Econometric Theory, **0**, 0000, 1–47.
doi:10.1017/S0266466625100091

THREE-DIMENSIONAL FACTOR MODELS WITH GLOBAL AND LOCAL FACTORS

XUN LU
Chinese University of Hong Kong

SAINAN JIN
Tsinghua University

LIANGJUN SU
Tsinghua University

This article considers a three-dimensional latent factor model in the presence of one set of global factors and two sets of local factors. We show that the numbers of global and local factors can be estimated uniformly and consistently. Given the number of global and local factors, we propose a two-step estimation procedure based on principal component analysis (PCA) and establish the asymptotic properties of the PCA estimators. Monte Carlo simulations demonstrate that they perform well in finite samples. An application to the dataset of international trade reveals the relative importance of different types of factors.

1. INTRODUCTION

As part of the big data revolution, there has been a rapid emergence of multi-dimensional panel data sets. In particular, the use of three-dimensional (3D) panel data sets has gained tremendous momentum in the last decade. It has been frequently employed in empirical research in many economic fields, such as international trade, transportation, labor, housing, and migration, among others (see Matyas, 2017 for a review). Latent factor models provide an effective way of modeling panel data. They allow for unobserved heterogeneity and cross-sectional dependence of unknown form, both of which are fundamental features of economic and financial data. So far, the theories for factor models have been developed mostly for traditional two-dimensional (2D) panel data (see Bai and Ng, 2002; Stock and Watson, 2002; Bai, 2003; Onatski, 2010; Onatski, 2012; Ahn and Horenstein, 2013; Giglio and Xiu, 2021, among others). This article develops

The authors thank the Editor, Peter C.B. Phillips, a Co-Editor, and two anonymous referees for their constructive comments. Lu acknowledges support from the Hong Kong Research Grants Council (RGC) under grant number 14500121 and Chinese University of Hong Kong for the start-up fund. Su thanks the National Natural Science Foundation of China (NSFC) for financial support under the grant number 72133002. All errors are the authors' sole responsibilities. Address correspondence to Liangjun Su, School of Economics and Management, Tsinghua University, Beijing, China, e-mail: sulj@sem.tsinghua.edu.cn.

theories for general 3D factor models with both global and local factors and applies them to empirical studies.

Specifically, we consider estimation and inference for the following 3D model:

$$y_{ijt} = \lambda_{ij}^{(0)\prime} f_t^{(0)} + \lambda_{ij}^{(1)\prime} f_{it}^{(1)} + \lambda_{ij}^{(2)\prime} f_{jt}^{(2)} + u_{ijt},\ i \in [N],\ j \in \mathcal{M}_i,\ t \in [T], \quad \textbf{(1.1)}$$

where y_{ijt} is the observable data, $f_t^{(0)}$ is the global factor, $f_{it}^{(1)}$ and $f_{jt}^{(2)}$ are the local factors that depend on i and j, respectively, $\lambda_{ij}^{(0)}, \lambda_{ij}^{(1)}$, and $\lambda_{ij}^{(2)}$ are the corresponding factor loadings, u_{ijt} is the idiosyncratic error, and $[a] = \{1, 2, ..., a\}$ for any positive integer a. Let $M_i = |\mathcal{M}_i|$, the cardinality of the set $\mathcal{M}_i \subset [M]$, where $M = \max\{M_i, i \in [N]\}$.[1] We assume that the dimensions of $f_t^{(0)}, f_{it}^{(1)}$, and $f_{jt}^{(2)}$ are $r^{(0)} \times 1, r_i^{(1)} \times 1$, and $r_j^{(2)} \times 1$, respectively. That is, there are $r^{(0)}, r_i^{(1)}$, and $r_j^{(2)}$ global factors, i-specific local factors and j-specific local factors, respectively. The factors, factor loadings, and factor numbers are all unknown. We treat the factors as random and factor loadings as non-random following the literature (see, e.g., Bai, 2003 Bai and Ng, 2023). We consider large panels where the three dimensions (N, M_i, T) go to infinity jointly. The goal of this article is to determine $(r^{(0)}, r_i^{(1)}, r_j^{(2)})$, and to estimate and conduct inferences for $(f_t^{(0)\prime}, f_{it}^{(1)\prime}, f_{jt}^{(2)\prime})$ and $(\lambda_{ij}^{(0)\prime}, \lambda_{ij}^{(1)\prime}, \lambda_{ij}^{(2)\prime})$ up to certain rotation matrices.

1.1. Examples of 3D Factor Models and Their Usefulness

There are numerous potential applications of this type of model. Below are three examples in economic growth, trade, and macroeconomics. Note that our model (1.1) nests some simple models that have already been employed in empirical research, e.g., in Koren and Tenreyro (2007) and Andrade and Zachariadis (2016) as discussed in Examples 1 and 3, respectively, which follows. However, the general estimation and inference methods for model (1.1) are missing in the literature.

Example 1 (Economic Growth). One important question in development economics is to decompose the volatility of economic growth into different sources, such as country factors and industry factors (see, e.g., Koren and Tenreyro, 2007).[2] Let y_{ijt} be the growth rate of value-added per worker for industry j in country i in year t. $f_t^{(0)}$ is the global factor that affects all industries and countries, e.g., the COVID-19 pandemic or a global financial crisis. $f_{it}^{(1)}$ and $f_{jt}^{(2)}$ are country and industry specific factors, respectively. For example, let i be China and j be the mining industry. $f_{it}^{(1)}$ is a China-specific factor that only affects all industries

[1] Here, i and j are symmetric. Equivalently, we can write the indices as $j \in [M], i \in \mathcal{N}_j, and\ t \in [T]$.

[2] Koren and Tenreyro, 2007 consider a simpler model than ours, where there are no global factors, and factor loadings of country factors and industry factors are assumed to be constant. Using our notations, their model is specified as $y_{ijt} = f_{it}^{(1)} + f_{jt}^{(2)} + u_{ijt}$.

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

in China (not other countries), e.g., China's entry into the WTO or a major earthquake. $f_{jt}^{(2)}$ is the mining-industry-specific factor that only affects the mining industry (not other industries) in all countries, e.g., giant oil discoveries. The effects of those factors are captured by their factor loadings, which depend on the country and industry. In this application, we could examine how the volatility of global factor components, country factor components and industry factor components contribute to the overall economic volatility of a country, which is important for industrial policy and risk management, as discussed in Koren and Tenreyro (2007).

Example 2 (International/Firm-to-Firm Trade). Consider the international trade data, where y_{ijt} is the volume of trade from source (export) country i to destination (import) country j in year t. $f_t^{(0)}$ is the global factor that affects every trade volume, and its effect ($\lambda_{ij}^{(0)}$) depends on the specific source country i and destination country j. For example, worldwide technological advancement in transportation can be thought of as a global factor, which is likely to affect all trade volumes between any pair of countries. The local factors $f_{it}^{(1)}$ and $f_{jt}^{(2)}$ represent source country and destination country factors, respectively: $f_{it}^{(1)}$ is the factor of the i th exporting country (say, China) that affects the trade volumes from China to all destination countries; $f_{jt}^{(2)}$ is the factor of the jth importing country (say, the U.S.) that affects all trade volumes to the U.S. from all other countries. If China's export price level increases (say, $f_{it}^{(1)}$ increases), its export volume to all other countries may decrease; if the U.S. economy booms (say, $f_{jt}^{(2)}$ increases), it may increase its imports from all other countries.[3] Here, the local factors are unobservable, and they can also be thought of as time-varying country fixed effects, which have been argued to be important both theoretically and empirically for the gravity models (see, e.g., Feenstra, 2015). In Section S8 of the Supplementary Material, we show that our model matches the structural gravity model considered in Anderson and van Wincoop (2003) and each element in our model is economically meaningful. In particular, we show that the global factor $f_t^{(0)}$ represents the technological improvement in reducing trading costs, the source country factor $f_{it}^{(1)}$ represents the export price level and the destination country factor $f_{jt}^{(2)}$ includes the destination country's income and its multi-lateral resistance level.[4] The global factor loading includes elements of an individual country's heterogeneous preference parameter (the elasticity of substitution) and geographic distance between countries. The two local factor loadings include elements of the heterogeneous preference parameters. The same framework can be applied to firm-to-firm trading, where y_{ijt} denotes the sales from firm i (supplier) to firm j (purchaser), $f_t^{(0)}$ represents the factor

[3]Here, we discuss the total export volume from China to the U.S. We may also study more detailed industry-level exports. Then, the data will become four-dimensional with the fourth dimension being industry.

[4]Multilateral resistance is a term proposed in Anderson and van Wincoop (2003) and can be thought of as an index of bilateral trade costs. Its precise definition can be found in Section S8 of the Supplementary Material or Anderson and van Wincoop (2003).

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

that affects every firm's sales (e.g., a macroeconomic variable), and $f_{it}^{(1)}$ and $f_{jt}^{(2)}$ represent firm specific supply-side and demand-side factors, respectively. Dhyne et al. (2021) and Bernard et al. (2022) consider such buyer-supplier data among all firms in Belgium during 2002–2014.

Example 3 (Macroeconomics). Factor models are often used to study the global co-movements of inflation rates. For example, following Andrade and Zachariadis (2016) (2016, (AZ) hereafter) but with our notation, let y_{ijt} be the inflation rate of product item j in location i at date t. For example, j can be apple, mineral water, and annual premium for car insurance, among others. Table 1 in (AZ) list 276 products (goods and services) in their sample. The location i refers to city i in different countries. Table 1 in (AZ) includes 88 cities in 59 countries. The date t in (AZ)'s sample is semi-annual from 1990 to 2010. (AZ) decomposes the global inflation rates exactly into the same four components as in our model (1.1) and provide interpretation for the three different types of factors. $f_t^{(0)}$ is the global factor that affects every price in every location. (AZ) explain that "typical examples of such global factors would be oil prices or global liquidity shocks associated with worldwide money supply". $f_{it}^{(1)}$ is the location-level local factor that affects every price in a given location i. (AZ) also explain that "typical examples of such local macro factors are monetary or fiscal policies". $f_{jt}^{(2)}$ is the product-level local factors that affect a given product j in every location, and an example includes "technological innovation specific to a given product" (AZ). The factor loadings measure the effects of those factors, which are product and location-specific. (AZ) assume that the numbers of factors are all 1 and use different sample averages to proxy different factors in their estimation. We allow the number of factors to be data-driven and propose a PCA-type estimation method. In addition, (AZ) is an applied paper without much econometric theory.

There are many other examples for y_{ijt}, including the retail price at a supermarket chain i in region j, the foreign direct investment from region i to region j, the total value of bilateral asset flows (assets of region i bought by agents of region j), and the number of immigrants from region i to region j, among others. The model considered here can be used in various ways. First, it provides an effective way of reducing the dimensionality and summarizing information for large data sets. Here, the dimensionality of the original time series data ($\sum_{i=1}^{N} M_i$) is reduced to the total number of factors: $r^{(0)} + \sum_{i=1}^{N} r_i^{(1)} + \sum_{j=1}^{M} r_j^{(2)}$, where $M = \max_{1 \le i \le N} M_i$. For example, for a balanced panel with $N = 60, M_i = M = 60, r^{(0)} = 1, r_i^{(1)} = 1$, and $r_j^{(2)} = 1$, the dimensionality is reduced from 3600 to 121. Second, it provides useful variance decomposition in the spirit of analysis of variance (ANOVA). Assuming that the global factors and local factors are uncorrelated, and two types of local factors are uncorrelated, we can decompose the variance of y_{ijt} into four parts.

$$\mathrm{Var}\left(y_{ijt}\right) = \mathrm{Var}(c_{ijt}^{(0)}) + \mathrm{Var}(c_{ijt}^{(1)}) + \mathrm{Var}(c_{ijt}^{(2)}) + \mathrm{Var}(u_{ijt}),$$

TABLE 1. Comparison of estimators of global factors

	$\text{Bias}^2 \times 10^4$			$\text{Variance} \times 10^4$			$\text{MSE} \times 10^4$		
	1st-step (initial)	2nd-step (final)	Orcale	1st-step (initial)	2nd-step (final)	Orcale	1st-step (initial)	2nd-step (final)	Orcale
				DGP 1					
(50,50,50)	0.19	0.03	0.02	31.40	4.98	4.07	31.59	5.01	4.09
(50,100,50)	0.09	0.02	0.01	19.36	2.39	1.96	19.46	2.40	1.97
(100,100,50)	0.05	0.01	0.00	10.56	1.18	0.99	10.61	1.19	1.00
(50,50,100)	0.10	0.02	0.02	24.56	4.56	4.04	24.66	4.58	4.05
(50,100,100)	0.06	0.01	0.01	14.19	2.16	1.95	14.25	2.17	1.96
(100,100,100)	0.04	0.01	0.00	7.46	1.07	0.99	7.51	1.08	0.99
				DGP 2					
(50,50,50)	1.49	0.51	0.01	173.87	5.23	1.98	175.36	5.74	1.99
(50,100,50)	0.97	0.36	0.00	144.07	3.03	0.95	145.04	3.39	0.96
(100,100,50)	0.61	0.13	0.00	83.10	1.46	0.49	83.71	1.58	0.49
(50,50,100)	1.52	0.56	0.01	166.83	3.98	1.97	168.35	4.54	1.98
(50,100,100)	1.04	0.39	0.00	136.75	2.06	0.96	137.79	2.45	0.96
(100,100,100)	0.54	0.15	0.00	83.34	1.03	0.50	83.88	1.17	0.50

Note: Numbers in the main entries are the bias, variance, and MSE of the three estimators: the 1st-step estimator, the 2nd-step estimator, and the oracle estimator. The two steps are described in Algorithm 2.3. The oracle estimator is the infeasible one assuming that the local factor components are absent and the number of global factors is known.

where $c_{ijt}^{(0)} = \lambda_{ij}^{(0)\prime} f_t^{(0)}, c_{ijt}^{(1)} = \lambda_{ij}^{(1)\prime} f_{it}^{(1)}$, and $c_{ijt}^{(2)} = \lambda_{ij}^{(2)\prime} f_{jt}^{(2)}$ are three factor components, and study their relative contributions. In the above economic growth example, we can examine whether the country factors or industry factors are important in terms of explaining the variations of economic growth of countries. Third, the underlying unobservable factors may contain useful economic information. For example, in the international trade application discussed above, $f_t^{(0)}$ can be interpreted as the global driving force of international trade flows, and $f_{it}^{(1)}$ and $f_{jt}^{(2)}$ are the specific country factors, such as the export price level, the income, and the multilateral resistance level. Fourth, the estimated factors can be thought of as diffusion indices and used to improve forecasting accuracy (see, e.g., Stock and Watson, 2002; Bai and Ng, 2006 Cheng and Hansen, 2015). Fifth, our model provides foundations for the more general model with observable exogenous regressors, i.e.,

$$y_{ijt} = \beta_{ij}' x_{ijt} + \lambda_{ij}^{(0)\prime} f_t^{(0)} + \lambda_{ij}^{(1)\prime} f_{it}^{(1)} + \lambda_{ij}^{(2)\prime} f_{jt}^{(2)} + u_{ijt},\ i \in [N],\ j \in \mathcal{M}_i,\ t \in [T], \tag{1.2}$$

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

where x_{ijt} is a $K \times 1$ vector of observable regressors and β_{ij} is the corresponding vector of slope coefficients. For example, in the international trade application above, x_{ijt} could be the observed trading costs from country i to country j in year t. Another example of x_{ijt} is the lagged dependent variable, that is, $x_{ijt} = y_{ij,t-1}$ in the dynamic model. This model is often referred to as panel data models with interactive fixed effects, as we allow the regressor x_{ijt} to be correlated with the unobservable factors. Kapetanios, Serlenga, and Shin (2021) consider model (1.2) and propose a common correlated effects (CCE) estimation approach to estimate β_{ij}, as in Pesaran (2006).

