline{\boldsymbol{y}}_{-1})$, which arise in our context by setting $\boldsymbol{V}_i = (\boldsymbol{y}_i, \boldsymbol{y}_{i,-1})$ and $w_i = 1$, satisfy Assumption 3.1 provided that $\mathrm{E}_{\mathcal{F}}[\lambda_i] \neq 0$. In this setup, none of the two factor proxies in $\widehat{\boldsymbol{F}}_e$ alone is able to estimate $\boldsymbol{f}$ consistently; instead only a linear combination of both columns can be estimated consistently. In particular, if we define $\boldsymbol{F}_e \equiv \mathrm{E}_{\mathcal{F}}[\boldsymbol{V}_i]$, then $\boldsymbol{F}_e = (\boldsymbol{f}_e^{(1)},\boldsymbol{f}_e^{(2)})$ with $\boldsymbol{f}_e^{(1)} = \alpha\boldsymbol{f}_e^{(2)} + \mathrm{E}_{\mathcal{F}}[\lambda_i]\boldsymbol{f}$, while $\boldsymbol{f}_e^{(2)}$ is a cumulative function of the lags of $\boldsymbol{f}$ and the initial condition, that is, $f_{1,e}^{(2)} = \mathrm{E}_{\mathcal{F}}[y_{i,0}]$, and $f_{t,e}^{(2)} = \mathrm{E}_{\mathcal{F}}[\lambda_i]f_{t-1} + \alpha f_{t-1,e}^{(2)}$ for

$t = 2, \ldots, T$. Notice that in this case $\boldsymbol{R} = (1, -\alpha)'/\,\mathrm{E}_{\mathcal{F}}[\lambda_i]$, and the extended vector of factor loadings $\boldsymbol{\lambda}_{i,e}$ is defined accordingly. As has been discussed, alternative choices for $w_i = 1$ do exist. For instance, when $\mathrm{E}_{\mathcal{F}}[\lambda_i] = 0$, a plausible strategy is to set $w_i = y_{i,0}$ and drop the first observation from estimation of the AR(1) model in Equation (17).

Remark 3. Method II is motivated by, although it is not nested within, the seminal CCE approach of Pesaran (2006). In particular, CCE considers deterministic weights (mostly $w_i = 1$), and assumes that *all* K regressors, $\boldsymbol{X}_i$, are strictly exogenous with respect to $\boldsymbol{\varepsilon}_i$. In contrast, here w_i can be stochastic, as Example 2 points out. Moreover, not all regressors need to be strictly exogenous. Finally, the restriction $\mathrm{E}_{\mathcal{F}}[\lambda_i] \neq 0$ is testable within our framework, see Section 3.4. In the supplementary appendix, we study a stylized setup in which both the proposed linear GMM approach and CCE are consistent for T fixed; specifically, we focus on a model with one strictly exogenous regressor, which is driven by the same (single) factor that enters into the error term of the equation for y, such that $\mathrm{E}_{\mathcal{F}}[\gamma_i] \neq 0$. We demonstrate that under spherical error components, the CCE estimator and the linear GMM estimator based on $w_i = 1$ have the same asymptotic variance.

Remark 4. Depending on the model at hand, one can also combine multiple time-varying variables, $\boldsymbol{V}_i$, with multiple time-invariant weights, $\boldsymbol{w}_i$, to construct $\widehat{\boldsymbol{F}}_e$. As an example, let $\boldsymbol{V}_i = (\mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)})$ and $\boldsymbol{w}_i = \left(w_i^{(1)}, w_i^{(2)}\right)'$, in which case potential factor proxies can be constructed by using all possible combinations of $\boldsymbol{V}_i$ and $\boldsymbol{w}_i$, that is,

$$\widehat{\boldsymbol{F}}_e = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{V}_i \otimes \boldsymbol{w}_i'. \tag{19}$$

Alternatively, one can consider only a subset of available proxies. For instance,

$$\widehat{\boldsymbol{F}}_e = \frac{1}{N}\sum_{i=1}^{N} (\mathbf{v}_i^{(1)} w_i^{(1)}, \mathbf{v}_i^{(2)} w_i^{(2)}). \tag{20}$$

This flexibility provides a further advantage of the factor-proxies formulation used in this article over that of Pesaran (2006). The practical question of how to actually select among different values of $\boldsymbol{w}_i$ (and/or different values of $\mathbf{v}_i$) is discussed in Section 3.5.

Irrespective of the method that is considered to construct factor proxies, notice that $\widehat{\boldsymbol{F}}_e$ can be expressed as

$$\widehat{\boldsymbol{F}}_e = \boldsymbol{F}_e \boldsymbol{A} + \overline{\boldsymbol{\Psi}}; \quad \overline{\boldsymbol{\Psi}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\Psi}_i, \tag{21}$$

where here $\boldsymbol{\Psi}_i$ is defined as

$$\boldsymbol{\Psi}_i = \left(\boldsymbol{F}_e(\boldsymbol{A}_i - \boldsymbol{A}) + \boldsymbol{E}_i\right), \tag{22}$$

for some $\boldsymbol{A}_i$. Thus, $\widehat{\boldsymbol{F}}_e$ is a linear estimator of the column space of $\boldsymbol{F}_e$. In what follows we put forward an estimator that is only asymptotically linear, but not linear for fixed N and T.

3.3. Regularized Factor Proxies

This section analyzes the important case where the model is "fundamentally identified," but the practitioner includes more weights, or more variables than necessary in the approximation of $\boldsymbol{F}_e$. In the context of Assumption 3.1, this implies that Assumption 3.1.3 is violated but Assumption 3.1.4 is not.

To formalize this idea, let $R \geq L_e$ be the total number of factor proxies and $\widehat{\boldsymbol{F}}_R$ denote the corresponding factor proxies with dimension $[T \times R]$. Following Equation (21), irrespective of the method considered to construct factor proxies, $\widehat{\boldsymbol{F}}_R$ can be decomposed as

$$\widehat{\boldsymbol{F}}_R = \boldsymbol{F}_e \boldsymbol{A}_R + \overline{\boldsymbol{\Psi}}^{(R)}; \quad \overline{\boldsymbol{\Psi}}^{(R)} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\Psi}_i^{(R)}, \tag{23}$$

where $\overline{\boldsymbol{\Psi}} = (\overline{\boldsymbol{\psi}}_1^{(R)}, \ldots, \overline{\boldsymbol{\psi}}_T^{(R)})'$ is $[T \times R]$, and $\boldsymbol{A}_R$ is $[L_e \times R]$ with rank L_e. As a result, $\mathrm{rk}(\mathrm{E}_{\mathcal{F}}[\widehat{\boldsymbol{F}}_R]) = L_e$ and $\boldsymbol{\Gamma}_{\boldsymbol{g}}$, defined in Assumption 3.2, is of reduced rank.

Example 3. Consider the following model in vector form:

$$\boldsymbol{y}_i = \beta_1 \boldsymbol{x}_i^{(1)} + \beta_2 \boldsymbol{x}_i^{(2)} + \boldsymbol{f}^{(1)} \lambda_i^{(1)} + \boldsymbol{\varepsilon}_i, \tag{24}$$

where $\boldsymbol{f}^{(1)}$ is $[T \times 1]$. Thus, $K = 2$ and $L = 1$. Let

$$\boldsymbol{x}_i^{(1)} = \boldsymbol{f}^{(1)} \pi_i + \boldsymbol{u}_i^{(1)}, \tag{25}$$

and $L_e = 2$, such that

$$\boldsymbol{x}_i^{(2)} = \boldsymbol{F}_e \boldsymbol{\xi}_i + \boldsymbol{u}_i^{(2)}, \tag{26}$$

where $\boldsymbol{F}_e = \left(\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)}\right)$ is of order $[T \times 2]$, and $\boldsymbol{\xi}_i$ is a vector of order $[2 \times 1]$. One possible strategy is to proxy $\boldsymbol{F}_e$ using the full set of observables, by setting $\boldsymbol{V}_i = \left(\boldsymbol{x}_i^{(1)}, \boldsymbol{x}_i^{(2)}, \boldsymbol{y}_i\right)$ and $w_i = 1$. This is standard practice in CCE estimation, with the aim to avoid the need for estimating the number of factors in the model. In terms of the notation used in Equation (15), we have $\boldsymbol{\Upsilon}_i = \left(\boldsymbol{\gamma}_i^{(1)}, \boldsymbol{\gamma}_i^{(2)}, \boldsymbol{\gamma}_i^{(3)}\right)$, which is a $[2 \times 3]$ matrix with $\boldsymbol{\gamma}_i^{(1)} = (\pi_i, 0)'$, $\boldsymbol{\gamma}_i^{(2)} = \boldsymbol{\xi}_i$ and $\boldsymbol{\gamma}_i^{(3)} = \boldsymbol{\lambda}_i + \beta_1 \boldsymbol{\gamma}_i^{(1)} + \beta_2 \boldsymbol{\xi}_i$, where $\boldsymbol{\lambda}_i = (\lambda_i^{(1)}, 0)'$. In this case, $\mathrm{E}_{\mathcal{F}}[\boldsymbol{\Upsilon}_i w_i]$ does not have full column rank. Essentially, too many variables are used for approximating $\boldsymbol{F}_e$. In this section, we put forward a method that overcomes this problem within our GMM approach.

Under these circumstances, it is straightforward to show that the GMM estimator considered thus far remains consistent. However, it turns out that the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ can be highly nonstandard, due to a degeneracy of the $\boldsymbol{g}$ estimates. For this reason, it is essential to use factor proxies that are nondegenerate, that is, none of the columns of $\widehat{\boldsymbol{F}}_e$ are asymptotically collinear.

In what follows, we put forward a regularization approach for constructing $\widehat{\boldsymbol{F}}_e$ such that Assumptions 3.1.1–Assumption 3.1.3 of Assumption 3.1 are satisfied. Our regularization method uses the singular value decomposition of $\widehat{\boldsymbol{F}}_R$, or, equivalently, the principal components of $\widehat{\boldsymbol{F}}_R \widehat{\boldsymbol{F}}_R'$.

To begin with, let $\boldsymbol{\Lambda}_N$ be a $[L_e \times L_e]$ diagonal matrix containing the L_e largest eigenvalues of $T^{-1}\widehat{\boldsymbol{F}}_R \widehat{\boldsymbol{F}}_R'$ in descending order. The following assumption ensures that the asymptotic distribution of the proposed regularized estimator is well defined (see, e.g., Bai 2003 for a similar condition):

Assumption 3.3. The eigenvalues of the $[L_e \times L_e]$ matrix $(\boldsymbol{A}_R\boldsymbol{A}_R')\left(\boldsymbol{F}_e'\boldsymbol{F}_e\right)$ are distinct a.s.

Let $\widetilde{\boldsymbol{F}} = \sqrt{T}\widehat{\boldsymbol{U}}_{L_e}$ denote the scaled regularized, principal components (PC) estimator for $\boldsymbol{F}_e$, where $\widehat{\boldsymbol{U}}_{L_e}$ denotes the associated eigenvectors (left singular vectors) corresponding to the L_e largest eigenvalues of $T^{-1}\widehat{\boldsymbol{F}}_R\widehat{\boldsymbol{F}}_R'$. The fixed T consistency of the PC factor estimator $\widetilde{\boldsymbol{F}}$, which is of dimension $[T \times L_e]$, follows intuitively from the results in Connor and Korajczyk (1986) and Bai (2003). As it is emphasized in Bai (2003), a necessary and sufficient condition for fixed T consistency is that $\widehat{\boldsymbol{F}}_R\widehat{\boldsymbol{F}}_R' \xrightarrow{p} \boldsymbol{F}_e\boldsymbol{M}\boldsymbol{F}_e' + \kappa\boldsymbol{I}_T$, for some matrix $\boldsymbol{M}$ and some scalar $\kappa \geq 0$. In our case, this necessary condition is satisfied trivially with $\kappa = 0$. However, a new proof for the asymptotic distribution of $\widetilde{\boldsymbol{F}} = (\tilde{\boldsymbol{f}}_1, \dots, \tilde{\boldsymbol{f}}_T)'$ is required when T is fixed. This is achieved by Theorem 2.

Theorem 2. Suppose that Assumption 3.3 holds true, and $\widehat{\boldsymbol{F}}_R$ satisfies Equation (21) together with Assumption 3.1, except that $\mathrm{rk}(\boldsymbol{A}_R) = L_e < R$. Then the regularized factor estimator $\widetilde{\boldsymbol{F}}$ satisfies Assumption 3.1, with

$$\boldsymbol{A}_N = (\boldsymbol{A}_R\boldsymbol{A}_R')\left(\frac{\widetilde{\boldsymbol{F}}'\boldsymbol{F}_e}{T}\right)\boldsymbol{\Lambda}_N^{-1}. \tag{27}$$

Furthermore, each row $\boldsymbol{\Psi}_i^{(t)}$ in $\boldsymbol{\Psi}_i$ is of the following form:

$$\boldsymbol{\Psi}_i^{(t)} = \boldsymbol{\Lambda}_N^{-1}\frac{1}{T}\left(\sum_{s=1}^{T}\tilde{\boldsymbol{f}}_s\left(\boldsymbol{f}_{s,e}'\boldsymbol{A}_R\boldsymbol{\psi}_{i,t}^{(R)} + \boldsymbol{f}_{t,e}'\boldsymbol{A}_R\boldsymbol{\psi}_{i,s}^{(R)}\right)\right); \quad t = 1, \dots, T. \tag{28}$$

Proof. See the supplementary appendix. □

It is straightforward to see from the expression in Equation (28) that the leading variance term of $\widetilde{\boldsymbol{F}}$ is affine in each $\{\boldsymbol{A}_R\boldsymbol{\psi}_{i,t}\}_{t=1}^{T}$. This implies that the inclusion of *uninformative* factor proxies, that is, proxies that do not identify $\boldsymbol{F}_e$, has no impact on the first-order asymptotic properties of $\widetilde{\boldsymbol{F}}$. Essentially, the PC estimator of factors performs estimation of the factor space and factor proxy selection at the same time.

Since $\widetilde{\boldsymbol{F}}$ satisfies Assumption 3.1, it can be used as a plug-in estimator in the GMM objective function. Consistent estimation of the variance-covariance matrix $\boldsymbol{\Delta}$, requires replacing unknown quantities in $\boldsymbol{\Psi}_i$ with their plug-in counterparts. In particular, the $\boldsymbol{f}_{s,e}'\boldsymbol{A}_R$ terms in Equation (28) can be consistently estimated by $(\widehat{\boldsymbol{f}}_s^{(R)})'$. Furthermore, depending on the method used to construct the factor proxies, the plug-in counterparts of $\boldsymbol{\psi}_{i,t}^{(R)}$ are given by either

$$\widehat{\boldsymbol{\psi}}_{i,t}^{(R)} = \boldsymbol{w}_i \mathrm{v}_{i,t} - \widehat{\boldsymbol{f}}_t^{(R)}, \tag{29}$$

or

$$\widehat{\boldsymbol{\psi}}_{i,t}^{(R)} = \mathbf{v}_{i,t} w_i - \widehat{\boldsymbol{f}}_t^{(R)}, \tag{30}$$

where $\mathbf{v}_{i,t} = (\mathrm{v}_{i,t}^{(1)}, \dots, \mathrm{v}_{i,t}^{(L_e)})'$.

Remark 5. The above strategy for selection of factor proxies is not optimal from the point of view of obtaining a GMM estimator for $\boldsymbol{\beta}_0$ with minimal asymptotic variance. Specifically, the PC estimator is a weighted average of all individual factor proxies, with corresponding weights being determined outside the GMM objective function. One could proxy $\boldsymbol{F}_e$ by combining factor proxies optimally. However, such an approach has a major drawback, in that the resulting moment conditions are nonlinear. To illustrate this, consider a linear combination of $\boldsymbol{w}_i = (w_i^{(1)}, w_i^{(2)})'$ that takes the form $\tilde{w}_i = aw_i^{(1)} + (1-a)w_i^{(2)}$. It is clear that in the absence of knowledge of a, the resulting moment conditions involve products of unknown parameters, namely $\boldsymbol{A}_R$ and a. Hence, the appealing linearity of the proposed approach no longer holds, and instead one may use, for example, the FIVU and FIVR estimators of Robertson and Sarafidis (2015), which are asymptotically more efficient because they involve joint estimation of $\boldsymbol{\beta}$ and $\boldsymbol{F}$ using the full set of moment conditions.

3.4. Identification

Assumption 3.1 plays a major role in characterizing the large sample properties of the proposed GMM estimator. In this section, we discuss several departures from Assumption 3.1, as well as diagnostic checks that can be used to detect these departures. To save space, we focus on factor proxies constructed using multiple $\boldsymbol{w}_i$, as in Equation (12).

At first, consider the case where $\boldsymbol{F} \notin \mathrm{Col}(\boldsymbol{F}_e)$ but otherwise all remaining parts of Assumption 3.1 are satisfied. For instance, suppose that some of the factors in $\boldsymbol{y}_i$ are entirely different from those that drive $\mathbf{v}_i$, or alternatively $L_e < L$. In neither case can the factor proxies approximate the column space of $\boldsymbol{F}$ asymptotically. As a result, it is straightforward to show that the GMM estimator for $\boldsymbol{\beta}_0$ is inconsistent. That is,

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \xrightarrow{p} \left(\boldsymbol{\Gamma}_\beta'\boldsymbol{\Gamma}_\beta - \boldsymbol{\Gamma}_\beta'\boldsymbol{\Gamma}_g\left(\boldsymbol{\Gamma}_g'\boldsymbol{\Gamma}_g\right)^{-1}\boldsymbol{\Gamma}_g'\boldsymbol{\Gamma}_\beta\right)^{-1} \times \left(\boldsymbol{\Gamma}_\beta' - \boldsymbol{\Gamma}_\beta'\boldsymbol{\Gamma}_g\left(\boldsymbol{\Gamma}_g'\boldsymbol{\Gamma}_g\right)^{-1}\boldsymbol{\Gamma}_g'\right)\boldsymbol{S}\,\mathrm{vec}\left(\boldsymbol{G}_{z,\lambda}\boldsymbol{F}'\right), \tag{31}$$

where $\boldsymbol{\Gamma}_\beta$ and $\boldsymbol{\Gamma}_g$ denote the two constituent blocks of the Jacobian matrix of the moment conditions, as defined in Assumption 3.2. Such identification failure can be detected using the usual overidentifying restrictions test, commonly referred to as J-statistic.

Next, consider the case where $\boldsymbol{G}_{\gamma,w}$ is not of full rank, such that $\boldsymbol{F} \notin \mathrm{Col}(\boldsymbol{F}_e\boldsymbol{G}_{\gamma,w})$. As an example, let $L_e = L = 1$ and $w_i = 1$, such that $\mathrm{E}_{\mathcal{F}}(\gamma_i w_i) = \mu_\gamma$. The full rank condition on $\boldsymbol{G}_{\gamma,w}$ is violated when $\mu_\gamma = 0$. More generally, suppose that $\mathrm{rk}(\boldsymbol{G}_{\gamma,w}) = Q < L_e \leq L$ a.s. Let $\boldsymbol{G}_{\gamma,w} = \boldsymbol{C}\boldsymbol{D}'$, where both $\boldsymbol{C}$ and $\boldsymbol{D}$ are $[L_e \times Q]$ matrices of rank Q a.s. Furthermore, let $\boldsymbol{D}_\perp$ denote the orthogonal complement of $\boldsymbol{D}$, that is, $\boldsymbol{D}_\perp$ satisfies $\mathrm{rk}(\boldsymbol{D}_\perp) = L_e - Q$ and $\boldsymbol{D}'\boldsymbol{D}_\perp = \mathbf{O}_{Q\times(L_e-Q)}$. Theorem 3 summarizes the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ when $\boldsymbol{G}_{\gamma,w}$ is rank-deficient.