Despite the advantages of our model discussed above, several limitations warrant mention. First, like all factor models, ours is fundamentally statistical and "agnostic" in nature and may not fully capture complex economic relationships implied by theoretical frameworks. For instance, while we attempt to align our model with a structural gravity framework in Section S8 of the Supplementary Material, it cannot fully account for intricate cross-country economic linkages. Second, our current specification is a pure factor model that excludes observable variables. This limitation could be addressed through extensions such as Model (1.2), which incorporates observable variables linearly. Alternatively, one might allow the factors or their loadings to depend on observable variables (e.g., Fan, Ke, and Liao, 2016a; Fan, Liao, and Wang, 2016b and Kelly, Pruitt, and Su, 2019). The optimal approach for integrating observable variables while maintaining theoretical coherence remains an open question that likely depends on specific empirical applications.

1.2. Related Literature

There is an enormous literature on the 2D factor models. Omitting the j index, model (1.1) reduces to $y_{it} = \lambda_i' f_t + u_{it}$, where f_t and λ_i are $r \times 1$ vectors of factors and factor loadings, respectively. There is only one type of factor here. The literature on 2D factor models has been developing rapidly (see Bai and Wang (2016) for a detailed review). For the pure factor models, Bai (2003) considers estimation based on PCA and develops the inference theory by assuming the number of factors r is known. Various methods have been proposed to determine r (see, e.g., the information criterion (IC) method of Bai and Ng (2002)), the edge-distribution method of Onatski (2010), and the eigenvalue ratio (ER) and growth ratio (GR) methods of Ahn and Horenstein (2013) (2013, (AH) hereafter). For the model with exogenous regressors, Pesaran (2006) and Chudik and Pesaran (2015) consider CCE estimation; Bai (2009), Moon and Weidner (2015), Lu and Su (2016), and Moon and Weidner (2017) study PCA-based Gaussian quasi-maximum likelihood estimators.

There are a limited number of papers on 3D pure factor modelsFirst, some papers assume that numbers of global factors (see Dias, Pinheiro, and Rua, 2013; Wang, 2014; Breitung and Eickmeier, 2016;; Ando and Bai, 2017; Choi et al., 2018; Andreou et al., 2019; Han, 2021; Chen, 2023; Choi, Lin, and Shin, 2023 Gao and

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

Tsay, 2023. In all these papers, there is only one local factor component,[5] i.e.,

$$y_{ijt} = \lambda_{ij}^{(0)\prime} f_t^{(0)} + \lambda_{ij}^{(1)\prime} f_{it}^{(1)} + u_{ijt},\ i \in [N],\ j \in \mathcal{M}_i,\ t \in [T]. \tag{1.3}$$

This model is often referred to as a multidimensional or multilevel/hierarchical factor model depending on whether there is a nested relationship between the two crosssection indices. According to Yang and Schmidt (2021), the concept of nesting corresponds to the distinction in the literature between a multidimensional model (not nested) and a multilevel or hierarchical model (nested). For the nested case, for example, i and j may refer to industries and firms, respectively; any firm j must belong to certain industry i that may have M_i firms so that $\sum_{i=1}^{N} M_i$ denotes the total number of firms. Even for the model in (1.3), there are certain limitations of the existing methods. First, some papers assume that numbers of global factors $\left(r^{(0)}\right)$ and local factors $(r_i^{(1)})$ are known (see, e.g., Wang, 2014; Breitung and Eickmeier, 2016. Choi et al. (2018) use IC to select the number of local factors $(r_i^{(1)})$ assuming that the number of global factors $\left(r^{(0)}\right)$ is known. Dias et al. (2013) consider IC for both $r^{(0)}$ and $r_i^{(1)}$, but their method is not practical when N is large. Choi et al. (2023) develop a novel method to determine both $r^{(0)}$ and $r_i^{(1)}$ in (1.3) consistently using canonical correlation analysis (CCA). Second, most of the papers do not provide inference theory for the factors and factor loadings except Andreou et al. (2019).[6] They only establish the consistency of their estimators, and it is unclear how to conduct inferences. Third, some papers impose strong assumptions on local factors $f_{it}^{(1)}$. For example, Choi et al. (2018) and Han (2021) assume that the local factors are uncorrelated, i.e., $\mathrm{Cov}(f_{i_1,t}^{(1)}, f_{i_2,t}^{(1)}) = 0$ for $i_1 \neq i_2$, which may not be satisfied in practice. In contrast, Andreou et al. (2019), Chen (2023), and Choi et al. (2023) allow non-zero correlations in the local factors. All the methods mentioned above are based on PCA or CCA. Alternatively, Moench and Ng (2011) and Moench, Ng, and Potter (2013) propose estimation methods based on the MCMC algorithm for dynamic hierarchical factor models. As discussed above, multidimensional and hierarchical factor models are special cases of our model in terms of model specification. Our method can be applied to (1.3) with little modification if (N, M_i, T) pass to infinity jointly and $\sum_{i=1}^{N} M_i = o(T^2)$. However, our method is not applicable if one cross-sectional dimension is fixed, say, in a dataset with a small number of industries but a large number of firms.

There is a rapidly growing literature on 3D panel data models with exogenous regressors. Matyas (2017) provides an excellent review on both econometric

[5]Superficially, in terms of model specification, this model is a special case of ours. Nevertheless, the detailed conditions are different. In particular, we need all three dimensions, $N, M,$ and T to pass to infinity in order to identify all types of factors and loadings. In the model with only one type of local factors, often one cross-sectional dimension is fixed.

[6]Andreou et al. (2019) consider model (1.3) with $N = 2$ and they consider both estimation and inference theory. In addition, there are inference theories for other types of 3D factor models. For example, Freeman (2022) and Babii, Ghysels, and Pan (2024) consider inference for tensor factor models which we review below.

theories and empirical applications. Most of the research on 3D panel data models focuses on linear models with different specifications of fixed effects. Lu, Miao, and Su (2021) consider seven commonly used specifications of fixed effects and provide a cross-validation method to determine the correct specifications. The most general 3D linear fixed effect model considered in Lu et al. (2021) is $y_{ijt} = \beta' x_{ijt} + \gamma_{ij} + \alpha_{it} + \alpha^*_{jt} + u_{ijt}$, where x_{ijt} is a vector of observable regressors and β is the corresponding slope coefficients, γ_{ij}, α_{it}, and α^*_{jt} are fixed effects. The fixed effect specification in this model can be thought of as a special case of our model (1.1). Chiang, Rodrigue, and Sasaki (2023) consider post-selection inferences for a subset of the models studied in Lu et al. (2021). Kapetanios et al. (2021), Feng et al. (2024), and Jin et al.2025 consider panels with exogenous regressors and factor structures.

In the recent literature, there are studies on 3D panel data model with tensor factor structures or matrix-valued time series. There are two types of tensor factor models: the canonical polyadic (CP) tensor factor model (see, e.g., Freeman, 2022; Babii et al., 2024; Chen, Han, and Yu, 2024 Han et al., 2024) and the Tucker tensor factor model (see, e.g., Wang, Liu, and Chen, 2019; Chen, Tsay, and Chen, 2020; Chen, Yang, and Zhang, 2022; Lettau, 2022; Yu et al., 2022 He et al., 2024). The CP model can be written as

$$y_{ijt} = \sum\nolimits_{\ell=1}^{r} \varphi_{i\ell} \psi_{j\ell} f_{t\ell} + u_{ijt}, \tag{1.4}$$

where $\varphi_{i\ell}$, $\psi_{j\ell}$, and $f_{t\ell}$ are all scalars and r is generally assumed to be fixed. So in this model, there are r global factors $\{f_{t\ell}\}_{\ell=1}^{r}$ with factor loadings $\{\lambda_{ij,\ell} \equiv \varphi_{i\ell}\psi_{j\ell}\}_{\ell=1}^{r}$. The Tucker model can be written as

$$y_{ijt} = \Phi_i' f_t \Psi_j + u_{ijt},$$

where the factors f_t is an $r \times R$ matrix, Φ_i and Ψ_j are $r \times 1$ and $R \times 1$ factor loadings, respectively. Note that CP model is a special case of Tucker model if we impose $r = R$ and that f_t is a diagonal matrix. See, e.g., Babii et al., 2024 and Han et al., 2024 who discuss the differences between these two tensor factor models. There are three main differences between these two types of tensor models and ours. First, they only consider global factors in our framework, while we allow both global factors and two types of local factors. Second, for the global factor loadings, they impose a certain multiplicative form, while we do not impose any restrictions. Third, the number of parameters in the tensor models is of order $O(N+M+T)$ while it is $O(NM+NT+MT)$ in our model. In general, the tensor factor models are more parsimonious than ours. There is a trade-off between model generality and parsimony, and which model is more appropriate depends on specific applications. In Section S1.1 of the Supplementary Material, we provide more discussions on the model specification.

1.3. Contributions of the Article

To the best of our knowledge, this article is the first systematic study of the general 3D factor model in (1.1), which has many potential applications. We make both methodological and empirical contributions, which we discuss separately below.

On the methodological side, we develop theories to determine the number of factors, to estimate the factors and loadings, and to make inferences on them. First, we show how the global and local factors can be identified and extend (AH)'s ER or GR statistics to determine the number of global factors and local factors sequentially. Intuitively, for the complicated model in (1.1), the singular values corresponding to the global factors are larger than those from the local factors and error terms in order of magnitude so that the global factors are well separated from the local factors and error terms. See Remark 1 in Section 2.1 for the discussion and the proofs of Lemma S4.1 in the Supplementary Material and Theorem 3.1 in the for details. Once the global factor component is identified, it can be treated as known, allowing for the identification of the two types of local factors, similar to the standard 2D case. We establish the consistency of the ER and GR estimators for the number of global factors, as well as the uniform consistency of the estimators for the number of local factors, a result that requires a more intricate analysis as the set of the total number of local factors in $\{r_i^{(1)}\}_{i=1}^N$ and $\{r_j^{(2)}\}_{j=1}^M$ increases with the sample size.

Second, we provide estimators for all the factors and loadings and show their consistency up to a certain rotation matrices. Our estimation method is a PCA-based two-step procedure. In the first step, we disregard the presence of local factors and consistently estimate the global factor component at a slow rate. Subsequently, by subtracting the estimated global factor component, we obtain consistent estimators of the two local factor components. In the second step, we re-estimate the global and local factors by regressing y_{ijt} on all the loading estimators from the first step, after which we similarly update the loading estimators. We show that the mean squared error (MSE) rate for the first-step global factor estimator is given by $O_p(N^{-1}+M^{-1}+T^{-2})$ whereas that for the second-step global factor estimator is improved to $O_p((NM)^{-1}+T^{-2})$. See the proofs of Lemma S4.1(i) in the Supplementary Material and Theorem 4.1(i) in the Third, we develop the asymptotic normal and inference theories for our estimators in the presence of both cross-sectional and serial dependence in the error terms. The results are substantially different from those in the 2D case. For example, there are two main differences in the asymptotic properties of the estimator of the global factors: (1) we now have an $O(\frac{1}{T})$ bias term that is non-vanishing asymptotically unless $\frac{\sqrt{NM}}{T}=o(1)$, and (2) the estimation of the local factors may or may not have asymptotically nonnegligible effect on the estimator of the global factors, and in some extreme cases, the second-step estimator of the global factor enjoys the oracle property. See Remark 9 for details. In contrast, due to the fact that the global factors can be estimated at a fast rate, their estimation does not have asymptotic effect on

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

the estimation of the two types of local factors. We also propose a method to correct the bias in the global factor estimators and conduct valid inference based on the asymptotic normality. Given the complicated structure of the model, the theoretical development is obviously challenging.

On the empirical side, we apply our methods to the international trade flow data, where y_{ijt} is the growth of the trade flow from source country i to destination country j in year t. We find that there is one global factor and most local factor numbers are 0 or 1. The global factor, source country factor, and destination country factor contribute to 14.4%, 21.0%, and 28.1% of the total sample variance, respectively. The estimated global factor is closely related to the lagged world economic growth and the world openness index.