Theorem 3. Suppose that Assumption 2.1 is satisfied, and consider factor proxies with $\boldsymbol{A}_N = \boldsymbol{G}_{\gamma,w}$ with $\mathrm{rk}(\boldsymbol{G}_{\gamma,w}) = Q <$

$L_e \leq L$. Then for $\boldsymbol{\Omega}_N = \boldsymbol{I}$, as $N \to \infty$:

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \xrightarrow{d} \left(\boldsymbol{\Gamma}'_\beta \boldsymbol{\Gamma}_\beta - \boldsymbol{\Gamma}'_\beta \boldsymbol{\Xi}\left(\boldsymbol{\Xi}'\boldsymbol{\Xi}\right)^{-1}\boldsymbol{\Xi}'\boldsymbol{\Gamma}_\beta\right)^{-1} \times \left(\boldsymbol{\Gamma}'_\beta - \boldsymbol{\Gamma}'_\beta \boldsymbol{\Xi}\left(\boldsymbol{\Xi}'\boldsymbol{\Xi}\right)^{-1}\boldsymbol{\Xi}'\right)\boldsymbol{S}\,\mathrm{vec}\left(\boldsymbol{G}_{z,\lambda}\boldsymbol{F}'\right), \quad (32)$$

$\mathcal{F}$-stably, where $\boldsymbol{\Xi} = \boldsymbol{S}((\boldsymbol{\Xi}_d, \boldsymbol{\Xi}_s) \otimes \boldsymbol{I}_d)$, $\boldsymbol{\Xi}_d = \boldsymbol{F}_e\boldsymbol{C}(\boldsymbol{D}'\boldsymbol{D})$, and $\boldsymbol{\Xi}_s$ is such that $\mathrm{vec}(\boldsymbol{\Xi}_s) \sim \boldsymbol{\Delta}_F^{1/2}\boldsymbol{\pi}_F$ with $\boldsymbol{\Delta}_F = \mathrm{E}_{\mathcal{F}}[\mathrm{vec}(\mathbf{v}_i\boldsymbol{w}_i'\boldsymbol{D}_\perp)\,\mathrm{vec}(\mathbf{v}_i\boldsymbol{w}_i'\boldsymbol{D}_\perp)']$, and $\boldsymbol{\pi}_F \sim N(\boldsymbol{0}_{(L_e-Q)T}, \boldsymbol{I}_{(L_e-Q)T})$.

Proof. See the supplementary appendix. □

It is clear that the GMM estimator converges to a random limit because the $\boldsymbol{\Xi}_s$ matrix is stochastic. This result is in line with existing weak instruments results for IV/GMM estimators (see, e.g., Staiger and Stock 1997). However, a major difference between Theorem 3 and existing literature is that in the usual weak-IV setup, the limit remains random even after conditioning on the (random) Jacobian matrix. In contrast, in the present case, the right-hand side of Equation (32) is a nonzero constant vector, conditional on $\mathcal{F}$ and $\boldsymbol{\pi}_F$. As a result, failure of identification associated with $\mathrm{rk}(\boldsymbol{G}_{\gamma,w}) = Q < L_e \leq L$ a.s. implies model mis-specification. This can be detected using the usual J-statistic again.

Remark 6. The stochastic nature of the identification failure in Theorem 3 can be easily avoided using regularized factor proxies as in Section 3.3, obtained using a consistent estimate of L_e.

3.5. Implementation

We take it as given in the section that a set of potential factor proxies has been collected into a $[T \times R]$ matrix $\widehat{\boldsymbol{F}}_R$, using either Method I, Method II, or a combination of both. For example, under Method II one could construct $\widehat{\boldsymbol{F}}_R$ based on all available observables, that is, the dependent variable, the regressors, as well as possible external variables. Unfortunately, as we discussed in Section 3.3, it turns out that including more variables (or, equivalently for Method I, more weights) than necessary in the approximation of $\boldsymbol{F}_e$, can render the asymptotic distribution of the GMM estimator highly nonstandard. To circumvent this potential issue, in what follows we put forward two distinct methods that practitioners may use to implement the plug-in principle embedded in our approach. Firstly, a "regularization" method, then a "best-subset selection" method. Both methods are illustrated in the empirical section.

3.5.1. Regularization

This approach builds upon the regularized factor proxies analyzed in Section 3.3 and consists of the following steps:

1. Obtain a consistent estimate of the underlying number of factors in $\widehat{\boldsymbol{F}}_R$, given by $\widehat{L}_e$, using either a sequential pivotal rank testing as proposed by Kleibergen and Paap (2006), or the eigenvalue ratio (ER) and the growth ratio (GR) statistics as in Ahn and Horenstein (2013).
2. Use the regularized estimator $\tilde{\boldsymbol{F}}$ as the plug-in estimator for the GMM objective function to estimate $\widehat{\boldsymbol{\theta}}$. If the model is not rejected by the J-statistic, no further steps are required.

Assuming that $\widehat{L}_e$ is not large relative to $\widehat{L}$, the only disadvantage of this approach is ultimately some loss in terms of efficiency due to the fact that for $L_e > L$ one estimates the extended $\boldsymbol{G}_{z,\lambda_e}$, as for example in Example 2.

Remark 7. Let $r_{\max} = \min(T, R) - 1$. The estimator for L_e based on the ER-statistic is defined as

$$\widehat{L}_e = \underset{r\in\{1,\dots,r_{\max}\}}{\arg\max}\ \mathrm{ER}(r); \quad \mathrm{ER}(r) = \frac{\lambda_r(T^{-1}\widehat{\boldsymbol{F}}_R\widehat{\boldsymbol{F}}_R')}{\lambda_{r+1}(T^{-1}\widehat{\boldsymbol{F}}_R\widehat{\boldsymbol{F}}_R')}, \quad (33)$$

where $\lambda_r(T^{-1}\widehat{\boldsymbol{F}}_R\widehat{\boldsymbol{F}}_R')$ is the rth largest eigenvalue of $T^{-1}\widehat{\boldsymbol{F}}_R\widehat{\boldsymbol{F}}_R'$. This procedure yields a consistent estimate of L_e because for $r < L_e$ $ER(r)$ remains bounded a.s., whereas for $r = L_e$ $ER(r) \to \infty$ as $\lambda_{r+1}(T^{-1}\widehat{\boldsymbol{F}}_R\widehat{\boldsymbol{F}}_R') \xrightarrow{p} 0$ by the continuity of the eigenvalues. However, as it currently stands, the ER procedure excludes the possibility that $\widehat{L}_e = R < T$, that is, the dimension of the column space of $\widehat{\boldsymbol{F}}_R$ is exactly R. We fix this shortcoming by borrowing the mock-eigenvalue idea of Ahn and Horenstein (2013) and suggesting that a single, redundant column in $\widehat{\boldsymbol{F}}_R$ always be included. For example, one can easily construct a redundant factor proxy as $N^{-1}\sum_{i=1}^N \mathbf{v}_i w_i^+$, where w_i^+ is randomly drawn from any zero mean distribution (e.g., the Rademacher $\{-1; 1\}$ distribution) and $\mathbf{v}_i$ is either defined in Equation (11), or it is one of the columns of $\boldsymbol{V}_i$ in Equation (15). In this way, we extend the definition of R to $R+1$ manually, thus avoiding the boundary problem of the original ER-statistic. The GR-statistic is discussed in greater detail in the supplementary appendix.

3.5.2. Best-Subset Selection

The best-subset selection method is a model selection approach that is motivated by the machine learning literature (see, e.g., Hastie, Tibshirani, and Friedman 2017, sec. 3.3). In the present context, the method aims to determine the combination of factor proxies that yields the smallest BIC value. In particular, let R be the number of factor proxies at hand, $\widehat{\boldsymbol{F}}_R$, and $L_{\max}$ ($\geq L_e$) be the maximum number of unobserved factors considered in estimation. In practice, $L_{\max}$ could be set as the largest possible value of L that is feasible to allow in estimation. Furthermore, let $\mathcal{B}(L_{\max})$ denote the set of different combinations of columns in $\widehat{\boldsymbol{F}}_R$ of sizes $P = 1, \dots, L_{\max}$. To illustrate, consider Example 3 with $L_{\max} = 2$. Therein, $R = 3$ since $\widehat{\boldsymbol{F}}_R = N^{-1}\sum_{i=1}^N \left(\boldsymbol{x}_i^{(1)}, \boldsymbol{x}_i^{(2)}, \boldsymbol{y}_i\right) w_i \equiv \left(\widehat{\boldsymbol{f}}_R^{(1)}, \widehat{\boldsymbol{f}}_R^{(2)}, \widehat{\boldsymbol{f}}_R^{(3)}\right)$. For $P = 1$, one can use either one of $\widehat{\boldsymbol{f}}_R^{(1)}$, $\widehat{\boldsymbol{f}}_R^{(2)}$ and $\widehat{\boldsymbol{f}}_R^{(3)}$. For $P = 2$, possible combinations of factor proxies are $\left\{\widehat{\boldsymbol{f}}_R^{(1)}, \widehat{\boldsymbol{f}}_R^{(2)}\right\}$, $\left\{\widehat{\boldsymbol{f}}_R^{(1)}, \widehat{\boldsymbol{f}}_R^{(3)}\right\}$, and $\left\{\widehat{\boldsymbol{f}}_R^{(2)}, \widehat{\boldsymbol{f}}_R^{(3)}\right\}$. In total, the cardinality of the set $\mathcal{B}(L_{\max})$ is at most $|\mathcal{B}(L_{\max})| = \binom{3}{1} + \binom{3}{2} = 6$.

Within our framework, a consistent estimate of the number of factors and the selected combination of factor proxies can be determined using a Schwartz/Bayesian model information criterion (BIC), as proposed originally by Cragg and Donald (1997) and Ahn, Lee, and Schmidt (2013). This is formalized right below.

Proposition 1. Let $p = 1, \dots, |\mathcal{B}(L_{\max})|$ and $Q_{N,p}\left(\widehat{\boldsymbol{\theta}}\left(\boldsymbol{\Omega}_N\right)|\widehat{\boldsymbol{F}}^{(p)}\right)$ be the value of the objective function evaluated at $\widehat{\boldsymbol{\theta}}$ given $\boldsymbol{\Omega}_N$

and some $\widehat{\boldsymbol{F}}^{(p)} \in \mathcal{B}(L_{\max})$:

$$Q_{N,p}\left(\widehat{\boldsymbol{\theta}}\left(\boldsymbol{\Omega}_N\right)|\widehat{\boldsymbol{F}}^{(p)}\right) = \overline{\boldsymbol{\mu}}_N\left(\widehat{\boldsymbol{\theta}}\right)'\boldsymbol{\Omega}_N\overline{\boldsymbol{\mu}}_N\left(\widehat{\boldsymbol{\theta}}\right).$$

Consider the following BIC:

$$S_{N,p} = N \times Q_{N,p}\left(\widehat{\boldsymbol{\theta}}\left(\boldsymbol{\Omega}_N\right)|\widehat{\boldsymbol{F}}^{(p)}\right) - \ln(N) \times h\left(p\right), \qquad (34)$$

where $h\left(p\right) = \rho \times \left(\zeta - \dim\left(\widehat{\boldsymbol{\theta}}|p(P)\right)\right) = \mathcal{O}(1)$, a strictly decreasing function of P with $0 < \rho < \infty$. Under the set of our assumptions, we have

$$\widehat{P} = \underset{\widehat{\boldsymbol{F}}^{(p)} \in \mathcal{B}(L_{\max})}{\arg\min}\ S_{N,p} \xrightarrow{p} L \text{ as } N \to \infty.$$

Proof. The proof follows directly from Cragg and Donald (1997). The only difference is that for every value of L several weights are potentially available, that is, one needs to consider minimum BIC for each value of $P = 1, \ldots, L_{\max}$ first. □

Proposition 1 implies that practitioners can estimate models for all choices of $\mathcal{B}(L_{\max})$ and pick $\widehat{P}$ (together with the associated combination of factor proxies) as the value of L that corresponds to the smallest BIC value. The above result holds as long as for $P = L$ there exists an element in $\mathcal{B}(L_{\max})$ that ensures that the GMM estimator is consistent. For instance, in the context of Example 3 where $L = 1$, one such element in $\mathcal{B}(L_{\max})$ is given by $N^{-1}\sum_i^N \boldsymbol{x}_i^{(2)} = \widehat{\boldsymbol{f}}_R^{(2)}$. Notice that in this case, the GMM estimator based on best-subset selection is more efficient than the GMM estimator based on regularization because asymptotically the former employs 1 factor proxy and the latter employs 2 factor proxies. On the other hand, if no such element exists (as, e.g., in Example 2), but there exists at least one $\widehat{\boldsymbol{F}}^{(p)}$ that satisfies Assumption 3.1 with $L_e > L$, then BIC will consistently estimate L_e instead. Finally, if no such $\widehat{\boldsymbol{F}}^{(p)}$ exists, then BIC will not consistently estimate L or L_e due to lack of identification, as discussed in Section 3.4. Note that the selected model will be rejected with high probability by the J-statistic in this case.

4. Finite Sample Evidence

4.1. Setup

We consider the following dynamic model with one or two factors:

$$y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t} + \sum_{r=1}^{2}\lambda_{r,i}^{y} f_{r,t} + \varepsilon_{i,t}^{y};$$
$$x_{i,t} = \delta y_{i,t-1} + \alpha_x x_{i,t-1} + \lambda_{1,i}^{x} f_{1,t} + \varepsilon_{i,t}^{x},$$

for $t > 0$, while for $t = 0$ we set:

$$y_{i,0} = \sum_{r=1}^{2}\lambda_{r,i} f_{r,0} + \varepsilon_{i,0}^{y}; \quad x_{i,0} = \lambda_{1,i}^{x} f_{1,0} + \varepsilon_{i,0}^{x}.$$

Additional covariates to be used in the construction of factor proxies are generated as

$$\mathrm{v}_{i,t}^{(1)} = \lambda_{1,i}^{\mathrm{v}1} f_{1,t} + \varepsilon_{i,t}^{\mathrm{v}1}; \quad \mathrm{v}_{i,t}^{(2)} = \sum_{r=1}^{2}\lambda_{r,i}^{\mathrm{v}2} f_{r,t} + \varepsilon_{i,t}^{\mathrm{v}}.$$

The factor loadings for the first factor are normally distributed with mean equal to μ_λ and unit variance, such that

$$\lambda_{1,i}^{\psi} = \mu_\lambda + \rho(\lambda_{1,i}^{y} - \mu_\lambda) + \sqrt{1-\rho^2}\upsilon_{1,i}^{\psi}; \quad \upsilon_{1,i}^{\psi} \sim \mathcal{N}(0,1),$$

where $\psi \in \{x, \mathrm{v}1, \mathrm{v}2\}$, and $\lambda_{1,i}^{y} \sim \mathcal{N}(\mu_\lambda, 1)$. ρ denotes the correlation coefficient between the factor loadings of the $y_{i,t}$ and $x_{i,t}$, and $y_{i,t}$ and $\mathrm{v}_{i,t}^{(1)}, \mathrm{v}_{i,t}^{(2)}$ processes.

The properties of the factor loadings that correspond to $f_{2,t}$ depend on the setup we consider. In particular, in the case where $L = L_e = 1$ we simply set $\lambda_{2,i}^{y} = \lambda_{2,i}^{\mathrm{v}2} = 0$ for all i. In the case where $L = L_e = 2$ the corresponding factor loadings are drawn independently of other factors loadings, that is,

$$\lambda_{2,i}^{y} \sim \mathcal{N}(\mu_\lambda, 1); \quad \lambda_{2,i}^{\mathrm{v}2} \sim \mathcal{N}(\mu_\lambda^{\mathrm{v}2}, 1). \qquad (35)$$

Such a setup facilitates the interpretability of the simulation results, without overly parameterizing what is already a large set of nuisance parameters. In what follows, we fix $\mu_\lambda^{\mathrm{v}2} = 1$. Finally, all factors are drawn as $f_{r,t} \sim \mathcal{N}(0,1)$.

The idiosyncratic errors are generated as

$$\varepsilon_{i,t}^{y} \sim \mathcal{N}(0,1); \quad \varepsilon_{i,t}^{x} \sim \mathcal{N}\left(0, \sigma_x^2\right); \quad \varepsilon_{i,t}^{\mathrm{v}1} \sim \mathcal{N}(0,1);$$
$$\varepsilon_{i,t}^{\mathrm{v}2} \sim \mathcal{N}(0,1); \quad t \geq 0.$$

In all designs the value of σ_x^2 is fixed to ensure that the signal-to-noise ratio of the model

$$\mathrm{SNR} \equiv \frac{1}{T}\sum_{t=1}^{T}\frac{\mathrm{var}\left(y_{i,t}|\lambda_{1,i}^{y}, \lambda_{2,i}^{y}, \lambda_{1,i}^{x}, \lambda_{2,i}^{x}, \{f_{1,s}\}_{s=0}^{t}, \{f_{2,s}\}_{s=0}^{t}\right)}{\mathrm{var}(\varepsilon_{i,t}^{y})} - 1,$$

equals 5. The chosen SNR value lies within the range of values considered in the literature, for example, Bun and Kiviet (2006) specifies $\mathrm{SNR} \in \{3; 9\}$.

We consider $N \in \{200; 800\}$, $T \in \{4; 8\}$, $\alpha \in \{0.4; 0.8\}$, and we set $\beta = 1 - \alpha$, such that the "long-run" parameter equals 1. The values of the remaining parameters are as follows: $\delta \in \{0; 0.3\}$, $\mu_\lambda = 1$, $\rho = 0.6$, and $\alpha_x = 0.6$. The number of replications performed equals 2000 for each design and the factors are drawn in each replication.

Remark 8. The supplementary appendix to the present article provides further finite sample results that correspond to alternative designs. In particular, among other setups, we examine the case where $\mathbf{v}_i$ contains more factors than those already drive y (i.e., $L_e > L$); moreover, we examine the effect of lack of identification by setting $\mu_\lambda = 0$, and we also specify a model with an additional covariate. In addition, we report results with respect to the nonlinear GMM estimators of Ahn, Lee, and Schmidt (2013) and Robertson and Sarafidis (2015) for the one-factor design.

4.2. Results

We investigate the finite sample properties of the following four estimators: "F1" denotes the GMM estimator that uses $\mathbf{v}_i^{(1)}$ and $w_i = 1$; "F2" is the GMM estimator that uses $(\mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)})$ and $w_i = 1$; "Fr" denotes the regularized GMM estimator that uses $(\mathbf{v}_i^{(1)}, \mathbf{v}_i^{(2)})$ and $\boldsymbol{w}_i = (1, y_{i,0})'$, as described in the regularization approach outlined in Section 3.5. To isolate the effect of regularization on the construction of factor proxies, we take L_e as given. Results corresponding to Fr based on an estimate of L_e are reported in the supplementary appendix of the article; finally, "Fbic" denotes the estimator that employs different combinations of factor proxies based on the choices above, and picks up the proxies corresponding to the minimum BIC value. Thus, Fbic represents the best-subset selection approach outlined in Section 3.5. The value of ρ in the BIC function is set equal to 0.75, following BIC1 in Ahn, Lee, and Schmidt (2013), and Zhu, Sarafidis, and Silvapulle (2020).

All estimators use the maximum number of moment conditions with respect to lagged values of $y_{i,t}$ and $x_{i,t}$, although for comparison purposes $x_{i,t}$ is always treated as weakly exogenous regardless of the value of δ. Notice that, for reasons explained in the supplementary appendix, setting $w_i = 1$ leads to slightly better results compared to $w_i = y_{i,0}$ in the present design. Therefore for $L = L_e = 1$, F1 can be viewed as an "oracle" estimator because it makes use of the true number of factors, as well as the "right" choice of $\mathbf{v}_i$ and w_i. Likewise, F2 can be viewed as an oracle estimator for $L = L_e = 2$. Therefore, the performances of F1 and F2 may serve as good benchmarks for the performance of Fr and Fbic.

Tables A1 and A2 report results in terms of bias, RMSE, standard deviation and empirical size of the t-test for $L = L_e = 1$ and $L = L_e = 2$, respectively. Nominal size is set equal to 5%. We make use of corrected standard errors, which are computed based on Windmeijer (2005). This is important because, as is well known in the dynamic panel data literature, the two-step GMM estimator may exhibit substantial size distortions, especially when the number of moment conditions is large relative to N.

As can be seen from Table A1, all four estimators show negligible bias under all combinations of N, T, α, and δ. In addition, the RMSE values of all estimators are small and fall roughly at the rate of $\sqrt{N}$, as predicted by our theory. Fr performs very similar to F1 in terms of RMSE, which indicates that the proposed regularization approach works very well. On the other hand, RMSE is slightly higher for Fbic (especially when $N = 200$), which mostly reflects the fact that L is treated as unknown. Finally, the empirical size of all estimators is close to the nominal value for both α and β in most cases, unless N is relatively small and T is relatively large, in which case there appear to be upward size distortions. This outcome implies that using too many moment conditions when the cross-sectional dimension is small can result in size distortions, despite the standard error correction. However, it is worth emphasizing that in practice this problem can be mitigated substantially by using only a subset of the moment conditions available. Since this has already been demonstrated by Juodis and Sarafidis (2018) using simulated data for the factor-augmented model, we do not explore this possibility here.