1.4. Roadmap

The rest of the article is structured as follows. In Section 2, we discuss the identification of the model and propose the methods for determining the number of factors and estimating factors and factor loadings. In Section 3, we study the asymptotic properties of the estimators of the number of factors. In Section 4, we show the asymptotic distribution of the estimators of the factors and factor loadings. Section 5 reports Monte Carlo simulation results. In Section 6, we apply our method to study international trade flow data. Section 7 concludes. Section S1 of the Supplementary Material provides more discussions for the model specification and some additional notations. Sections S2–S7 of the Supplementary Material contain the proofs of the main results in the article. Section S8 of the Supplementary Material discusses a simple gravity model with heterogeneous preferences. Sections S9 and S10 of the Supplementary Material contain some additional simulation and empirical results, respectively.

Notation. By symmetry, we assume that for each $j, i \in \mathcal{N}_j \subset [N]$, where $\left|\mathcal{N}_j\right| = N_j, N \equiv \max\left\{N_j, j \in [M]\right\}$ and $\equiv$ signifies a definitional relationship. Let $\bar{N} = \frac{1}{M}\sum_{j=1}^{M} N_j$ and $\bar{M} = \frac{1}{N}\sum_{i=1}^{N} M_i$. For balanced panels, $M_i = \bar{M} = M, N_j = \bar{N} = N$. Let $\sum_{i,j,t} = \sum_{i=1}^{N}\sum_{j\in\mathcal{M}_i}\sum_{t=1}^{T} = \sum_{j=1}^{M}\sum_{i\in\mathcal{N}_j}\sum_{t=1}^{T}$. Define $\sum_{i,j}, \sum_{i,t}, \sum_{j,t}, \sum_{i}, \sum_{j}$, and $\sum_{t}$ analogously. Let $\max_i = \max_{1\le i\le N}$ and $\max_j = \max_{1\le j\le M}$. Define $\min_i$ and $\min_j$ analogously. Let $\mathbb{I}_a$ denote an $a \times a$ identity matrix. For a real $m \times n$ matrix $A = (a_{ij})$, we use $\|A\|$ and $\|A\|_{\text{sp}}$ to denote its Frobenius norm and spectral norm, respectively: $||A||_{\text{sp}} \equiv \mu_1(A)$, and $||A|| \equiv (\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2)^{1/2}$, where $\mu_k(A)$ denotes the kth largest singular value of A for $k \le m \wedge n \equiv \min(m,n)$ and $\equiv$ signifies a definitional relationship. For a real symmetric matrix A, we use $A \ge 0$ and $A > 0$ to denote that A is positive semidefinite (p.s.d.) and positive definite, respectively. We use $\psi_k(A)$ to denote the kth largest eigenvalue of a symmetric p.s.d. matrix A. For two numbers a and b, let $a \wedge b = \min(a,b)$ and $a \vee b = \max(a,b)$. We use $\text{bdiag}(A_1, ..., A_N)$ to denote a block diagonal matrix with diagonal blocks given by $A_1, ..., A_N$.

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

2. IDENTIFICATION AND ESTIMATION

2.1. Identification

Throughout, we assume that the factor loadings are nonstochastic while the factors are random. To identify the factors and factor loadings, we try to impose as minimal conditions as possible on $\{f_t^{(0)}\}, \{f_{it}^{(1)}\}$, and $\{f_{jt}^{(2)}\}$, as stated in Assumption I.

Assumption I.

(i) $\mathbb{E}[f_{it}^{(1)}] = 0$ and $\mathbb{E}[f_{jt}^{(2)}] = 0$ for all i,j and t;

(ii) $\mathbb{E}[f_t^{(0)} f_{it}^{(1)\prime}] = 0$ and $\mathbb{E}[f_t^{(0)} f_{jt}^{(2)\prime}] = 0$ for all i, j and t, and $\max_t\{\frac{1}{N}\sum_{i,i_1} ||\mathbb{E}(f_{it}^{(1)} f_{i_1t}^{(1)\prime})|| + \frac{1}{M}\sum_{j,j_1} ||\mathbb{E}(f_{jt}^{(2)} f_{j_1t}^{(2)\prime})||\} \leq C < \infty$;

(iii) $\mathbb{E}[f_t^{(0)} u_{ijt}] = 0, \mathbb{E}[f_{it}^{(1)} u_{ijt}] = 0,$ and $\mathbb{E}[f_{jt}^{(2)} u_{ijt}] = 0$ for all i, j and t;

(iv) $\{(f_t^{(0)}, f_{it}^{(1)}, f_{jt}^{(2)}), t \geq 1\}$ is weakly stationary, $\mathbb{E}[f_t^{(0)} f_t^{(0)\prime}] > 0, \mathbb{E}[f_{it}^{(1)} f_{it}^{(1)\prime}] > 0$ and $\mathbb{E}[f_{jt}^{(2)} f_{jt}^{(2)\prime}] > 0$ for all i, j and t;

(v) $\mathbb{E}[f_{it}^{(1)} f_{jt}^{(2)\prime}] = 0$ for all (i, j).

Assumption I(i) assumes that the local factors have mean zero while leaving the mean of the global factors unspecified. If either $\mathbb{E}(f_{it}^{(1)})$ or $\mathbb{E}(f_{jt}^{(2)})$ is nonzero but constant over time, we can recenter the local factors around their expectations to achieve zero mean. See Remark 2, which follows. The uncorrelation assumption between the global and local factors in Assumption I(ii) can be thought of as a normalization, as we discuss in Remark 3, which follows. A similar assumption has been made in the literature on multi-level factor models (see Wang, 2014; Breitung and Eickmeier 2016; Choi et al., 2018; Han, 2021, among others). In particular, Wang (2014) studies the identification of a two-level factor model where the two sets of factors are labeled as global and sector factors. He finds that the uncorrelation between the global and sector factors is necessary for the separate identification of the two sets of factors. In addition, the uncorrelation assumption makes it possible to conduct a variance decomposition in the spirit of ANOVA. This is important for the determination of the relative importance of the global and local factors in explaining the underlying response variable. Note that we do not require that $f_{i_1t}^{(1)}$ and $f_{i_2t}^{(1)}$ with $i_1 \neq i_2$ (or $f_{j_1t}^{(2)}$ and $f_{j_2t}^{(2)}$ with $j_1 \neq j_2$) be uncorrelated, which significantly relaxes the uncorrelation condition among local factors in Choi et al. (2018) and Han (2021). But as reflected in the second part of Assumption I(ii), we do require weak cross-sectional dependence among the local factors in order to separate the local factors from the global ones. In the presence of strong cross-sectional dependence in the local factors, it becomes difficult to separate the local factors from the global ones. For example, if $f_{it}^{(1)} = g_{1t}\mathbf{1}\{i \in G_1\} + g_{2t}\mathbf{1}\{i \in G_2\}$, where G_1 and G_2 form a partition of $[N], g_{1t}$ and g_{2t} have finite second moments, and $\mathbf{1}\{\cdot\}$ is the usual indicator function, then it is easy to see $\frac{1}{N}\sum_{i,i_1} ||\mathbb{E}(f_{it}^{(1)} f_{i_1t}^{(1)\prime})||$ is generally divergent to infinity at rate N so that the second part of Assumption I(ii) is violated. In this extreme case, $\{(g_{1t}, g_{2t})\}$ can be treated as a part of the global factors with sparse loadings.

Assumption I(iii) assumes orthogonality between the factors and the idiosyncratic error terms. Assumption I(iv) assumes weak stationarity and positive definiteness of certain matrices. Assumption I(v) requires the two types of local factors to be uncorrelated. See Remark 4 for further discussion.

We can rewrite model (1.1) as

$$y_{ijt} = \lambda_{ij}^{(0)\prime} f_t^{(0)} + u_{ijt}^{(0)}, \text{ where } u_{ijt}^{(0)} = \lambda_{ij}^{(1)\prime} f_{it}^{(1)} + \lambda_{ij}^{(2)\prime} f_{jt}^{(2)} + u_{ijt}, \tag{2.1}$$

$$u_{ijt}^{(0)} = \lambda_{ij}^{(1)\prime} f_{it}^{(1)} + u_{ijt}^{(1)}, \text{ where } u_{ijt}^{(1)} = \lambda_{ij}^{(2)\prime} f_{jt}^{(2)} + u_{ijt}, \tag{2.2}$$

$$u_{ijt}^{(0)} = \lambda_{ij}^{(2)\prime} f_{jt}^{(2)} + u_{ijt}^{(2)}, \text{ where } u_{ijt}^{(2)} = \lambda_{ij}^{(1)\prime} f_{it}^{(1)} + u_{ijt}. \tag{2.3}$$

In matrix form, we can write

$$Y = F^{(0)} \Lambda^{(0)\prime} + U^{(0)}, \tag{2.4}$$

$$U_i^{(0)} = F_i^{(1)} \Lambda_i^{(1)\prime} + U_i^{(1)} \text{ for } i \in [N], \tag{2.5}$$

$$U_{\cdot j}^{(0)} = F_j^{(2)} \Lambda_{\cdot j}^{(2)\prime} + U_{\cdot j}^{(2)} \text{ for } j \in [M], \tag{2.6}$$

where

$$\begin{aligned}
\underset{T\times 1}{y_{ij}} &= (y_{ij1},\ldots,y_{ijT})', \quad \underset{T\times N\bar{M}}{Y} = \{y_{ij}\}_{i\in[N],\, j\in\mathcal{M}_i}, \\
\underset{T\times r^{(0)}}{F^{(0)}} &= (f_1^{(0)},\ldots,f_T^{(0)})', \quad \underset{N\bar{M}\times r^{(0)}}{\Lambda^{(0)}} = \{\lambda_{ij}^{(0)}\}_{i\in[N],\, j\in\mathcal{M}_i}, \\
\underset{T\times 1}{u_{ij}^{(0)}} &= (u_{ij1}^{(0)},\ldots,u_{ijT}^{(0)})', \underset{T\times M_i}{U_i^{(0)}} = \{u_{ij}^{(0)}\}_{j\in\mathcal{M}_i,\, t\in[T]}, \underset{T\times N_j}{U_{\cdot j}^{(0)}} = \{u_{ij}^{(0)}\}_{i\in\mathcal{N}_j,\, t\in[T]}, \\
\underset{T\times r_i^{(1)}}{F_i^{(1)}} &= (f_{i1}^{(1)},\ldots,f_{iT}^{(1)})', \quad \underset{M_i\times r_i^{(1)}}{\Lambda_i^{(1)}} = \{\lambda_{ij}^{(1)}\}_{j\in\mathcal{M}_i}, \\
\underset{T\times r_j^{(2)}}{F_j^{(2)}} &= (f_{j1}^{(2)},,\ldots,f_{jT}^{(2)})', \quad \underset{N_j\times r_j^{(2)}}{\Lambda_{\cdot j}^{(2)}} = \{\lambda_{ij}^{(2)}\}_{i\in\mathcal{N}_j},
\end{aligned}$$

and the definitions of $U^{(0)}, U_i^{(1)}$, and $U_{\cdot j}^{(2)}$ are analogous to those of $Y, U_i^{(0)}$, and $U_{\cdot j}^{(0)}$, respectively. Below, we will write $Y = \left\{y_{ijt}\right\}$, $\Lambda^{(0)} = \{\lambda_{ij}^{(0)}\}$, $F^{(0)} = \left\{f_t^0\right\}$, $F_i^{(1)} = \{f_{it}^{(1)}\}$, $\Lambda_i^{(1)} = \{\lambda_{ij}^{(1)}\}$, etc., Similar notations apply to the estimators.

1. *Identification of* $r^{(0)}, f_t^{(0)}$, *and* $\lambda_{ij}^{(0)}$. We identify the global factors and global factor loadings based on model (2.1). Note that we can treat $u_{ijt}^{(0)}$ as a new error term because $c_{ijt}^{(1)} + c_{ijt}^{(2)} (= \lambda_{ij}^{(1)\prime} f_{it}^{(1)} + \lambda_{ij}^{(2)\prime} f_{jt}^{(2)})$ is uncorrelated with $c_{ijt}^{(0)} (= \lambda_{ij}^{(0)\prime} f_t^{(0)})$ and the cross-sectional dependence in $c_{ijt}^{(1)} + c_{ijt}^{(2)}$ is much weaker than that in $c_{ijt}^{(0)}$ under Assumption I(ii). If we stack the (i,j) indices to a single index as in (2.4), then we can view model (2.1) or (2.4) as a standard 2D factor model and apply the PCA method to estimate it as in Bai (2003). We can determine $r^{(0)}$ by maximizing the ratio of two adjacent eigenvalues of YY' when the eigenvalues

are ordered in descending order, as suggested by (AH). To identify $F^{(0)}$ and $\Lambda^{(0)}$, we need to impose certain normalization conditions, as discussed in Bai (2003). Following Bai (2003), we impose the normalization condition that $F^{(0)\prime}F^{(0)}/T=\mathbb{I}_{r^{(0)}}$ and $\Lambda^{(0)\prime}\Lambda^{(0)}$ is an $r^{(0)}\times r^{(0)}$ diagonal matrix with diagonal elements stacked in descending order along its main diagonal line. Then we can identify $r^{(0)}, \lambda_{ij}^{(0)\prime}f_t^{(0)}$, and a rotated version of $\lambda_{ij}^{(0)}$ and $f_t^{(0)}$.

2. *Identification of* $r_i^{(1)}$ *and*$(f_{it}^{(1)},\lambda_{ij}^{(1)})$. After identifying $\lambda_{ij}^{(0)\prime}f_t^{(0)}$, we can identify $u_{ijt}^{(0)}\equiv y_{ijt}-\lambda_{ij}^{(0)\prime}f_t^{(0)}$. Then, we can identify $r_i^{(1)}$ and $(f_{it}^{(1)},\lambda_{ij}^{(1)})$ for each i based on model (2.2) or (2.5) by treating $u_{ijt}^{(0)}$ as the response variable and noting that $u_{ijt}^{(1)}$ and $f_{it}^{(1)}$ are uncorrelated. For each $i,$ we impose the normalization condition that $F_i^{(1)\prime}F_i^{(1)}/T=\mathbb{I}_{r_i^{(1)}}$ and $\Lambda_i^{(1)\prime}\Lambda_i^{(1)}$ is an $r_i^{(1)}\times r_i^{(1)}$ diagonal matrix.
3. *Identification of* $r_j^{(2)}$ *and* $(f_{jt}^{(2)},\lambda_{ij}^{(2)})$. By exchanging the roles of the i and j indices in Step 2, we can identify $r_j^{(2)}$ and a rotated version of $(f_{jt}^{(2)},\lambda_{ij}^{(2)})$.