Similar conclusions can be drawn from Table A2 where $L = L_e = 2$, focusing upon the performance of Fr and Fbic vis-a-vis F2. It is worth noting that F1 is not consistent in this case because it estimates one factor only. Thus, it is not surprising that in most cases F1 exhibits large RMSE and substantial size distortions.

Table A3 reports results on the overidentifying restrictions (J) test statistic (nominal size equals 5%) for F1, F2, and Fr. In addition, this table reports selection frequencies of $\widehat{L}$ based on BIC, and $\widehat{L}_e$ using the ER-statistic, as outlined in Section 3.5. Notice that since $L = L_e$, $P = L$ and therefore we do not distinguish between $\widehat{P}$ and $\widehat{L}$. As we can see, for $L = L_e = 1$, the size of the J-statistic corresponding to both F1 and Fr is close to the 5% level in most circumstances, unless N is relatively small and T is relatively large, in which case there is a small downward size distortion, especially for F2. As has been mentioned, such small-sample distortions can be mitigated in practice by using a smaller number of moment conditions. Similar conclusions hold in regards to the performance of the J-statistic corresponding to F2 and Fr for $L = L_e = 2$. Note that since F1 is not consistent in the two-factor case, the results of the J-statistic corresponding to F1 reflect power, which appears to be high under all circumstances.

In regards to model selection, for $L = L_e = 1$ all methods appear to perform well and select the true number of factors with high frequency. As expected, model selection becomes less straightforward for $L = L_e = 2$, especially when both N and T are small. Partially, this may attributed to the fact that for small values of T the two factors can be highly collinear in some simulated samples, even if they are drawn independently. However, it is clear that the frequency of selecting $\widehat{L} = 2$ and $\widehat{L}_e = 2$ rises substantially with larger values of either N or T. Finally, the corresponding rows of ER-statistic in Table A3 need not sum to 1, due to rounding.

5. Application: Estimation of the Price Elasticity of Urban Water Demand

A large number of studies have focused on the estimation of the price elasticity of water usage demand (see, e.g., House-Peters and Chang 2011; Araral and Wang 2013 for excellent surveys). The vast majority of the literature assumes that the effect of weather is linear. However, as Maidment and Miaou (1986) and Gato, Jayasuriya, and Roberts (2007) pointed out, water usage is most likely to respond to changes in weather conditions in a nonlinear fashion. We address this concern in what follows by allowing for nonlinear effects of weather conditions, depending on household/property-specific unobserved characteristics, such as household size, garden size, and others. This setup is represented by a common factor structure.

5.1. *Data and Methodology*

We make use of publicly available multi-household level data from New South Wales, Australia, provided by the Sydney Water Corporation (SWC), see also Abrams et al. (2012). SWC is the largest water utility in Australia, serving more than 4 million people, while its area of operations covers around 12,700 km^2.

Our sample contains 4500 multi-household units, each one being observed over a period of 5 years, 2004–2008. Each unit represents an average of four to six households, to preserve privacy. Additional descriptive statistics of the data are reported in the supplementary appendix to this article. The model that we consider for studying the price elasticity of water demand is as follows:

$$y_{i,t} = \alpha y_{i,t-1} + \beta_1 \text{price}_{i,t} + \beta_2 \text{rain}_{i,t} + \beta_3 \text{temp}_{i,t} + \epsilon_{i,t};$$
$$\epsilon_{i,t} = \boldsymbol{\lambda}_i' \boldsymbol{f}_t + \varepsilon_{i,t}, \quad (36)$$

for $i = 1, \ldots, N(= 4500)$ and $t = 1, \ldots, T(= 4)$, where $y_{i,t}$ denotes the natural logarithm of the average daily water consumption for household i at year t, expressed in thousands of liters of water (kL); $\text{price}_{i,t}$ is the average real price paid per kiloliter of water used by household i at time t; and $\text{rain}_{i,t}$ and $\text{temp}_{i,t}$ denote the average amounts of daily rainfall (mm) and temperature (°C) during year t. The dynamic specification accommodates partial adjustment mechanisms in water demand. This is due to both habit formation in water consumption and energy efficiency constraints associated with the existing stock of durable goods within households. The log-linear functional form implies that price elasticity depends on the level of price itself. That is, the higher the level of the price, the more sensitive consumers become to changes in price. This is consistent with utility theory (see, e.g., Al-Qunaibet and Johnston 1985). The weather variables are unit-specific because they are determined by the physical proximity of each property to a total of thirteen weather stations that exist across Sydney, operated by the Australian Bureau of Meteorology. This reflects the fact that weather patterns can vary substantially across NSW and, more specifically, the coast generally has more rainfall and cooler conditions than many areas located inland.

The common factor structure allows for nonlinear unobserved heterogeneity across individual units, due to differences in the number of people living in a household, pool ownership, garden size and structure, and so on. As an example, consider two properties: one with two household members, no swimming pool and a small garden, and another one with five members, a large garden and a pool. The difference in average yearly consumption between the two properties is expected to be proportionately larger under extreme weather conditions than under normal conditions. To put this differently, the change in water consumption following an extreme weather event is likely to be smaller for the former property than the latter.

The formulation above requires that household size (and other property-specific characteristics) remains constant over the 5-year period of our analysis. While it is unreasonable to expect such condition to be fulfilled for all households in the sample, recall that each individual unit $i = 1, \ldots, N$ represents an average of four to six households, which is due to the data aggregation implemented by SWC in the original dataset. As a result, changes in household size over time are likely to be smoothed out to a large extent. If this is not true, our estimator will not be consistent. This implies that violation of such restriction can indeed be detected in practice using the usual J-statistic.

We implement our methodology by assuming that the unobserved factor component is approximated by an external variable, the average daily soil moisture index (smi) observed for unit i at period t. smi is computed based on a combination of precipitation, temperature and soil moisture, and is used widely as an index accounting for extreme weather and soil drought intensity (see, e.g., Hunt et al. 2009). Hence, in terms of the notation used in Section 3.1, we set $\mathrm{v}_{i,t} = \text{smi}_{i,t}$.

We estimate the model by fitting $L = \{0, 1, 2\}$ factors. We employ three weights, namely $\boldsymbol{w}_i = \left(1, y_{i,0}, y_{i,0}^2\right)'$. In terms of the notation introduced in Section 3, we have $L_{\max} = 2$, $R = 3$ and the set $\mathcal{B}(L_{\max})$ contains all 6 possible permutations of $\widehat{\boldsymbol{F}}_R = \frac{1}{N}\sum_{i=1}^{N} \mathbf{v}_i \boldsymbol{w}_i'$, where $\mathbf{v}_i$ is constructed by stacking $\text{smi}_{i,t}$ for $t = 1, \ldots, T$ in a $[T \times 1]$ vector. The implementation of our approach based on regularization follows closely the steps described in Section 3.5, with L_e estimated from $T^{-1}\widehat{\boldsymbol{F}}_R\widehat{\boldsymbol{F}}_R'$ using the ER-statistic outlined therein. The implementation based on best-subset selection estimates the model using six different combinations of weights, that is, those in $\mathcal{B}(L_{\max})$. The estimated number of factors, $\widehat{P}$, as well as the associated selected weight combination, are determined using the BIC criterion documented in Section 3.5. The value of ρ in the BIC function is set equal to 0.75, as in the Monte Carlo setup.

5.2. Results

Table 1 reports results for ten different specifications. In specific, $M0$ denotes the estimator that imposes $\epsilon_{i,t} = \varepsilon_{i,t}$, that is, it assumes unobserved heterogeneity away. M_{DIF} and M_{SYS} denote the first-differenced and system GMM estimators proposed by Arellano and Bond (1991) and Arellano and Bover (1995), respectively. Both estimators allow for an additive two-way error components structure and are obtained using the xtabond2 command in Stata 15 (see Roodman 2009). $M_{\widetilde{F}}$ denotes the GMM estimator that uses regularized factor proxies, $\widetilde{\boldsymbol{F}}$, based on $\widehat{L}_e = 1$ (the value of which is obtained from the ER-statistic). In addition, $M1_c$, $M1_{y_0}$, and $M1_{y_0^2}$ denote the GMM estimators that impose $L_e = 1$ and use $w_i = 1$, $w_i = y_{i,0}$ and $w_i = y_{i,0}^2$, respectively. Finally, $M2_{c,y_0}$ imposes $L_e = 2$ with weights given by $\boldsymbol{w}_i = \left(1, y_{i,0}\right)'$. The same holds for $M2_{c,y_0^2}$ and $M2_{y_0,y_0^2}$, except that $\boldsymbol{w}_i = \left(1, y_{i,0}^2\right)'$ and $\boldsymbol{w}_i = \left(y_{i0}, y_{i0}^2\right)'$, respectively.

In all models the price variable is treated as endogenous and is instrumented by appropriate lagged values of the same variable. This is because during the period of the analysis a two-tier pricing scheme was in place in NSW, such that consumers paid a higher price when their consumption levels exceeded a certain threshold. On the other hand, all weather variables are treated as exogenous with respect to $\varepsilon_{i,t}$.

All estimators that allow for a common factor structure, make use of $\zeta = 40$ moment conditions, whereas M_{DIF} and M_{SYS} make use of 17 and 21 moment conditions, respectively. The difference in the number of moment conditions between the GMM estimators that impose an additive structure and those that allow for a genuine factor model is mainly due to their treatment of the exogenous weather variables. In the former case, standard practice involves taking first differences to remove unobserved (linear) heterogeneity and then using the exogenous weather variables as Anderson–Hsiao type instruments. In the latter case, present and lagged values of the weather variables are included as instruments in each time period, to allow for possible arbitrary correlations between nonlinear heterogeneity and weather conditions.

Results are reported in terms of the estimated coefficients and standard errors (in parentheses), where $\widehat{\beta}_1/(1-\widehat{\alpha})$ corresponds to the long-run price coefficient, the standard error of which has been obtained using the Delta method. Furthermore, Table 1 reports the J-statistic and its p-value (in square brackets), the

Table 1. Results.