Remark 1. Assumption I only specifies some essential conditions for the separation between the global and local factors, which, in conjunction with Assumptions 1–3, which follows, are sufficient to identify the global factors. The singular values corresponding to global and local factors diverge at different rates, which is the key to achieve the separation between the global and two-types of local factors. Define the $T\times N\bar{M}$ matrices $U=\{u_{ijt}\}$ and $C^{(l)}=\{c_{ijt}^{(l)}\}$ for $l=0,1,2$. Then, $Y=C^{(0)}+C^{(1)}+C^{(2)}+U$. Intuitively, we can show that under the key conditions in Assumption I and Assumptions 1–3,

$$\begin{aligned}
\frac{1}{(N\bar{M}T)^{1/2}}\left\|C^{(0)}\right\|_{\mathrm{sp}} &= \Omega(1),\\
\frac{1}{(N\bar{M}T)^{1/2}}\left\|C^{(1)}\right\|_{\mathrm{sp}} &= \Omega\left(N^{-1/2}+T^{-1/2}\right),\\
\frac{1}{(N\bar{M}T)^{1/2}}\left\|C^{(2)}\right\|_{\mathrm{sp}} &= \Omega\left(M^{-1/2}+T^{-1/2}\right),\\
\frac{1}{(N\bar{M}T)^{1/2}}\left\|U\right\|_{\mathrm{sp}} &= \Omega\left(T^{-1/2}\right),
\end{aligned}$$

where $\Omega(\cdot)$ denotes the exact probability order. The smaller order of $\left\|C^{(1)}\right\|_{\mathrm{sp}}$ (similarly $\left\|C^{(2)}\right\|_{\mathrm{sp}}$) than $\left\|C^{(0)}\right\|_{\mathrm{sp}}$ is mainly due to two reasons: (1) $\mathbb{E}(f_{it}^{(1)})=0$, and (2) only weak cross-sectional correlations are allowed in $\{f_{it}^{(1)}\}_{i\in[N]}$ such that

$$\max_t \frac{1}{N}\sum_{i=1}^{N}\sum_{i_1=1}^{N}\left\|\mathbb{E}(f_{it}^{(1)}f_{i_1t}^{(1)\prime})\right\|\leq C<\infty. \tag{2.7}$$

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

See Assumption I(i) and (ii). In contrast, the global factor $f_t^{(0)}$ may or may not have zero mean, and it certainly does not satisfy a condition like (2.7).[7] This explains why we can separate the local factors from the global factors. Similarly, we can separate the effect of U from the global factors. This explains why we can rely on (2.4) to identify the global factor component $C^{(0)}$. For details, see the proof of Lemma S2.1 in the Supplementary Material. Given the identification of $C^{(0)}$, we can rely on (2.5) and (2.6) to identify the two types of local factors as in the 2D case.

Remark 2. Assumption I(i) imposes that the local factors have zero mean, which can be thought of as a normalization condition, as we argue below. Consider two cases. First, the original global factors do not contain a constant term. If either $\mathbb{E}[f_{it}^{(1)}]$ or $\mathbb{E}[f_{jt}^{(2)}]$ is nonzero but does not change over time, we can rewrite model (1.1) as

$$y_{ijt} = \lambda_{ij}^{(0*)\prime} f_t^{(0*)} + \lambda_{ij}^{(1)\prime} f_{it}^{(1*)} + \lambda_{ij}^{(2)\prime} f_{jt}^{(2*)} + u_{ijt},$$

where $\lambda_{ij}^{(0*)} = (\lambda_{ij}^{(0)\prime}, \lambda_{ij}^{(1)\prime}\mathbb{E}[f_{it}^{(1)}] + \lambda_{ij}^{(2)\prime}\mathbb{E}[f_{jt}^{(2)}])'$, $f_t^{(0*)} = (f_t^{(0)\prime}, 1)'$, $f_{it}^{(1*)} = f_{it}^{(1)} - \mathbb{E}[f_{it}^{(1)}]$, and $f_{jt}^{(2*)} = f_{jt}^{(2)} - \mathbb{E}[f_{jt}^{(2)}]$. By construction, the new local factors $f_{it}^{(1*)}$ and $f_{jt}^{(2*)}$ have zero mean, and the new global factor $f_t^{(0*)}$ is obtained by augmenting the original global factor $f_t^{(0)}$ with the constant 1. Similarly, consider the second case where the original global factors already contain a constant such that $f_t^{(0)} = [1, \mathfrak{f}_t^{(0)\prime}]'$ with the loadings $[\lambda_{ij}^{(00)}, \lambda_{ij}^{(01)\prime}]'$. Then, we can write

$$y_{ijt} = \left[\lambda_{ij}^{(00*)}, \lambda_{ij}^{(01)\prime}\right] f_t^{(0)\prime} + \lambda_{ij}^{(1)\prime} f_{it}^{(1*)} + \lambda_{ij}^{(2)\prime} f_{jt}^{(2*)} + u_{ijt},$$

where $\lambda_{ij}^{(00*)} = \lambda_{ij}^{(00)} + \lambda_{ij}^{(1)\prime}\mathbb{E}[f_{it}^{(1)}] + \lambda_{ij}^{(2)\prime}\mathbb{E}[f_{jt}^{(2)}]$. Again, the new local factors $f_{it}^{(1*)}$ and $f_{jt}^{(2*)}$ have zero mean.

Remark 3. The uncorrelation assumption in Assumption I(ii), is often assumed in multidimensional and multilevel/hierarchical factor models (see, e.g., Choi et al. (2018, Assump. 1(ii)) Han (2021, Assump. 1(a))). This assumption can also be thought of as a normalization, as it can be satisfied by linear projections and redefining factors and factor loadings. Specifically, suppose that the original factors $f_t^{(0)}$, $f_{it}^{(1)}$, and $f_{jt}^{(2)}$ do not satisfy Assumption I(ii) such that $\mathbb{E}[f_t^{(0)} f_{it}^{(1)\prime}] \neq 0$ and $\mathbb{E}[f_t^{(0)} f_{jt}^{(2)\prime}] \neq 0$. Then, by linear projections, we can find $\delta_i^{(1)}$ (an $r_i^{(1)} \times r^{(0)}$ matrix) and $\delta_j^{(2)}$ (an $r_j^{(2)} \times r^{(0)}$ matrix) such that

$$f_{it}^{(1)} = \delta_i^{(1)} f_t^{(0)} + e_{it}^{(1)} \text{ for } i \in [N], \quad f_{jt}^{(2)} = \delta_j^{(2)} f_t^{(0)} + e_{jt}^{(2)} \text{ for } j \in [M],$$

$\mathbb{E}[f_t^{(0)} e_{it}^{(1)\prime}] = 0$, and $\mathbb{E}[f_t^{(0)} e_{jt}^{(2)\prime}] = 0$. Then, model (1.1) implies that

$$y_{ijt} = (\lambda_{ij}^{(0)} + \delta_i^{(1)\prime}\lambda_{ij}^{(1)} + \delta_j^{(2)\prime}\lambda_{ij}^{(2)})' f_t^{(0)} + \lambda_{ij}^{(1)\prime} e_{it}^{(1)} + \lambda_{ij}^{(2)\prime} e_{jt}^{(2)} + u_{ijt}.$$

[7] If one wrote $f_t^{(0)}$ as $f_{it}^{(0)}$, then $f_{it}^{(0)}$'s would be fully dependent across i.

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

Then, we can redefine $f_t^{(0)}, e_{it}^{(1)}$ and $e_{jt}^{(2)}$ as new factors and their coefficients as factor loadings. By construction, the new factors satisfy Assumption I(ii). Note that $r^{(0)}, r_i^{(1)}$, and $r_j^{(2)}$ are unchanged through this normalization. So the identification of $r^{(0)}, r_i^{(1)}$, and $r_j^{(2)}$ does not rely on Assumption I(ii).

Remark 4. Assumption I(v) is not innocuous and it can be relaxed at the cost of much more lengthy arguments. In particular, our Algorithms 2.2 and 2.3 which follows can still be applied even when the two types of local factors are correlated. However, to show the asymptotic results, we need to introduce rather complicated notations and argue with projections of one type of local factors onto the other. In an earlier version of this article, we showed the detailed results. Intuitively, suppose that Assumption I(v) is not satisfied, we could consider the linear projections for $f_{jt}^{(2)}$ and $f_{it}^{(1)}$ under Assumption I(iv):

$$f_{jt}^{(2)} = \Pi_{ij}^{(1)\prime} f_{it}^{(1)} + e_{ijt}^{(1)} \text{ and } f_{it}^{(1)} = \Pi_{ij}^{(2)\prime} f_{jt}^{(2)} + e_{ijt}^{(2)}, \tag{2.8}$$

where $\Pi_{ij}^{(1)} = [\mathbb{E}(f_{it}^{(1)} f_{it}^{(1)\prime})]^{-1} \mathbb{E}(f_{it}^{(1)} f_{jt}^{(2)\prime}), \Pi_{ij}^{(2)} = [\mathbb{E}(f_{jt}^{(2)} f_{jt}^{(2)\prime})]^{-1} \mathbb{E}(f_{jt}^{(2)} f_{it}^{(1)\prime})$, and $e_{ijt}^{(1)}$ and $e_{ijt}^{(2)}$ are the least squares projection errors. By construction, $\mathbb{E}[f_{it}^{(1)} e_{ijt}^{(1)\prime}] = 0$ and $\mathbb{E}[f_{jt}^{(2)} e_{ijt}^{(2)\prime}] = 0$. Substituting (2.8) into the error terms in (2.2) and (2.3), respectively, yields

$$u_{ijt}^{(0)} = \lambda_{ij}^{(1\diamond)\prime} f_{it}^{(1)} + u_{ijt}^{(1\diamond)}, \text{ and} \tag{2.9}$$

$$u_{ijt}^{(0)} = \lambda_{ij}^{(2\diamond)\prime} f_{jt}^{(2)} + u_{ijt}^{(2\diamond)}, \tag{2.10}$$

where $\lambda_{ij}^{(1\diamond)} = \lambda_{ij}^{(1)} + \Pi_{ij}^{(1)} \lambda_{ij}^{(2)}, \lambda_{ij}^{(2\diamond)} = \lambda_{ij}^{(2)} + \Pi_{ij}^{(2)} \lambda_{ij}^{(1)}, u_{ijt}^{(1\diamond)} = \lambda_{ij}^{(2)\prime} e_{ijt}^{(1)} + u_{ijt}$, and $u_{ijt}^{(2\diamond)} = \lambda_{ij}^{(1)\prime} e_{ijt}^{(2)} + u_{ijt}$. Then by construction, $u_{ijt}^{(1\diamond)}$ and $f_{it}^{(1)}$ are uncorrelated and $u_{ijt}^{(2\diamond)}$ and $f_{jt}^{(2)}$ are uncorrelated. Comparing models (2.9) and (2.10) with models (2.2) and (2.3), we see that local factors $f_{it}^{(1)}$ and $f_{jt}^{(2)}$ are not affected. In other words, $f_{it}^{(1)}$ and $f_{jt}^{(2)}$ can still be consistently estimated in Steps 1.2 and 1.3 of Algorithm 2.3, which follows, respectively. After obtaining the consistent estimators of $f_{it}^{(1)}$ and $f_{jt}^{(2)}$, we can consistently estimate the factor loadings $\lambda_{ij}^{(1)}$ and $\lambda_{ij}^{(2)}$ in Step 1.4 of Algorithm 2.3, which follows. Again, for ease of exposition, we impose Assumption I(v) here.

2.2. Determination of the Number of Global and Local Factors

Our method to determine the number of factors is a two-step method. The first step obtains the initial consistent estimators based on the above three-step identification strategy, and the second step updates the estimators. In theory, we show that the first-step estimators and second-step estimators are both consistent. However, in simulations, we find that the second step can improve the accuracy of the estimators in terms of the probability of determining the correct number of factors.

Here, we propose to extend (AH)'s method to determine $r^{(0)}, r_i^{(1)}$, and $r_j^{(2)}$ and remark that other methods can also be applied. Pick a reasonably large integer $r_{\max}$. We first introduce (AH)'s ER and GR estimators for a generic 2D data matrix.

Algorithm 2.1. ***ER and GR estimators of the number of factors.***

Consider a generic $\mathcal{T} \times \mathcal{N}$ data matrix $\mathcal{Y}$. Let $\mu_1, ..., \mu_{r_{\max}}$ be the first $r_{\max}$ largest eigenvalues of $\mathcal{Y}\mathcal{Y}'/(\mathcal{N}\mathcal{T})$ (a $\mathcal{T} \times \mathcal{T}$ matrix) in descending order. Define the "zero-th" eigenvalue as $\mu_0 = \sum_{k=1}^{\min(\mathcal{N},\mathcal{T})} \mu_k / \ln(\min(\mathcal{N},\mathcal{T}))$. Then the ER and GR estimators of factor numbers are defined respectively as

$$r_{ER} = \arg \max_{0 \le k \le r_{\max}} \frac{\mu_k}{\mu_{k+1}} \text{ and } r_{GR} = \arg \max_{0 \le k \le r_{\max}} \frac{\ln\left(1+\mu_k^*\right)}{\ln\left(1+\mu_{k+1}^*\right)},$$

where $\mu_k^* = \frac{\mu_k}{V(k)}$ and $V(k) = \sum_{j=k+1}^{\min(\mathcal{N},\mathcal{T})} \mu_j$.

It is well known that the GR estimator performs slightly better than the ER estimator as it uses the information on all eigenvalues.

Algorithm 2.2. ***Determination of the numbers of global and local factors.***

Step 1: Obtain the initial consistent estimators of $\{r^{(0)}, r_i^{(1)}, r_j^{(1)}\}$.