	$M0$	M_{DIF}	M_{SYS}	$M_{\tilde{F}}$	$M1_c$	$M1_{y_0}$	$M1_{y_0^2}$	$M2_{c,y_0}$	$M2_{c,y_0^2}$	$M2_{y_0,y_0^2}$
$\widehat{\alpha}$	0.942	0.504	0.771	0.405	0.414	0.405	0.393	0.354	0.382	0.395
	(0.004)	(0.045)	(0.018)	(0.047)	(0.048)	(0.048)	(0.047)	(0.082)	(0.082)	(0.087)
$\widehat{\beta}_1$	0.170	−0.032	0.261	−0.185	−0.178	−0.187	−0.192	−0.247	−0.225	−0.190
	(0.015)	(0.361)	(0.419)	(0.034)	(0.035)	(0.035)	(0.034)	(0.061)	(0.054)	(0.051)
$\widehat{\beta}_2$	−0.051	−0.001	−0.003	−0.013	−0.013	−0.021	−0.013	−0.030	−0.011	−0.018
	(0.003)	(0.008)	(0.009)	(0.006)	(0.006)	(0.010)	(0.006)	(0.009)	(0.010)	(0.011)
$\widehat{\beta}_3$	−0.008	0.034	0.029	0.050	0.050	0.084	0.050	0.032	0.060	0.051
	(0.001)	(0.012)	(0.013)	(0.010)	(0.010)	(0.019)	(0.010)	(0.013)	(0.015)	(0.016)
$\frac{\widehat{\beta}_1}{1-\widehat{\alpha}}$	2.93	−0.065	1.14	−0.310	−0.303	−0.314	−0.316	−0.383	−0.364	−0.314
	(0.075)	(0.726)	(1.86)	(0.078)	(0.079)	(0.079)	(0.077)	(0.137)	(0.128)	(0.113)
J-test	156.3	27.6	49.2	28.8	28.7	28.9	28.9	13.6	16.1	15.6
p-value	[0.000]	[0.002]	[0.000]	[0.092]	[0.094]	[0.092]	[0.091]	[0.096]	[0.042]	[0.050]
ζ	40	17	21	40	40	40	40	40	40	40
dim ($\boldsymbol{\theta}$)	5	9	10	20	20	20	20	32	32	32
BIC	10.6	−17.8	−4.91	−54.48	−54.54	−54.46	−54.39	−19.69	−17.24	−17.74

number of moment conditions and of parameters estimated for each model, and finally the value of the BIC.

It is clear that the p-value of the J-statistic is close to zero when we fit either zero factors or the two-way (additive) error components structure, which implies that the model is misspecified. This is also reflected in the estimated price coefficient, which is largely statistically insignificant for both M_{DIF} and M_{SYS}, and for the latter it even has the wrong sign. We note that these results are not sensitive to the number of instruments employed. For example, using only the two most recent lags of the dependent variable as instruments, that is, $y_{i,t-2}$ and $y_{i,t-3}$, along with the *collapse* option in Stata, the J-statistic for M_{SYS} roughly equals 27.6, and so the p-value remains close to zero.

On the other hand, fitting one or two genuine factors fails to reject the specified model at the 1% level of significance. This finding provides evidence that the factor structure is supported by the data compared to the additive error components model, and demonstrates the importance of controlling for nonlinear heterogeneity.

More specifically, recall that the ER-statistic indicates that $\widehat{L}_e = 1$. In addition, according to the BIC criterion, $\widehat{P} = 1$. Furthermore, $M1_c$, the estimator that uses a constant weight, $w_i = 1$, minimizes the overall BIC criterion. The results from adopting the regularization approach ($M_{\tilde{F}}$) and the best-subset selection approach ($M1_c$) are similar. While this may be partially attributed to $\widehat{L}_e = \widehat{P} = 1$, it is evident that the estimated price elasticity of demand appears to be robust across different factor proxies and different values of L. This is a desirable outcome.

In what follows we discuss further findings based on $M_{\tilde{F}}$. A unit (dollar) increase in the price of water is estimated to reduce water consumption by approximately 18.5% and 31.0% in the short- and long-run, respectively. Similarly, a unit increase in rain (temperature) is expected to reduce (increase) water consumption by approximately 1.13% (5.00%) in the short-run.

The price elasticity of demand is computed by multiplying the relevant price coefficients with a range of values for price. Table 2 presents elasticity estimates at four different values of price; namely, mean and median price, as well as the 10th and 90th percentiles. For example, at the median price of \$1.37 per kL, a 1% increase in the price of water lowers demand by about 0.25% in the short-run and 0.43% in the long-run.

Table 2. Point-wise predicted elasticities for M_{F_R}.

	10th perc.	Mean	Median	90th perc.
Price	1.17	1.35	1.37	1.56
SR elasticity	−0.216	−0.249	−0.253	−0.288
LR elasticity	−0.362	−0.419	−0.426	−0.483

As expected, urban water demand appears to be much more elastic in the long-run than in the short-run. This may be attributed to habit formation and technological constraints of water appliance efficiency. Moreover, the value of $\widehat{\alpha} = 0.405$ implies that it takes about 2.5 years for 90% of the total price effect to be realized, all other things being constant. This outcome casts doubt on the potential effectiveness of scarcity pricing to balance demand and supply of water in periods of transitory droughts.

In comparison to other studies in the literature, the estimated price elasticity of demand is in the low-to-middle range of results. For instance, the long-run price coefficient is not statistically different to the value obtained by Nauges and Thomas (2003) (see Table III in their article) although theirs is derived from the constant-elasticity model using municipal-level data and includes average income but not weather conditions.

7. Concluding Remarks

This article puts forward a novel method-of-moments approach for estimation of factor-augmented panel data models with endogenous regressors and T fixed. The underlying idea is to proxy the factors by observed variables, so that the resulting moment conditions are linear in the parameters. The proposed methodology addresses several issues that arise with existing nonlinear GMM estimators, such as local minima-related problems and a potential lack of global identification. At the same time, the proposed methodology retains the appealing features of the method of moments in that it accommodates weakly exogenous and endogenous regressors without the need for bias correction.

We note that although this article has explicitly assumed that $\mathbf{v}_i$ (or $\boldsymbol{V}_i$) has an additive factor structure with corresponding factor loadings $\boldsymbol{\gamma}_i$ ($\boldsymbol{\Upsilon}_i$), in practice, as it is clear from Equation (21), the additive factor specification is sufficient but not necessary. In particular, since the proposed estimation method uses information from $\mathrm{E}_{\mathcal{F}}[\widehat{\boldsymbol{F}}_e]$ only, it is sufficient that this expected value is of reduced rank structure, that is, $\mathrm{E}_{\mathcal{F}}[\widehat{\boldsymbol{F}}_e] = \boldsymbol{F}_e\boldsymbol{A}$. The deviations from the mean, that is, $\overline{\boldsymbol{\Psi}} = N^{-1}\sum_{i=1}^{N}\boldsymbol{\Psi}_i$, can still contain unobserved common shocks so long as they satisfy the conditional independence in Assumption 2.1. Finally, $\boldsymbol{F}_e$ and $\boldsymbol{A}$ should be simply regarded as correspondingly the left and right singular vectors (up to a scaling) of $\mathrm{E}_{\mathcal{F}}[\widehat{\boldsymbol{F}}_e]$. Thus, the $\boldsymbol{F}_e\boldsymbol{A}$ decomposition can be assumed completely without loss of generality, provided that L_e is defined accordingly.

We hope that the proposed methodology will enhance the application of multifactor error structures in panels involving micro level data, and encourage empirical researchers to implement such approaches in practice. Furthermore, since the resulting method of moments estimator has a close form solution, our approach can be extended straightforwardly to several different set ups, such as multidimensional panels, spatial panels, pseudo panels, and threshold models, to mention a few. We leave these avenues for future research.

Supplementary Materials

The supplementary appendix to this article provides additional results about the method developed in the present article. In particular, Section S1 analyses several extensions of the model analyzed in the main text, including unbalanced panels, observed factors, and consistency of the GMM estimator under an alternative set of assumptions, in which the factor loadings are treated as a sequence of constants. Section S2 provides descriptive statistics for the data used in the empirical illustration. Section S3 reports additional Monte Carlo results. Finally, Section S4 provides proofs of the main theoretical results put forward in the article.

Acknowledgments

We would like to thank the editor, Jianqing Fan, and three anonymous referees for their constructive comments and suggestions, which have helped to improve our article substantially. We are also grateful to Pavel Čížek, Mervyn Silvapulle, Rutger Teulings, and Frank Windmeijer for helpful comments at an earlier stage of this project.

Funding

Financial support from the NWO VENI grant number 451-17-002 is gratefully acknowledged by the first author. The second author acknowledges financial support from the Australian Research Council, under research grant number DP-170103135.

Appendix A: Monte Carlo Results

Table A1. Estimation results for $L = 1$ and $L_e = 1$ setup.