1. *Apply Algorithm 2.1 to $Y = \left\{y_{ijt}\right\}$ (a $T \times N\bar{M}$ matrix) based on (2.4) to obtain $\tilde{r}^{(0)}$, the ER/GR estimator of $r^{(0)}$. The PCA estimator $\tilde{F}^{(0)} = (\tilde{f}_1^{(0)}, ..., \tilde{f}_T^{(0)})'$ of $F^{(0)}$ is obtained as $\sqrt{T}$ times the normalized eigenvectors corresponding to the $\tilde{r}^{(0)}$ largest eigenvalues of YY'. Let $\tilde{\Lambda}^{(0)} = Y'\tilde{F}^{(0)}/T = \{\tilde{\lambda}_{ij}^{(0)}\}_{i\in[N],\, j\in\mathcal{M}_i}$. Let $\tilde{u}_{ijt}^{(0)} \equiv y_{ijt} - \tilde{\lambda}_{ij}^{(0)\prime}\tilde{f}_t^{(0)}, \tilde{u}_{ij}^{(0)} = (\tilde{u}_{ij1}^{(0)}, ..., \tilde{u}_{ijT}^{(0)})'$, $\tilde{U}_i^{(0)} = \{\tilde{u}_{ij}^{(0)}\}_{j\in\mathcal{M}_i}$, and $\tilde{U}_{\cdot j}^{(0)} = \{\tilde{u}_{ij}^{(0)}\}_{i\in\mathcal{N}_j}$.*
2. *For each i, apply Algorithm 2.1 and PCA to $\tilde{U}_i^{(0)}$ based on (2.5). Let $\tilde{r}_i^{(1)}$ and $(\tilde{F}_i^{(1)\prime}, \tilde{\Lambda}_i^{(1)\prime})$ be the ER/GR estimator of $r_i^{(1)}$ and PCA estimators of $(F_i^{(1)\prime}, \Lambda_i^{(1)\prime})$, respectively, where $\tilde{F}_i^{(1)} = (\tilde{f}_{i1}^{(1)}, ..., \tilde{f}_{iT}^{(1)})'$ and $\tilde{\Lambda}_i^{(1)} = (\tilde{\lambda}_{i1}^{(1)}, ..., \tilde{\lambda}_{iM_i}^{(1)})'$.*
3. *For each j, apply Algorithm 2.1 and PCA to $\tilde{U}_{\cdot j}^{(0)}$ based on (2.6). Let $\tilde{r}_j^{(2)}$ and $(\tilde{F}_j^{(2)\prime}, \tilde{\Lambda}_{\cdot j}^{(2)\prime})$ be the ER/GR estimator of $r_j^{(2)}$ and PCA estimators of $(F_j^{(2)\prime}, \Lambda_{\cdot j}^{(2)\prime})$, respectively, where $\tilde{F}_j^{(2)} = (\tilde{f}_{j1}^{(2)}, ..., \tilde{f}_{jT}^{(2)})'$ and $\Lambda_{\cdot j}^{(2)} = (\tilde{\lambda}_{1j}^{(2)}, ..., \tilde{\lambda}_{N_j j}^{(2)})'$. Let $\tilde{y}_{ijt}^{(0)} \equiv y_{ijt} - \tilde{\lambda}_{ij}^{(1)\prime}\tilde{f}_{it}^{(1)} - \tilde{\lambda}_{ij}^{(2)\prime}\tilde{f}_{jt}^{(2)}$.*

Step 2 Obtain the updated estimators of $\{r^{(0)}, r_i^{(1)}, r_j^{(1)}\}$.

1. *Apply the Algorithm 2.1 and PCA to $\tilde{Y}^{(0)} = \{\tilde{y}_{ijt}^{(0)}\}$ (a $T \times N\bar{M}$ matrix) to obtain the updated estimator $(\hat{r}^{(0)\prime}, \hat{F}^{(0)\prime}, \hat{\Lambda}^{(0)\prime})$ of $(r^{(0)}, F^{(0)\prime}, \Lambda^{(0)\prime})$, where $\hat{F}^{(0)} = \{\hat{f}_t^{(0)}\}$ and $\hat{\Lambda}^{(0)} = \{\hat{\lambda}_{ij}^{(0)}\}$. Let $\tilde{y}_{ijt}^{(1)} \equiv y_{ijt} - \hat{\lambda}_{ij}^{(0)\prime}\hat{f}_t^{(0)} - \tilde{\lambda}_{ij}^{(2)\prime}\tilde{f}_{jt}^{(2)}$ and $\tilde{y}_{ijt}^{(2)} \equiv y_{ijt} - \hat{\lambda}_{ij}^{(0)\prime}\hat{f}_t^{(0)} - \tilde{\lambda}_{ij}^{(1)\prime}\tilde{f}_{it}^{(1)}$.*
2. *For each i, apply Algorithm 2.1 to $\tilde{Y}_i^{(1)} = \{\tilde{y}_{ijt}^{(1)}\}$ (a $T \times M_i$ matrix) to obtain the updated estimator $\hat{r}_i^{(1)}$ of $r_i^{(1)}$.*

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

3. *For each j, apply Algorithm 2.1 to $\tilde{Y}_j^{(2)} = \{\tilde{y}_{ijt}^{(2)}\}$ (a $T \times N_j$ matrix) to obtain the updated estimator $\hat{r}_j^{(2)}$ of $r_j^{(2)}$.*

Our final estimators of $(r^{(0)}, r_i^{(1)}, r_j^{(2)})$ are $(\hat{r}^{(0)}, \hat{r}_i^{(1)}, \hat{r}_j^{(2)})$, respectively. Note that in Algorithm 2.2, there is no tuning parameter involved and users do not need to choose anything other than $r_{\max}$. The requirement on $r_{\max}$ is that $r_{\max} \geq \max(r^{(0)}, \max_i r_i^{(1)}, \max_j r_j^{(2)})$ and it be fixed. For the practical choice of $r_{\max}$, we want to avoid a too small number and recommend $r_{\max} = 8$ following (AH) and Bai and Ng (2002).

2.3. Estimation of the Global and Local Factors and Factor Loadings

After determining the number of factors, we apply Algorithm 2.3, which follows to estimate all factors and factor loadings. Here, we propose a two-step approach to estimate the factors and factor loadings for any given $(r^{(0)}, r_i^{(1)}, r_j^{(2)})$ and remark that the factor numbers can be replaced by their consistent estimates, say, $(\hat{r}^{(0)}, \hat{r}_i^{(1)}, \hat{r}_j^{(2)})$ obtained in Algorithm 2.2.

Algorithm 2.3. ***Estimation of factors and factor loadings for given*** $(r^{(0)}, r_i^{(1)}, r_j^{(2)})$**.**

Step 1: Obtain the initial consistent estimators of the factors and factor loadings.

1. *Apply the PCA to $Y = \{y_{ijt}\}$ to obtain the initial estimators $\tilde{F}^{(0)} = \{\tilde{f}_t^{(0)}\}$ and $\tilde{\Lambda}^{(0)} = \{\tilde{\lambda}_{ij}^{(0)}\}$ of $F^{(0)}$ and $\Lambda^{(0)}$ as in Step 1(1) in Algorithm 2.2. Define $\tilde{u}_{ijt}^{(0)}$, $\tilde{U}_i^{(0)}$, and $\tilde{U}_{\cdot j}^{(0)}$ as in Algorithm 2.2.*
2. *For each i, apply PCA to $\tilde{U}_i^{(0)}$ to obtain the PCA estimator $\tilde{F}_i^{(1)} = \{\tilde{f}_{it}^{(1)}\}$ of $F_i^{(1)}$, and run the OLS regression of $\tilde{F}_i^{(1)}$ on $\tilde{F}^{(0)}$ to obtain the residual $\vec{F}_i^{(1)} = \{\vec{f}_{it}^{(1)}\}$.*
3. *For each j, apply PCA to $\tilde{U}_{\cdot j}^{(0)}$ to obtain the PCA estimator $\tilde{F}_j^{(2)} = \{\tilde{f}_{jt}^{(2)}\}$ of $F_j^{(2)}$, and run the OLS regression of $\tilde{F}_j^{(2)}$ on $\tilde{F}^{(0)}$ to obtain the residual $\vec{F}_j^{(2)} = \{\vec{f}_{jt}^{(1)}\}$.*[8]
4. *For each (i,j) pair, run the OLS regression of y_{ij} on $\vec{F}_{ij} = (\tilde{F}^{(0)}, \vec{F}_i^{(1)}, \vec{F}_j^{(2)})$ to obtain the estimator $\vec{\lambda}_{ij} = (\vec{\lambda}_{ij}^{(0)\prime}, \vec{\lambda}_{ij}^{(1)\prime}, \vec{\lambda}_{ij}^{(2)\prime})'$ of $\lambda_{ij} = (\lambda_{ij}^{(0)\prime}, \lambda_{ij}^{(1)\prime}, \lambda_{ij}^{(2)\prime})'$.*

Step 2: Obtain the more efficient estimators of the factors and factor loadings

1. *For each t, run the OLS regression of $Y_{\cdot\cdot t} = \{y_{ijt}\}$ (an $N\bar{M} \times 1$ vector) on $\{\vec{\lambda}_{ij}\}$ to obtain the updated estimator $\breve{f}_t = (\breve{f}_t^{(0)\prime}, \breve{F}_{\cdot t}^{(1)\prime}, \breve{F}_{\cdot t}^{(2)\prime})'$ of $f_t = (f_t^{(0)\prime}, F_{\cdot t}^{(1)\prime}, F_{\cdot t}^{(2)\prime})'$, where $F_{\cdot t}^{(1)} = (f_{1t}^{(1)\prime}, \ldots, f_{Nt}^{(1)\prime})'$, $F_{\cdot t}^{(2)} = (f_{1t}^{(2)\prime}, \ldots, f_{Mt}^{(2)\prime})'$, $\breve{F}_{\cdot t}^{(1)} = (\breve{f}_{1t}^{(1)\prime}, \ldots, \breve{f}_{Nt}^{(1)\prime})'$ and $\breve{F}_{\cdot t}^{(2)} = (\breve{f}_{1t}^{(2)\prime}, \ldots, \breve{f}_{Mt}^{(2)\prime})'$. Let $\breve{F}^{(0)} = \{\breve{f}_t^{(0)}\}$, $\breve{F}_i^{(1)} = \{\breve{f}_{it}^{(1)}\}$, and $\breve{F}_j^{(2)} = \{\breve{f}_{jt}^{(2)}\}$.*

---

[8]The OLS regressions in Step 1(2) (resp. Step 1(3)) ensure that $\vec{F}_i^{(1)}$ (resp. $\vec{F}_j^{(2)}$) and $\tilde{F}^{(0)}$ are orthogonal in finite samples.

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

2. *For each (i,j) pair, run the OLS regression of y_{ij} on $\breve{F}_{ij} = (\breve{F}^{(0)}, \breve{F}_i^{(1)}, \breve{F}_j^{(2)})$ to obtain the estimator $\breve{\lambda}_{ij} = (\breve{\lambda}_{ij}^{(0)\prime}, \breve{\lambda}_{ij}^{(1)\prime}, \breve{\lambda}_{ij}^{(2)\prime})'$ of λ_{ij}.*

Our final estimators of $(f_t^{(0)\prime}, f_{it}^{(1)\prime}, f_{jt}^{(2)\prime})'$ and $(\lambda_{ij}^{(0)\prime}, \lambda_{ij}^{(1)\prime}, \lambda_{ij}^{(2)\prime})'$ are $(\breve{f}_t^{(0)\prime}, \breve{f}_{it}^{(1)\prime}, \breve{f}_{jt}^{(2)\prime})'$ and $(\breve{\lambda}_{ij}^{(0)\prime}, \breve{\lambda}_{ij}^{(1)\prime}, \breve{\lambda}_{ij}^{(2)\prime})'$, respectively. Note that there is no tuning parameter involved in Algorithm 2.3. To improve the finite sample performance, we can iterate Steps 2.1 and 2.2 in Algorithm 2.3, by updating the estimators of factors and loadings.

Remark 5. The first step in Algorithm 2.3, basically follows the identification strategy. We can show that the first-step estimators are consistent. The purpose of the second step is to improve the estimation efficiency. To see this, for example, the first-step estimation of the global factor is based on

$$y_{ijt} = \lambda_{ij}^{(0)\prime} f_t^{(0)} + u_{ijt}^{(0)}, \text{ where } u_{ijt}^{(0)} = \lambda_{ij}^{(1)\prime} f_{it}^{(1)} + \lambda_{ij}^{(2)\prime} f_{jt}^{(2)} + u_{ijt}, \tag{2.11}$$

while the second-step estimation is based on

$$y_{ijt} = \lambda_{ij}^{(0)\prime} f_t^{(0)} + \lambda_{ij}^{(1)\prime} f_{it}^{(1)} + \lambda_{ij}^{(2)\prime} f_{jt}^{(2)} + u_{ijt}, \tag{2.12}$$

where $(\lambda_{ij}^{(0)\prime}, \lambda_{ij}^{(1)\prime}, \lambda_{ij}^{(2)\prime})'$ are estimated. Since u_{ijt} has a smaller variance than $u_{ijt}^{(0)}$ under Assumption I(iii), the estimator of the global factor based on (2.12) is asymptotically more efficient than that based on (2.11). The local factor components in $u_{ijt}^{(0)}$ create strong cross-sectional dependence and thus, slow down the convergence of the estimator of $f_t^{(0)}$ in general. In fact, our theory suggests that the first-step estimator of the global factor has a slower convergence rate than the second-step estimator except for some special cases. Even in the special case where the first- and second-step estimators have the same convergence rate, the second-step estimator usually has a smaller variance. In our simulations, we find that the second-step estimator of the global factor has a much smaller MSE than the first-step one.

Remark 6. The specification of our model (1.1) is quite general and includes many interesting special cases. Model (1.3) discussed above is one example. Another example is

$$y_{ijt} = \lambda_{ij}^{(0)\prime} f_t^{(0)} + \lambda_j^{(1)\prime} f_{it}^{(1)} + \lambda_i^{(2)\prime} f_{jt}^{(2)} + u_{ijt},$$

where the local factor loadings only depend on one cross-sectional dimension. Another example is that we can further impose a factor structure on the factor loadings. For notational simplicity, we assume that $r^{(0)} = r_i^{(1)} = r_j^{(2)} = 1$. Then, the model can be written as

$$y_{ijt} = (\lambda_{i\cdot}^{(0)\prime} \lambda_{\cdot j}^{(0)}) f_t^{(0)} + (\lambda_{i\cdot}^{(1)\prime} \lambda_{\cdot j}^{(1)}) f_{it}^{(1)} + (\lambda_{i\cdot}^{(2)\prime} \lambda_{\cdot j}^{(2)}) f_{jt}^{(2)} + u_{ijt}, \tag{2.13}$$

where $\lambda_{i\cdot}^{(0)}, \lambda_{\cdot j}^{(0)}, \lambda_{i\cdot}^{(1)}, \lambda_{\cdot j}^{(1)}, \lambda_{i\cdot}^{(2)}$, and $\lambda_{\cdot j}^{(2)}$ are the factor loadings. Our estimation procedure can be easily modified to incorporate the restrictions in these special

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

cases. For example, to estimate model (2.13), we can first apply our method and then apply PCA to the estimated factor loadings.

Remark 7. One limitation of our model, and factor models in general, is that we can only identify the factor and factor loadings up to a rotation matrix. In this article, we impose the standard normalization ($F'F/T$ is an identity matrix and $\Lambda'\Lambda$ is a diagonal matrix where F and Λ denote the factor matrix and the factor loading matrix, respectively), as we discuss the identification part in Section 2.1. Consequently, we can only identify a specific linear combination of the factors and their loadings. Even with normalization, the sign of both the factors and loadings remains indeterminate. To resolve this ambiguity, we rely on empirical context and subjective judgment. In our application, for example, we determine the sign of the estimated global factor to align with historical global recessions, as illustrated in Figure 2. Specifically, we assign the sign such that the global factor exhibits lower values during recession periods. This approach is economically intuitive given that our dependent variable measures export growth. As Bai and Ng (2013) discuss, practitioners may alternatively consider some other normalization conditions, though these require careful implementation.

3. ASYMPTOTIC THEORY: DETERMINATION OF $r^{(0)}, r_i^{(1)}$, AND $r_j^{(2)}$

In this section, we study the asymptotic properties of the estimators of $r^{(0)}, r_i^{(1)}$ and $r_j^{(2)}$. We first study the consistency of the first-step estimators $\tilde{r}^{(0)}$, $\tilde{r}_i^{(1)}$, and $\tilde{r}_j^{(2)}$ in Algorithm 2.2, and then study that of the second-step estimators $\hat{r}^{(0)}$, $\hat{r}_i^{(1)}$, and $\hat{r}_j^{(2)}$.

3.1. Consistency of $\tilde{r}^{(0)}, \tilde{r}_i^{(1)}$ and $\tilde{r}_j^{(2)}$

To proceed, we define some notations. Let $\underline{m} = \min(N,M,T)$, $\bar{m} = \max(N,M,T)$, $R^{(1)} = \sum_{i=1}^{N} r_i^{(1)}$, and $R^{(2)} = \sum_{j=1}^{M} r_j^{(2)}$. Let $\boldsymbol{F}^{(1)} = (F_1^{(1)}, ..., F_N^{(1)})$ and $\boldsymbol{F}^{(2)} = (F_1^{(2)}, ..., F_M^{(2)})$, which are $T \times R^{(1)}$ and $T \times R^{(2)}$ matrices, respectively. Let $U_{ij} = \left(u_{ij1}, ..., u_{ijT}\right)'$ and $U_i = \{U_{ij}\}_{j \in \mathcal{M}_i}$ for $i \in [N]$. Note that $U = (U_1, ..., U_N)$. Analogously, we can define $U^{\dagger} = (U_{\cdot 1}, ..., U_{\cdot M})$, where $U_{\cdot j} = \{U_{ij}\}_{i \in \mathcal{N}_j}$ for $j \in [M]$. Note that $U = U^{\dagger} S$, where S is an $N\bar{M} \times N\bar{M}$ permutation matrix that permutes the columns of the $T \times N\bar{M}$ matrix $U^{\dagger}$ to obtain the $T \times N\bar{M}$ matrix U. It is well known that S can be obtained by permuting the columns of the identity matrix $\mathbb{I}_{N\bar{M}}$ and all permutation matrices are orthogonal matrices. As a result, we have $SS' = \mathbb{I}_{N\bar{M}}$ and $U^{\dagger} = US'$. Define

$$\boldsymbol{\Lambda}^{(1)} = \text{bdiag}(\Lambda_1^{(1)}, ..., \Lambda_N^{(1)}) \text{ and } \boldsymbol{\Lambda}^{(2)} = \text{bdiag}(\Lambda_{\cdot 1}^{(2)}, ..., \Lambda_{\cdot M}^{(2)}).$$

Note that $\boldsymbol{\Lambda}^{(1)}$ and $\boldsymbol{\Lambda}^{(2)}$ are $N\bar{M} \times R^{(1)}$ and $N\bar{M} \times R^{(2)}$ matrices, respectively. Then, we have

$$U^{(0)} = \boldsymbol{F}^{(1)} \boldsymbol{\Lambda}^{(1)\prime} + \boldsymbol{F}^{(2)} \boldsymbol{\Lambda}^{(2)\prime} S + U = \boldsymbol{F}^{(1,2)} \boldsymbol{\Lambda}^{(1,2)\prime} + U,$$

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

where $\boldsymbol{F}^{(1,2)} = (\boldsymbol{F}^{(1)}, \boldsymbol{F}^{(2)})$ and $\boldsymbol{\Lambda}^{(1,2)} = (\boldsymbol{\Lambda}^{(1)}, S'\boldsymbol{\Lambda}^{(2)})$. Let $\tilde{\Sigma}_{\Lambda^{(0)}} \equiv \frac{1}{N\bar{M}}\Lambda^{(0)\prime}\Lambda^{(0)}$. Define

$$\bar{\mu}_l^{(0)} = \frac{1}{N\bar{M}T}\psi_l(\Lambda^{(0)\prime}\Lambda^{(0)}F^{(0)\prime}F^{(0)}) \text{ for } l \in [r^{(0)}],$$

$$\bar{c}_{\boldsymbol{\Lambda}^{(1)}} = \max_{i\in[N]} \frac{1}{\bar{M}}\psi_1(\Lambda_i^{(1)\prime}\Lambda_i^{(1)}),\ \underline{c}_{\boldsymbol{\Lambda}^{(1)}} = \min_{i\in[N],\ r_i^{(1)}>0} \frac{1}{\bar{M}}\psi_{\min}(\Lambda_i^{(1)\prime}\Lambda_i^{(1)}),$$

$$\bar{c}_{\boldsymbol{F}^{(1)}} = \frac{1}{T\vee N}\psi_1(\boldsymbol{F}^{(1)}\boldsymbol{F}^{(1)\prime}),$$

$$\bar{c}_{\boldsymbol{\Lambda}^{(2)}} = \max_{j\in[M]} \frac{1}{\bar{N}}\psi_1(\Lambda_{.j}^{(2)\prime}\Lambda_{.j}^{(2)}),\ \underline{c}_{\boldsymbol{\Lambda}^{(2)}} = \min_{j\in[M],\ r_j^{(2)}>0} \frac{1}{\bar{N}}\psi_{\min}(\Lambda_{.j}^{(2)\prime}\Lambda_{.j}^{(2)}),$$

$$\bar{c}_{\boldsymbol{F}^{(2)}} = \frac{1}{T\vee M}\psi_1(\boldsymbol{F}^{(2)}\boldsymbol{F}^{(2)\prime}).$$

Let c and C be generic positive constants that may change over places. We make the following assumptions.

Assumption 1.

(i) For each $(i,j), N_j/N \to \tau_{1j}$ and $M_i/M \to \tau_{2i}$ as $\underline{m} \to \infty$, where τ_{1j} and τ_{2i} are bounded away from zero uniformly in j and i.
(ii) $\bar{m}(\ln T)^4/\underline{m}^2 \to 0$ as $\underline{m} \to \infty$.
(iii) $0 \le r^{(0)}, r_i^{(1)}, r_j^{(2)} \le r_{\max}$, where $r_{\max}$ is a fixed integer.

Assumption 2.

(i) $\text{plim}_{\underline{m}\to\infty}\bar{\mu}_l^{(0)} = \mu_l^{(0)} \in (0,\infty)$ for $l \in [r^{(0)}]$.
(ii) $\text{plim}_{\underline{m}\to\infty}\bar{c}_{\boldsymbol{F}^{(\ell)}} = c_{\boldsymbol{F}^{(\ell)}} < \infty$ for $\ell = 1,2$.
(iii) $\max_{0\le\ell\le2}\max_{i,j}||\lambda_{ij}^{(\ell)}|| \le \bar{c}_\lambda < \infty, \tilde{\Sigma}_{\Lambda^{(0)}} \to \Sigma_{\Lambda^{(0)}} > 0$, and $\lim_{\underline{m}\to\infty}(\underline{c}_{\boldsymbol{\Lambda}^{(1)}} \wedge \underline{c}_{\boldsymbol{\Lambda}^{(2)}}) \ge c > 0$.
(iv) $\max_t\{\mathbb{E}||f_t^{(0)}||^4 + \frac{1}{NM}\sum_{i,i_1,j,j_1}\left\|\mathbb{E}\left(u_{ijt}u_{i_1j_1t}\right)\right\|\} \le C$.
(v) Let $\bar{c}_U = \frac{1}{N\bar{M}}\psi_1\left(UU'\right)$ and $\bar{c}_U^0 = \frac{1}{T}\psi_1\left(\mathbb{E}\left[U'U\right]\right).\bar{c}_U^0 \le C$ and $\bar{c}_U$ is stochastically bounded.

Assumption 3.

(i) $\frac{1}{NM^2T}\sum_{i_1,i_2,j_1,j_2,j_3,j_4,t,s}\left|\mathbb{E}\left(\xi_{i_1j_1j_2t}\xi_{i_2j_3j_4s}\right)\right| \le C$ for $\xi_{i_1j_1j_2t} = u_{i_1j_1t}u_{i_1j_2t} - \mathbb{E}(u_{i_1j_1t}u_{i_1j_2t})$.
(ii) There exists an integer $L_0 \ge r_{\max} + 2r^{(0)}$ such that either $\frac{1}{N\bar{M}T}\psi_{L_0}(\boldsymbol{F}^{(1,2)}\boldsymbol{\Lambda}^{(1,2)\prime}\boldsymbol{\Lambda}^{(1,2)}\boldsymbol{F}^{(1,2)\prime}) \ge \frac{1}{\bar{m}}\underline{c}_{L_0}^{(0)}$ or $\frac{1}{N\bar{M}T}\psi_{L_0}\left(UU'\right) \ge \frac{1}{T}\underline{c}_{L_0}^{(0)}$, where $\underline{c}_{L_0}^{(0)}$ is bounded away from zero in probability.

Assumption 1 imposes some general conditions on $M_i, N_j, M, N, T, r^{(0)}, r_i^{(1)}$, and $r_j^{(2)}$. Note that we assume that M_i's diverge to infinity at the same rate as M, N_j's diverge to infinity at the same rate as N, $r^{(0)}, r_i^{(1)}$, and $r_j^{(2)}$ are uniformly bounded above by a finite integer $r_{\max}$. These conditions can be relaxed at the cost of more

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

complicated notations and lengthy arguments. We do not restrict $N, M,$ and T to diverge to infinity at the same rate but do require that one should not diverge as fast as the square of the other. This condition appears reasonable for most macro, finance, and trade applications.

Assumption 2 imposes conditions on the factors, factor loadings, and error terms. Assumption 2(i) is a strong factor assumption which is commonly adopted in the literature despite the fact that it rules out weak factors considered by Onatski (2012) and Bai and Ng (2023). Assumption 2(ii) implies that $||\boldsymbol{F}^{(1)}||_{\mathrm{sp}} = O_p\left(N^{1/2}+T^{1/2}\right)$ and $||\boldsymbol{F}^{(2)}||_{\mathrm{sp}} = O_p\left(M^{1/2}+T^{1/2}\right)$, which can be verified under some primitive conditions on the zero-mean local factors. Assumption 2 (iii) assumes that the factor loadings are nonrandom and uniformly bounded. It is possible to relax this assumption to allow the factor loadings to be random with uniform finite fourth moments. Assumption 2(iv) allows error terms to be weakly cross-sectionally and serially dependent. Assumption 2(v) imposes high level conditions on the spectral norm of UU' and $\mathbb{E}\left(U'U\right)$. Noting that U is a $T \times N\bar{M}$ matrix, it is standard to assume that

$$\|U\|_{\mathrm{sp}} = O_p(T^{1/2} + (N\bar{M})^{1/2}),$$

which is equivalent to the requirement $\frac{1}{N\bar{M}}\psi_1\left(UU'\right) = O_p(1)$ under Assumption 1 (ii). (See, e.g., Latała, 2005; Moon and Weidner, 2017; Su and Ju, 2018. In fact, if we follow Bai and Saranadasa (1996), Chen and Qin (2010), Vershynin (2011), and Ma et al. (2020) and assume that $U = AE,$ where A is a $T \times n$ nonrandom matrix such that $\|A\|_{\mathrm{sp}} \leq C < \infty$ and E is an $n \times N\bar{M}$ random matrix whose entries are independent random variables with mean zero and $(4+\epsilon)$th moment for some $\epsilon > 0$, then $\mathbb{E}\,\|U\|_{\mathrm{sp}} = O\left(T^{1/2} + (N\bar{M})^{1/2}\right)$ (see, e.g., (Vershynin, 2011, Thm. 1.2)). Alternatively, we can follow the literature (e.g., (AH)) and assume that $U = A_T^{1/2} V B_{N\bar{M}}^{1/2},$ where $A_T^{1/2}$ and $B_{N\bar{M}}^{1/2}$ are the symmetric square roots of the $T \times T$ and $N\bar{M} \times N\bar{M}$ p.s.d. deterministic matrices A_T and $B_{N\bar{M}}$, respectively and V is a $T \times N\bar{M}$ matrix with elements v_{ijt} being i.i.d. random variables with zero mean, unit variance, and finite fourth moment, $\limsup_{T\to\infty}\psi_1(A_T) \leq \bar{c}_A < \infty$, and $\limsup_{\underline{m}\to\infty}\psi_1\left(B_{N\bar{M}}\right) \leq \bar{c}_B < \infty$. Noting that V' is a "tall" random matrix (i.e., $N\bar{M} \gg T$) under Assumption 1(ii), Vershynin (2012, Thm. 5.31) implies that $\psi_1(\frac{1}{N\bar{M}}VV') \overset{\text{a.s.}}{\to} 1$ and $\psi_T(\frac{1}{N\bar{M}}V'V) \overset{\text{a.s.}}{\to} 1$. Then, we have

$$\begin{aligned}
\bar{c}_U &= \frac{1}{N\bar{M}}\psi_1\left(UU'\right) \\
&= \frac{1}{N\bar{M}}\psi_1\left(A_T^{1/2} V B_{N\bar{M}} V' A_T^{1/2}\right) \leq \psi_1\left(B_{N\bar{M}}\right)\frac{1}{N\bar{M}}\psi_1\left(A_T^{1/2} VV' A_T^{1/2}\right) \\
&\leq \frac{\psi_1\left(B_{N\bar{M}}\right)\psi_1(A_T)}{N\bar{M}}\psi_1\left(VV'\right) \leq \bar{c}_A\bar{c}_B\{1+o_{\text{a.s.}}(1)\}.
\end{aligned}$$