Designs				F1				F2				Fr				Fbic			
N	*T*	α	δ	*Bias*	RMSE	Std	Size	*Bias*	RMSE	Std	Size	*Bias*	RMSE	Std	Size	*Bias*	RMSE	Std	Size
									α										
200	4	0.4	0.0	0.00	0.02	0.02	0.06	0.00	0.05	0.05	0.02	0.00	0.02	0.02	0.06	0.00	0.04	0.04	0.07
200	4	0.4	0.3	0.00	0.03	0.03	0.05	−0.01	0.07	0.07	0.02	0.00	0.03	0.03	0.07	0.00	0.04	0.04	0.08
200	4	0.8	0.0	0.00	0.03	0.03	0.06	0.00	0.05	0.05	0.02	0.00	0.02	0.02	0.07	0.00	0.03	0.03	0.07
200	4	0.8	0.3	0.00	0.03	0.03	0.07	0.00	0.06	0.06	0.03	0.00	0.03	0.03	0.07	0.00	0.03	0.03	0.08
200	8	0.4	0.0	0.00	0.02	0.02	0.11	0.00	0.02	0.02	0.06	0.00	0.01	0.01	0.09	0.00	0.02	0.02	0.11
200	8	0.4	0.3	−0.01	0.03	0.03	0.18	−0.01	0.03	0.03	0.09	−0.01	0.03	0.03	0.15	−0.01	0.04	0.03	0.18
200	8	0.8	0.0	0.00	0.01	0.01	0.10	0.00	0.02	0.02	0.07	0.00	0.01	0.01	0.12	0.00	0.02	0.02	0.13
200	8	0.8	0.3	0.00	0.02	0.02	0.13	0.00	0.02	0.02	0.07	0.00	0.02	0.02	0.13	0.00	0.02	0.02	0.13
800	4	0.4	0.0	0.00	0.01	0.01	0.06	0.00	0.03	0.03	0.02	0.00	0.01	0.01	0.04	0.00	0.01	0.01	0.07
800	4	0.4	0.3	0.00	0.02	0.02	0.05	0.00	0.04	0.04	0.02	0.00	0.02	0.02	0.06	0.00	0.02	0.02	0.06
800	4	0.8	0.0	0.00	0.01	0.01	0.06	0.00	0.03	0.03	0.02	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.07
800	4	0.8	0.3	0.00	0.01	0.01	0.05	0.00	0.04	0.04	0.02	0.00	0.01	0.01	0.05	0.00	0.02	0.02	0.06
800	8	0.4	0.0	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.05	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.06
800	8	0.4	0.3	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.05	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.06
800	8	0.8	0.0	0.00	0.01	0.01	0.07	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.07	0.00	0.01	0.01	0.07
800	8	0.8	0.3	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.05	0.00	0.01	0.01	0.07	0.00	0.01	0.01	0.07
									β										
200	4	0.4	0.0	0.00	0.03	0.03	0.07	0.00	0.07	0.07	0.02	0.00	0.02	0.02	0.07	0.00	0.06	0.06	0.06
200	4	0.4	0.3	0.00	0.03	0.03	0.06	0.01	0.08	0.08	0.02	0.00	0.03	0.03	0.07	0.00	0.04	0.04	0.06
200	4	0.8	0.0	0.00	0.01	0.01	0.06	0.00	0.03	0.03	0.02	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.06
200	4	0.8	0.3	0.00	0.01	0.01	0.05	0.00	0.04	0.04	0.03	0.00	0.01	0.01	0.06	0.00	0.02	0.02	0.06
200	8	0.4	0.0	0.00	0.02	0.02	0.11	0.00	0.02	0.02	0.06	0.00	0.02	0.02	0.11	0.01	0.02	0.02	0.13
200	8	0.4	0.3	0.01	0.04	0.03	0.19	0.01	0.03	0.03	0.10	0.01	0.03	0.03	0.16	0.01	0.04	0.04	0.19
200	8	0.8	0.0	0.00	0.01	0.01	0.10	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.10	0.00	0.01	0.01	0.10
200	8	0.8	0.3	0.00	0.02	0.02	0.13	0.00	0.02	0.02	0.07	0.00	0.01	0.01	0.12	0.00	0.02	0.02	0.14
800	4	0.4	0.0	0.00	0.01	0.01	0.06	0.00	0.05	0.05	0.02	0.00	0.01	0.01	0.06	0.00	0.02	0.02	0.06
800	4	0.4	0.3	0.00	0.02	0.02	0.05	0.00	0.05	0.05	0.03	0.00	0.02	0.02	0.05	0.00	0.02	0.02	0.05
800	4	0.8	0.0	0.00	0.01	0.01	0.06	0.00	0.02	0.02	0.02	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.06
800	4	0.8	0.3	0.00	0.01	0.01	0.05	0.00	0.03	0.03	0.02	0.00	0.01	0.01	0.05	0.00	0.01	0.01	0.06
800	8	0.4	0.0	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.05	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.07
800	8	0.4	0.3	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.05	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.06
800	8	0.8	0.0	0.00	0.00	0.00	0.06	0.00	0.01	0.01	0.04	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.06
800	8	0.8	0.3	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.05	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.06

Table A2. Estimation results for $L = 2$ and $L_e = 2$ setup.

Designs				F1				F2				Fr				Fbic			
N	T	α	δ	Bias	RMSE	Std	Size	Bias	RMSE	Std	Size	Bias	RMSE	Std	Size	Bias	RMSE	Std	Size
									α										
200	4	0.4	0.0	−0.02	0.14	0.14	0.60	0.00	0.04	0.04	0.05	0.00	0.04	0.04	0.05	−0.01	0.09	0.09	0.09
200	4	0.4	0.3	−0.05	0.26	0.26	0.66	0.00	0.06	0.06	0.04	0.00	0.06	0.06	0.05	−0.01	0.22	0.22	0.09
200	4	0.8	0.0	−0.02	0.14	0.13	0.60	0.00	0.04	0.04	0.05	0.00	0.04	0.04	0.05	−0.01	0.09	0.09	0.10
200	4	0.8	0.3	−0.03	0.17	0.17	0.64	0.00	0.05	0.05	0.05	0.00	0.05	0.05	0.06	−0.01	0.15	0.15	0.11
200	8	0.4	0.0	−0.03	0.11	0.11	0.76	0.00	0.02	0.02	0.08	0.00	0.02	0.02	0.08	0.00	0.02	0.02	0.09
200	8	0.4	0.3	−0.10	0.33	0.31	0.83	0.00	0.03	0.03	0.11	0.00	0.03	0.03	0.11	0.00	0.04	0.04	0.11
200	8	0.8	0.0	−0.02	0.09	0.09	0.73	0.00	0.01	0.01	0.09	0.00	0.01	0.01	0.09	0.00	0.02	0.02	0.09
200	8	0.8	0.3	−0.08	0.20	0.18	0.79	0.00	0.02	0.02	0.10	0.00	0.02	0.02	0.10	0.00	0.02	0.02	0.11
800	4	0.4	0.0	−0.02	0.14	0.14	0.79	0.00	0.02	0.02	0.05	0.00	0.02	0.02	0.04	0.00	0.06	0.06	0.07
800	4	0.4	0.3	−0.05	0.25	0.24	0.80	0.00	0.03	0.03	0.06	0.00	0.03	0.03	0.05	−0.01	0.13	0.13	0.08
800	4	0.8	0.0	−0.02	0.13	0.13	0.78	0.00	0.02	0.02	0.05	0.00	0.02	0.02	0.05	0.00	0.04	0.04	0.08
800	4	0.8	0.3	−0.04	0.17	0.17	0.80	0.00	0.02	0.02	0.05	0.00	0.02	0.02	0.06	−0.01	0.12	0.12	0.08
800	8	0.4	0.0	−0.02	0.11	0.10	0.85	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.06
800	8	0.4	0.3	−0.09	0.31	0.30	0.91	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.07
800	8	0.8	0.0	−0.02	0.08	0.08	0.84	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.06
800	8	0.8	0.3	−0.07	0.17	0.16	0.87	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.07	0.00	0.01	0.01	0.08
									β										
200	4	0.4	0.0	0.02	0.14	0.14	0.42	0.00	0.06	0.06	0.06	0.00	0.06	0.06	0.05	0.01	0.21	0.21	0.08
200	4	0.4	0.3	0.04	0.26	0.26	0.54	0.00	0.07	0.07	0.05	0.00	0.08	0.08	0.06	0.00	0.28	0.28	0.08
200	4	0.8	0.0	0.01	0.05	0.05	0.32	0.00	0.03	0.03	0.05	0.00	0.03	0.03	0.06	0.00	0.07	0.07	0.06
200	4	0.8	0.3	0.01	0.07	0.07	0.35	0.00	0.04	0.04	0.05	0.00	0.04	0.04	0.04	0.01	0.13	0.13	0.07
200	8	0.4	0.0	0.03	0.13	0.13	0.71	0.00	0.02	0.02	0.09	0.00	0.02	0.02	0.09	0.00	0.02	0.02	0.09
200	8	0.4	0.3	0.11	0.39	0.37	0.81	0.00	0.03	0.03	0.11	0.00	0.03	0.03	0.11	0.00	0.04	0.04	0.12
200	8	0.8	0.0	0.01	0.05	0.05	0.61	0.00	0.01	0.01	0.09	0.00	0.01	0.01	0.09	0.00	0.01	0.01	0.10
200	8	0.8	0.3	0.07	0.19	0.18	0.74	0.00	0.02	0.02	0.10	0.00	0.02	0.02	0.10	0.00	0.02	0.02	0.11
800	4	0.4	0.0	0.02	0.14	0.14	0.64	0.00	0.03	0.03	0.05	0.00	0.03	0.03	0.05	0.00	0.18	0.18	0.07
800	4	0.4	0.3	0.04	0.24	0.24	0.72	0.00	0.03	0.03	0.05	0.00	0.04	0.04	0.05	0.01	0.12	0.12	0.07
800	4	0.8	0.0	0.00	0.04	0.04	0.53	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.05	0.00	0.13	0.13	0.07
800	4	0.8	0.3	0.01	0.08	0.08	0.56	0.00	0.02	0.02	0.05	0.00	0.02	0.02	0.06	0.00	0.12	0.12	0.07
800	8	0.4	0.0	0.03	0.12	0.11	0.81	0.00	0.01	0.01	0.05	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.06
800	8	0.4	0.3	0.09	0.37	0.36	0.90	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.07
800	8	0.8	0.0	0.01	0.04	0.04	0.74	0.00	0.01	0.01	0.05	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.06
800	8	0.8	0.3	0.06	0.16	0.15	0.83	0.00	0.01	0.01	0.07	0.00	0.01	0.01	0.06	0.00	0.01	0.01	0.07

Table A3. Model selection and specification testing results.

				$L = L_e = 1$								$L = L_e = 2$							
Designs				J-statistic			BIC		ER L_e			J-statistic			BIC		ER L_e		
N	T	α	δ	F1	F2	Fr	$\#\hat{L}=1$	$\#\hat{L}=2$	$\#\hat{L}_e=1$	$\#\hat{L}_e=2$	$\#\hat{L}_e=3$	F1	F2	Fr	$\#\hat{L}=1$	$\#\hat{L}=2$	$\#\hat{L}_e=1$	$\#\hat{L}_e=2$	$\#\hat{L}_e=3$
200	4	0.4	0.0	0.03	0.01	0.05	0.98	0.02	0.98	0.00	0.02	0.97	0.05	0.05	0.16	0.84	0.16	0.76	0.08
200	4	0.4	0.3	0.03	0.01	0.04	0.98	0.02	0.97	0.00	0.02	0.96	0.04	0.05	0.16	0.84	0.17	0.76	0.07
200	4	0.8	0.0	0.05	0.02	0.05	0.98	0.02	0.98	0.00	0.02	0.96	0.04	0.04	0.20	0.80	0.17	0.76	0.08
200	4	0.8	0.3	0.04	0.02	0.04	0.98	0.02	0.97	0.00	0.03	0.95	0.04	0.04	0.20	0.80	0.15	0.77	0.08
200	8	0.4	0.0	0.03	0.01	0.03	0.96	0.04	1	0.00	0.00	1	0.03	0.02	0.01	0.99	0.06	0.93	0.01
200	8	0.4	0.3	0.03	0.01	0.03	0.96	0.04	1	0.00	0.00	1	0.02	0.02	0.02	0.98	0.07	0.93	0.01
200	8	0.8	0.0	0.02	0.01	0.03	0.97	0.03	1	0.00	0.00	1	0.03	0.03	0.01	0.99	0.07	0.92	0.01
200	8	0.8	0.3	0.03	0.01	0.03	0.97	0.03	1	0.00	0.00	1	0.02	0.02	0.02	0.98	0.06	0.93	0.01
800	4	0.4	0.0	0.06	0.02	0.06	0.99	0.01	0.99	0.00	0.01	1	0.05	0.05	0.05	0.95	0.07	0.91	0.03
800	4	0.4	0.3	0.04	0.02	0.04	1	0.00	0.99	0.00	0.01	1	0.05	0.05	0.04	0.96	0.07	0.90	0.03
800	4	0.8	0.0	0.04	0.02	0.04	1	0.00	0.99	0.00	0.01	0.99	0.04	0.04	0.06	0.94	0.08	0.89	0.03
800	4	0.8	0.3	0.05	0.02	0.06	1	0.00	0.99	0.00	0.01	1	0.05	0.06	0.05	0.95	0.07	0.90	0.03
800	8	0.4	0.0	0.05	0.02	0.04	1	0.01	1	0.00	0.00	1	0.05	0.05	0.00	1	0.01	0.99	0.00
800	8	0.4	0.3	0.04	0.02	0.04	0.99	0.01	1	0.00	0.00	1	0.05	0.04	0.00	1	0.01	0.99	0.00
800	8	0.8	0.0	0.05	0.02	0.05	1	0.00	1	0.00	0.00	1	0.05	0.05	0.00	1	0.01	0.99	0.00
800	8	0.8	0.3	0.05	0.02	0.05	0.99	0.01	1	0.00	0.00	1	0.05	0.04	0.00	1	0.01	0.99	0.00

References

Abrams, B., Kumaradevan, S., Sarafidis, V., and Spaninks, F. (2012), "An Econometric Assessment of Pricing Sydney's Residential Water Use," *Economic Record*, 88, 89–105. [10]

Ahn, S. C., and Horenstein, A. R. (2013), "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81, 1203–1227. [8]

Ahn, S. C., Lee, Y. H., and Schmidt, P. (2013), "Panel Data Models With Multiple Time-Varying Individual Effects," *Journal of Econometrics*, 174, 1–14. [1,2,3,8,9,10]

Al-Qunaibet, M. H., and Johnston, R. S. (1985), "Municipal Demand for Water in Kuwait: Methodological Issues and Empirical Results," *Water Resources Research*, 21, 433–438. [11]

Araral, E., and Wang, Y. (2013), "Water Demand Management: Review of Literature and Comparison in South-East Asia," *International Journal of Water Resources Development*, 29, 434–450. [10]

Arellano, M., and Bond, S. (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297. [11]

Arellano, M., and Bover, O. (1995), "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68, 29–51. [11]

Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171. [6,7]

——— (2009), "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279. [1,4]

Bun, M. J. G., and Kiviet, J. F. (2006), "The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models," *Journal of Econometrics*, 132, 409–444. [9]

Connor, G., and Korajczyk, R. A. (1986), "Performance Measurement With the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15, 373–394. [7]

Cragg, J. C., and Donald, S. G. (1997), "Inferring the Rank of a Matrix," *Journal of Econometrics*, 76, 223–250. [8,9]

De Vos, I., and Everaert, G. (2019), "Bias-Corrected Common Correlated Effects Pooled Estimation in Homogeneous Dynamic Panels," *Journal of Business & Economic Statistics*, DOI: 10.1080/07350015.2019.1654879. [2]

Everaert, G., and De Groote, T. (2016), "Common Correlated Effects Estimation of Dynamic Panels With Cross-Sectional Dependence," *Econometric Reviews*, 35, 428–463. [5]

Fan, J., and Liao, Y. (2019), "Learning Latent Factors From Diversified Projections and Its Applications to Over-Estimated and Weak Factors," arXiv no. 1908.01252. [5]

Gagliardini, P., and Gouriéroux, C. (2017), "Double Instrumental Variable Estimation of Interaction Models With Big Data," *Journal of Econometrics*, 201, 176–197. [5]

Gagliardini, P., Ossola, E., and Scaillet, O. (2016), "Time-Varying Risk Premium in Large Cross-Sectional Equity Data Sets," *Econometrica*, 84, 985–1046. [3]

Gato, S., Jayasuriya, N., and Roberts, P. (2007), "Temperature and Rainfall Thresholds for Base Use Urban Water Demand Modelling," *Journal of Hydrology*, 337, 364–376. [10]

Hansen, C., and Liao, Y. (2019), "The Factor-Lasso and k-Step Bootstrap Approach for Inference in High-Dimensional Economic Applications," *Econometric Theory*, 35, 465–509. [4]

Hastie, T., Tibshirani, R., and Friedman, J. (2017), *The Elements of Statistical Learning* (2nd ed.), New York: Springer. [8]

Hayakawa, K. (2016), "Identification Problem of GMM Estimators for Short Panel Data Models With Interactive Fixed Effects," *Economics Letters*, 139, 22–26. [1]

Holtz-Eakin, D., Newey, W. K., and Rosen, H. S. (1988), "Estimating Vector Autoregressions With Panel Data," *Econometrica*, 56, 1371–1395. [1]

House-Peters, L. A., and Chang, H. (2011), "Urban Water Demand Modeling: Review of Concepts, Methods, and Organizing Principles," *Water Resources Research*, 47, W05401. [10]

Hsiao, C., Pesaran, M. H., and Tahmiscioglu, A. K. (2002), "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," *Journal of Econometrics*, 109, 107–150. [2]

Hunt, E., Hubbard, K., Wilhite, D., Arkebauer, T., and Dutcher, A. (2009), "The Development and Evaluation of a Soil Moisture Index," *International Journal of Climatology*, 29, 747–759. [11]

Jiang, B., Yang, Y., Gao, J., and Hsiao, C. (2017), "Recursive Estimation in Large Panel Data Models: Theory and Practice," Mimeo. [1]

Juodis, A. (2018), "Pseudo Panel Data Models With Cohort Interactive Effects," *Journal of Business & Economic Statistics*, 36, 47–61. [1]

Juodis, A., Karabıyık, H., and Westerlund, J. (2020), "On the Robustness of the Pooled CCE Estimator," *Journal of Econometrics* (forthcoming). [5]

Juodis, A., and Sarafidis, V. (2018), "Fixed T Dynamic Panel Data Estimators With Multifactor Errors," *Econometric Reviews*, 37, 893–929. [1,4,10]

Kapetanios, G., Serlenga, L., and Shin, Y. (2020), "Estimation and Inference for Multidimensional Heterogeneous Panel Datasets With Hierarchical Multi-Factor Error Structure," *Journal of Econometrics* (forthcoming). [1]

Karabıyık, H., Urbain, J.-P., and Westerlund, J. (2019), "CCE Estimation of Factor-Augmented Regression Models With More Factors Than Observables," *Journal of Applied Econometrics*, 34, 268–284. [4]

Kleibergen, F. R., and Paap, R. (2006), "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126. [8]

Kruiniger, H. (2008), "Not So Fixed Effects: Correlated Structural Breaks in Panel Data," Mimeo. [1]

Kuersteiner, G., and Prucha, I. R. (2013), "Limit Theory for Panel Data Models With Cross Sectional Dependence and Sequential Exogeneity," *Journal of Econometrics*, 174, 107–126. [4]

Maidment, D. R., and Miaou, S. P. (1986), "Daily Water Use in Nine Cities," *Water Resources Research*, 22, 845–851. [10]

Moon, H. R., and Weidner, M. (2015), "Linear Regression for Panel With Unknown Number of Factors as Interactive Fixed Effects," *Econometrica*, 83, 1543–1579. [4]

Nauges, C., and Thomas, A. (2003), "Long-Run Study of Residential Water Consumption," *Environmental and Resource Economics*, 26, 25–43. [12]

Norkute, M., Sarafidis, V., Yamagata, T., and Cui, G. (2020), "Instrumental Variable Estimation of Dynamic Linear Panel Data Models With Defactored Regressors and a Multifactor Error Structure," *Journal of Econometrics* (forthcoming). [1]

Pesaran, M. H. (2006), "Estimation and Inference in Large Heterogeneous Panels With a Multifactor Error Structure," *Econometrica*, 74, 967–1012. [1,6]

Pesaran, M. H., Smith, L. V., and Yamagata, T. (2013), "Panel Unit Root Tests in the Presence of a Multifactor Error Structure," *Journal of Econometrics*, 175, 94–115. [4]

Robertson, D., and Sarafidis, V. (2015), "IV Estimation of Panels with Factor Residuals," *Journal of Econometrics*, 185, 526–541. [1,3,4,7,9]

Robertson, D., Sarafidis, V., and Westerlund, J. (2018), "Unit Root Inference in Generally Trending and Cross-Correlated Dynamic Panels," *Journal of Business & Economic Statistics*, 36, 493–504. [1]

Roodman, D. (2009), "How To Do xtabond2: An Introduction to Difference and System GMM in Stata," *The Stata Journal*, 9, 86–136. [11]

Sarafidis, V., and Wansbeek, T. J. (2012), "Cross-Sectional Dependence in Panel Data Analysis," *Econometric Reviews*, 31, 483–531. [1]

——— (2020), "Celebrating 40 Years of Panel Data Analysis: Past, Present and Future," *Journal of Econometrics* (forthcoming). [1]

Staiger, D., and Stock, J. H. (1997), "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65, 557–586. [8]

Westerlund, J., Petrova, Y., and Norkute, M. (2019), "CCE in Fixed-T Panels," *Journal of Applied Econometrics*, 34, 746–761. [1]

Windmeijer, F. (2005), "A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators," *Journal of Econometrics*, 126, 25–51. [10]

Zhu, H., Sarafidis, V., and Silvapulle, M. (2020), "A New Structural Break Test for Panels With Common Factors," *Econometrics Journal*, 23, 137–155. [10]