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

On the other hand, using the fact that $\frac{1}{T}\mathbb{E}\left[V'V\right]=\mathbb{I}_{N\bar{M}}$, we have

$$\begin{aligned}\bar{c}_U^0 =&\frac{1}{T}\psi_1\left(\mathbb{E}\left(U'U\right)\right)=\frac{1}{T}\psi_1\left(B_{N\bar{M}}^{1/2}\mathbb{E}\left[V'A_TV\right]B_{N\bar{M}}^{1/2}\right)\\ \leq&\psi_1(A_T)\frac{1}{T}\psi_1\left(B_{N\bar{M}}^{1/2}\mathbb{E}\left[V'V\right]B_{N\bar{M}}^{1/2}\right)\leq\psi_1(A_T)\psi_1\left(B_{N\bar{M}}\right)\leq\bar{c}_A\bar{c}_B.\end{aligned}$$

Thus, Assumption 2(v) is satisfied.

Assumption 3(i) imposes further conditions on the error term $\left\{u_{ijt}\right\}$ requiring that the dependence along either the two cross-sectional dimensions or the time dimension should not be too strong. In particular, it would be satisfied if u_{ijt}'s are independent along the two cross-sectional dimensions and weakly dependent along the time dimension (say, satisfying certain strong mixing and moment conditions). Assumption 3(ii) imposes further conditions on $\boldsymbol{F}^{(1,2)}\boldsymbol{\Lambda}^{(1,2)\prime}$ and U. In Section S7.1 of the Supplementary Material, we give some primitive conditions such that either case in Assumption 3(ii) can be satisfied. As shown in the proof of Lemma S2.1(ii) in the Supplementary Material, Assumption 3 ensures that $\frac{1}{N\bar{M}T}\psi_{L_0}\left(U^{(0)\prime}U^{(0)}\right)\geq\frac{1}{\underline{m}}\bar{c}_{L_0}$. Alternatively, we can assume such a high level condition directly and emphasize that it may be satisfied under some primitive conditions, including those specified in Assumption 3.

The following theorem states the first main result in the article.

THEOREM 3.1. *Suppose that Assumption I holds. Suppose Assumptions 1–3 hold. Let $\tilde{r}^{(0)}$ be the first-step ER or GR estimator of $r^{(0)}$. Let $\underline{m}=\min(N,M,T)$. Then, $P\left(\tilde{r}^{(0)}=r^{(0)}\right)\to 1$ as $\underline{m}\to\infty$.*

Theorem 3.1 states that the first-step ER or GR estimator $\tilde{r}^{(0)}$ of $r^{(0)}$ is consistent. In the proof, we allow $r^{(0)}\geq 1$ as well as $r^{(0)}=0$, but have to treat these two cases separately. Given the consistent estimate of $r^{(0)}$, it is possible to estimate the number of the local factors consistently.

To study the consistency of the preliminary estimators $\tilde{r}_i^{(1)}$ and $\tilde{r}_j^{(2)}$ of the number of local factors, we introduce the following notations. Let $C_j^{(1)}$ and $C_i^{(2)}$ be the $T\times N_j$ and $T\times M_i$ matrices with typical elements $c_{ijt}^{(1)}$ and $c_{ijt}^{(2)}$, respectively. Let $U_i^{(1)}=U_i+C_i^{(2)}, U_{\cdot j}^{(2)}=U_{\cdot j}+C_j^{(1)}, \mathbb{F}_j^{(1)}=\{F_i^{(1)}\}_{i\in\mathcal{N}_j}$, and $\mathbb{F}_i^{(2)}=\{F_j^{(2)}\}_{j\in\mathcal{M}_i}$. Define

$$\begin{aligned}\gamma_{st}^{(0)}=&\frac{1}{N\bar{M}}\sum_{i,j}\mathbb{E}(u_{ijt}^{(0)}u_{ijs}^{(0)}),\ \zeta_{st}^{(0)}=\frac{1}{N\bar{M}}\sum_{i,j}u_{ijt}^{(0)}u_{ijs}^{(0)}-\gamma_{st}^{(0)},\\ c_{i,u}=&\frac{1}{M_i\vee T}\psi_1\left(U_i'U_i\right),\ c_{\cdot j,u}=\frac{1}{N_j\vee T}\psi_1(U_{\cdot j}'U_{\cdot j}),\\ c_{\mathbb{F}_j^{(1)}}=&\frac{1}{N_j\vee T}\psi_1(\mathbb{F}_j^{(1)}\mathbb{F}_j^{(1)\prime}),\ c_{\mathbb{F}_i^{(2)}}=\frac{1}{M_i\vee T}\psi_1(\mathbb{F}_i^{(2)}\mathbb{F}_i^{(2)\prime}),\\ c_i^{(0)}=&\frac{1}{M_iT}\sum_t\sum_{j\in\mathcal{M}_i}\mathbb{E}(u_{ijt}^{(0)})^2,\text{ and } c_{\cdot j}^{(0)}=\frac{1}{N_jT}\sum_t\sum_{i\in\mathcal{N}_j}\mathbb{E}(u_{ijt}^{(0)})^2.\end{aligned}$$

Let $\underline{c}$and $\bar{c}$ (with $\underline{c} \leq \bar{c}$) be generic positive constants that do not depend on (N,M,T). The following three assumptions are needed to study the asymptotic properties of the first-step estimators of the global and local factors and their associated loadings and the consistency of $\tilde{r}_i^{(1)}$ and $\tilde{r}_j^{(2)}$ in Theorem 3.2, which follows.

Assumption 4.

(i) The eigenvalues of $\Sigma_{\Lambda^{(0)}}^{1/2}\Sigma_{F^{(0)}}\Sigma_{\Lambda^{(0)}}^{1/2}$ are distinct.

(ii) $\max_t \sum_s |\gamma_{st}^{(0)}| \leq C$.

(iii) $\max_{s,t} \mathbb{E}|\zeta_{st}^{(0)}|^4 \leq C\left(N^{-2}+M^{-2}\right)$.

(iv) $\max_t \mathbb{E}||\frac{1}{N\bar{M}\sqrt{T}}\sum_{i,j,s} f_s^{(0)}[u_{ijs}^{(0)}u_{ijt}^{(0)} - \mathbb{E}(u_{ijs}^{(0)}u_{ijt}^{(0)})]||^2 \leq C\left(N^{-1}+M^{-1}\right)$.

(v) $\mathbb{E}||\frac{1}{N\bar{M}\sqrt{T}}\sum_{i,j,t}\lambda_{ij}^{(0)}u_{ijt}^{(0)}f_t^{(0)\prime}||^2 \leq C(N^{-1}+M^{-1})$.

Assumption 5. For any fixed $\epsilon > 0$, we have the following.

(i) $P(||\frac{1}{T}F^{(0)\prime}F^{(0)} - \Sigma_{F^{(0)}}|| \geq \epsilon T^{-1/2}\ln T) = o(\bar{m}^{-1})$, $P(\max_i ||\frac{1}{T}F_i^{(1)\prime}F_i^{(1)} - \Sigma_{F_i^{(1)}}|| \geq \epsilon T^{-1/2}\ln T) = o(\bar{m}^{-1})$, $P(\max_j ||\frac{1}{T}F_j^{(2)\prime}F_j^{(2)} - \Sigma_{F_j^{(2)}}|| \geq \epsilon T^{-1/2}\ln T) = o(\bar{m}^{-1})$, and $\max_i ||\frac{1}{M_i}\Lambda_i^{(1)\prime}\Lambda_i^{(1)} - \Sigma_{\Lambda_i^{(1)}}|| + \max_j ||\frac{1}{N_j}\Lambda_{\cdot j}^{(2)\prime}\Lambda_{\cdot j}^{(2)} - \Sigma_{\Lambda_{\cdot j}^{(2)}}|| = o(1)$, where the eigenvalues of $\Sigma_{F_i^{(1)}}, \Sigma_{F_j^{(2)}}, \Sigma_{\Lambda_i^{(1)}}$, and $\Sigma_{\Lambda_{\cdot j}^{(2)}}$ are all bounded below and above, respectively, by $\underline{c}$ and $\bar{c}$ for all i with $r_i^{(1)} \geq 1$ and all j with $r_j^{(2)} \geq 1$.

(ii) $P(\max_{s,t} ||\zeta_{st}^{(0)}|| \geq \epsilon(N^{-1/2} + M^{-1/2})\ln T) = o(\bar{m}^{-1})$ and $P(\max_t ||\frac{1}{N\bar{M}}\sum_{i,j}\lambda_{ij}^{(0)}u_{ijt}^{(0)}|| \geq \epsilon(N^{-1/2}+M^{-1/2})\ln T) = o(\bar{m}^{-1})$.

(iii) $P(\max_{i,j}||\frac{1}{T}\sum_t f_t^{(0)}u_{ijt}^{(0)}|| \geq \epsilon T^{-1/2}\ln T) = o(\bar{m}^{-1})$, $P(\max_i |\frac{1}{M_iT}\sum_t\sum_{j\in\mathcal{M}_i}(u_{ijt}^{(0)})^2 - c_i^{(0)}| \geq \epsilon) = o(\bar{m}^{-1})$ and $P(\max_j |\frac{1}{N_jT}\sum_t\sum_{i\in\mathcal{N}_j}(u_{ijt}^{(0)})^2 - c_{\cdot j}^{(0)}| \geq \epsilon) = o(\bar{m}^{-1})$, where $\underline{c} \leq c_i^{(0)}, c_{\cdot j}^{(0)} \leq \bar{c}\forall(i,j)$.

Assumption 6.

(i) $P(|\bar{c}_U - \mathbb{E}(\bar{c}_U)| \geq \epsilon) + P\left(\left|\bar{c}_{\boldsymbol{F}^{(\ell)}} - \mathbb{E}(\bar{c}_{\boldsymbol{F}^{(\ell)}})\right| \geq \epsilon\right) = o(\bar{m}^{-1})$for $\ell \in [2]$and $\forall\epsilon > 0$.

(ii) $P(\max_i \left|c_{i,u} - \mathbb{E}(c_{i,u})\right| \geq \epsilon) + P(\max_j \left|c_{\cdot j,u} - \mathbb{E}(c_{\cdot j,u})\right| \geq \epsilon) = o(\bar{m}^{-1})$, and $P(\max_i |c_{\mathbb{F}_i^{(2)}} - \mathbb{E}(c_{\mathbb{F}_i^{(2)}})| \geq \epsilon) + P(\max_j |c_{\mathbb{F}_j^{(1)}} - \mathbb{E}(c_{\mathbb{F}_j^{(1)}})| \geq \epsilon) = o(\bar{m}^{-1})\forall\epsilon > 0$, where $\max_i \mathbb{E}\left(c_{i,u}\right) + \max_j \mathbb{E}\left(c_{j,u}\right) + \max_i \mathbb{E}(c_{\mathbb{F}_i^{(2)}}) + \max_j \mathbb{E}(c_{\mathbb{F}_j^{(1)}}) \leq C$.

(iii) Let $\bar{r} = \max_i r_i^{(1)} \vee \max_j r_j^{(2)}$. There exists an integer $L \geq r_{\max} + 2(\bar{r} + r^{(0)})$ and a fixed constant $\underline{c}_L > 0$ such that $P(\min_i c_{iL}^{(1)} \geq \underline{c}_L) = 1 - o(\bar{m}^{-1})$ and $P(\min_j c_{jL}^{(2)} \geq \underline{c}_L) = 1 - o(\bar{m}^{-1})$, where $c_{iL}^{(1)} = \frac{1}{M_i \vee T}\psi_L(U_iU_i')$and $c_{jL}^{(2)} = \frac{1}{N_j \vee T}\psi_L(U_{\cdot j}U_{\cdot j}')$.

Assumption 4(i) is commonly assumed in the literature and is required for the consistent estimation of certain eigenvectors. Assumptions 4(ii) and 4(iii) are

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

analogous to Assumptions C.2 and C.5 in Bai and Ng (2002), respectively. Bai and Ng (2002) assume that $\max_{s,t}|\gamma_{st}^{(0)}| \leq C$ and $\frac{1}{T}\sum_{s,t}|\gamma_{st}^{(0)}| \leq C$, in parallel with our Assumption 4(ii), both of which can be verified under weak serial dependence conditions. If $u_{ijt}^{(0)}$ does not contain the two local factor components, i.e., $u_{ijt}^{(0)} = u_{ijt}$, we can strengthen the term $N^{-2}+M^{-2}$ to $N^{-2}M^{-2}$ in Assumption 4(iii) and the term $N^{-1}+M^{-1}$ to $N^{-1}M^{-1}$ in Assumption 4(iv) and (v). So the presence of the local factor components slows down the convergence rate of our first-step global factor estimator.

Assumption 5(i)–(iii) imposes some (uniform) convergence rates for a variety of objects. These assumptions can be verified via the use of various Bernstein-type exponential inequalities under various weak dependence and moment conditions. See Section S7.2 of the Supplementary Material for the remark on the verification of Assumption 5(iii). Assumption 6 imposes conditions on the eigenvalues of certain random matrices, which can be verified under some primitive conditions on the elements of the random matrices. In Sections S7.3–S7.4 of the Supplementary Material, we state some primitive conditions to verify Assumption 6 (i) and (iii) and the other conditions in Assumption 6 can be verified similarly.

The following theorem states the uniform consistency of $\tilde{r}_i^{(1)}$ and $\tilde{r}_j^{(2)}$.

THEOREM 3.2. *Suppose that Assumption I holds. Suppose Assumptions 1–6 hold. Let $\tilde{r}_i^{(1)}$ and $\tilde{r}_j^{(2)}$ be the first-step ER or GR estimator of $r_i^{(1)}$ and $r_j^{(2)}$, respectively. Then, as $\underline{m} \to \infty$,*

(i) $P(\tilde{r}_i^{(1)} = r_i^{(1)} \forall i = 1,...,N) \to 1,$
(ii) $P(\tilde{r}_j^{(2)} = r_j^{(2)} \forall j = 1,...,M) \to 1.$

Theorem 3.2 states that the first-step ER or GR estimators $\tilde{r}_i^{(1)}$ and $\tilde{r}_j^{(2)}$ are uniformly consistent. To prove it, we need to study the asymptotic properties of the first-step estimators of the global factors and factor loadings. Let $\Lambda_i^{(0)} = \{\lambda_{ij}^{(0)}\}_{j\in\mathcal{M}_i}$, $\Lambda_{\cdot j}^{(0)} = \{\lambda_{ij}^{(0)}\}_{i\in\mathcal{N}_j}$, $\tilde{\Lambda}_i^{(0)} = \{\tilde{\lambda}_{ij}^{(0)}\}_{j\in\mathcal{M}_i}$ and $\tilde{\Lambda}_{\cdot j}^{(0)} = \{\tilde{\lambda}_{ij}^{(0)}\}_{i\in\mathcal{N}_j}$. As we have demonstrated in Lemma S4.1 in the Supplementary Material, the estimators $\tilde{F}^{(0)}$ and $\{\tilde{\lambda}_{ij}^{(0)}\}$ have the following mean square (MS) convergence rates:

(i) $\frac{1}{T}\left\|\tilde{F}^{(0)} - F^{(0)}\tilde{H}^{(0)}\right\|^2 = O_p(M^{-1}+N^{-1}+T^{-2}),$
(ii) $\frac{1}{M_i}\left\|\tilde{\Lambda}_i^{(0)} - \Lambda_i^{(0)}[\tilde{H}^{(0)\prime}]^{-1}\right\|^2 = O_p(T^{-1}),$
(iii) $\frac{1}{N_j}\left\|\tilde{\Lambda}_{\cdot j}^{(0)} - \Lambda_{\cdot j}^{(0)}[\tilde{H}^{(0)\prime}]^{-1}\right\|^2 = O_p(T^{-1}),$
(iv) $\frac{1}{N\bar{M}}\sum_{i,j}\left\|\tilde{\lambda}_{ij}^{(0)} - [\tilde{H}^{(0)}]^{-1}\lambda_{ij}^{(0)}\right\|^2 = O_p(T^{-1}),$

where $\tilde{H}^{(0)} = [\frac{1}{N\bar{M}}\Lambda^{(0)\prime}\Lambda^{(0)}]\frac{1}{T}F^{(0)\prime}\tilde{F}^{(0)}[\tilde{W}^{(0)}]^{-1}$ and $\tilde{W}^{(0)}$ is the $r^{(0)} \times r^{(0)}$ diagonal matrix of the first $r^{(0)}$ largest eigenvalues of $\frac{1}{N\bar{M}T}YY'$. Essentially, these results indicate that the first-step procedure is able to estimate the global factor loadings at

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

the desirable T^{-1} MS convergence rate; but the global factors can only be estimated at the $(M^{-1}+N^{-1}+T^{-2})$-rate, instead of $[(MN)^{-1}+T^{-2}]$ -rate. In addition, to study the uniform consistency of $\tilde{r}_i^{(1)}$ and $\tilde{r}_j^{(2)}$, we need certain uniform results on $\tilde{\Lambda}_i^{(0)}$ and $\tilde{\Lambda}_j^{(0)}$, which are studied in Lemma S4.2 in the Supplementary Material.

3.2. Consistency of $\hat{r}^{(0)}$, $\hat{r}_i^{(1)}$, and $\hat{r}_j^{(2)}$

To study the uniform consistency of the second-step estimators $\hat{r}^{(0)}$, $\hat{r}_i^{(1)}$, and $\hat{r}_j^{(2)}$, we add some notations. Let $f_{ijt}=(f_t^{(0)\prime},f_{it}^{(1)\prime},f_{jt}^{(2)\prime})^{\prime}$, $\lambda_{ij}=(\lambda_{ij}^{(0)\prime},\lambda_{ij}^{(1)\prime},\lambda_{ij}^{(2)\prime})^{\prime}$, $u_{ijt}^{(1\diamond)}=u_{ijt}+\lambda_{ij}^{(2)\prime}f_{jt}^{(2)}$, and $u_{ijt}^{(2\diamond)}=u_{ijt}+\lambda_{ij}^{(1)\prime}f_{it}^{(1)}$. Let $U_i^{(1\diamond)}$ and $U_j^{(2\diamond)}$ denote the $T\times M_i$ and $T\times N_j$ with typical elements $u_{ijt}^{(1\diamond)}$ and $u_{ijt}^{(2\diamond)}$, respectively. Define

$$\gamma_{i,st}^{(1)}=\frac{1}{M_i}\sum_{j\in\mathcal{M}_i}\mathbb{E}(u_{ijt}^{(1\diamond)}u_{ijs}^{(1\diamond)}),\ \zeta_{i,st}^{(1)}=\frac{1}{M_i}\sum_{j\in\mathcal{M}_i}u_{ijt}^{(1\diamond)}u_{ijs}^{(1\diamond)}-\gamma_{i,st}^{(1)},$$

$$\gamma_{j,st}^{(2)}=\frac{1}{N_j}\sum_{i\in\mathcal{N}_j}\mathbb{E}(u_{ijt}^{(2\diamond)}u_{ijs}^{(2\diamond)}),\ \zeta_{j,st}^{(2)}=\frac{1}{N_j}\sum_{i\in\mathcal{N}_j}u_{ijt}^{(2\diamond)}u_{ijs}^{(2\diamond)}-\gamma_{j,st}^{(2)},$$

$$\gamma_{st}=\frac{1}{N\bar{M}}\sum_{i=1}^{N}\sum_{j\in\mathcal{M}_i}\mathbb{E}(u_{ijt}u_{ijs}),\ \text{and}\ \zeta_{st}=\frac{1}{N\bar{M}}\sum_{i=1}^{N}\sum_{j\in\mathcal{M}_i}u_{ijt}u_{ijs}-\gamma_{st}.$$

The following two assumptions are needed for the study of uniform consistency of $\hat{r}^{(0)}$, $\hat{r}_i^{(1)}$ and $\hat{r}_j^{(2)}$ in Theorem 3.3, which follows.

Assumption 7.

(i) The eigenvalues of $\Sigma_{\Lambda_i^{(1)}}^{1/2}\Sigma_{F_i^{(1)}}\Sigma_{\Lambda_i^{(1)}}^{1/2}$ (resp. $\Sigma_{\Lambda_{\cdot j}^{(2)}}^{1/2}\Sigma_{F_j^{(2)}}\Sigma_{\Lambda_{\cdot j}^{(2)}}^{1/2}$) are distinct for all $r_i^{(1)}\geq 2$ (resp. $r_j^{(2)}\geq 2$).
(ii) $\max_{i,t}\sum_s|\gamma_{i,st}^{(1)}|+\max_{j,t}\sum_s|\gamma_{j,st}^{(2)}|+\max_t\sum_s|\gamma_{st}|\leq C$.
(iii) $\max_{i,s,t}\mathbb{E}|\zeta_{i,st}^{(1)}|^4\leq CM^{-2}$, $\max_{j,s,t}\mathbb{E}|\zeta_{j,st}^{(2)}|^4\leq CN^{-2}$ and $\max_{s,t}\mathbb{E}|\zeta_{st}|^4\leq C(N^{-2}+M^{-2})$.
(iv) $\max_t\mathbb{E}||\sum_{i,j,s}f_s^{(0)}[u_{ijs}u_{ijt}-\mathbb{E}(u_{ijs}u_{ijt})]||^2\leq CN\bar{M}T$ and $\mathbb{E}||\sum_{i,j,t}\lambda_{ij}^{(0)}f_t^{(0)\prime}u_{ijt}||^2\leq CN\bar{M}T$.
(v) $\max_{i,j}\frac{1}{T}\mathbb{E}||\sum_t\varsigma_{ijt}||^2+\max_{i,t}\frac{1}{M_i}\mathbb{E}||\sum_j\varsigma_{ijt}^{(1)}||^2+\max_{j,t}\frac{1}{N_j}\mathbb{E}||\sum_i\varsigma_{ijt}^{(2)}||^2\leq C$ for $\varsigma_{ijt}\in\{f_t^{(0)}f_{it}^{(1)\prime},f_t^{(0)}f_{jt}^{(2)\prime},f_{it}^{(1)}f_{jt}^{(2)\prime},f_{ijt}u_{ijt}\}$, $\varsigma_{ijt}^{(1)}\in\{\lambda_{ij}^{(0)}u_{ijt},\lambda_{ij}^{(1)}u_{ijt},\lambda_{ij}^{(1)}u_{ijt}^{(1\diamond)}\}$, and $\varsigma_{ijt}^{(2)}\in\{\lambda_{ij}^{(0)}u_{ijt},\lambda_{ij}^{(2)}u_{ijt},\lambda_{ij}^{(2)}u_{ijt}^{(2\diamond)}\}$.

Assumption 8.

(i) $\max_i\mathbb{E}||\frac{1}{\sqrt{TM_i}}\sum_{t,j}\lambda_{ij}f_{ijt}^{\prime}u_{ijt}||^2+\max_j\mathbb{E}||\frac{1}{\sqrt{TN_j}}\sum_{t,i}\lambda_{ij}f_{ijt}^{\prime}u_{ijt}||^2\leq C$.

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

(ii) $\max_{i,t}\frac{1}{TM_i}\sum_{j,s}\left\|\mathbb{E}\left(\varsigma_{ijst}\right)\right\| + \max_{j,t}\frac{1}{TN_j}\sum_{i,s}||\mathbb{E}\left(\varsigma_{ijst}\right)|| + \max_{i,t}\mathbb{E}||\frac{1}{\sqrt{TM_i}}\sum_{s,j}[\varsigma_{ijst} - \mathbb{E}\left(\varsigma_{ijst}\right)]||^2 + \max_{j,t}\mathbb{E}||\frac{1}{\sqrt{TN_j}}\sum_{s,i}\left[\varsigma_{ijst} - \mathbb{E}\left(\varsigma_{ijst}\right)\right]||^2 \leq C$ for $\varsigma_{ijst} = f_{ijs}u_{ijs}u_{ijt}$.

(iii) $\max_{i}\frac{1}{TM_i}\sum_{j,s,t}\left\|\mathbb{E}\left(\varsigma_{ijst}\right)\right\| + \max_{j}\frac{1}{TN_j}\sum_{i,s,t}\left\|\mathbb{E}\left(\varsigma_{ijst}\right)\right\| + \max_{i}\mathbb{E}||\frac{1}{TM_i^{1/2}}\sum_{j,s,t}[\varsigma_{ijst} - \mathbb{E}\left(\varsigma_{ijst}\right)]||^2 + \max_{i}\mathbb{E}||\frac{1}{TN_j^{1/2}}\sum_{i,s,t}\left[\varsigma_{ijst} - \mathbb{E}\left(\varsigma_{ijst}\right)\right]||^2 \leq C$ for $\varsigma_{ijst} \in \{f_{it}^{(1)}f_t^{(0)\prime}u_{ijs}u_{ijt}, f_{jt}^{(2)}f_t^{(0)\prime}u_{ijs}u_{ijt}\}$.

(iv) $\max_{i,j}||\frac{1}{T}\sum_t[\varsigma_{ijt} - \mathbb{E}\left(\varsigma_{ijt}\right)]|| = O_p((\ln T/T)^{1/2})$ for $\varsigma_{ijt} \in \{f_{ijt}u_{ijt}, f_{ijt}f_{ijt}^{\prime}, (c_{ijt}^{(1)})^2, (c_{ijt}^{(2)})^2, (u_{ijt}^{(1\diamond)})^2, (u_{ijt}^{(2\diamond)})^2\}$, $\max_{i,s,t}||\varsigma_{i,st}^{(1)}|| = O_p((\ln T/M)^{1/2})$, and $\max_{j,s,t}||\varsigma_{j,st}^{(2)}|| = O_p((\ln T/N)^{1/2})$.

(v) $\max_{i,t}||\frac{1}{M_i}\sum_{j\in\mathcal{M}_i}\varsigma_{ijt}^{(1)}|| = O_p\left((\ln T/M)^{1/2}\right)$ and $\max_{j,t}||\frac{1}{N_j}\sum_{i\in\mathcal{N}_j}\varsigma_{ijt}^{(2)}|| = O_p\left((\ln T/N)^{1/2}\right)$ for $\varsigma_{ijt}^{(l)} \in \{\lambda_{ij}^{(l)}u_{ijt}, \lambda_{ij}^{(l)}u_{ijt}^{(l\diamond)}\}$ for $l \in [2]$.

(vi) $P(\max_i|c_{iU}^{(1)} - \mathbb{E}(c_{iU}^{(1)})| \geq \epsilon) + P(\max_j|c_{jU}^{(2)} - \mathbb{E}(c_{jU}^{(2)})| \geq \epsilon) = o(\bar{m}^{-1})$, where $c_{iU}^{(1)} = \frac{1}{TM_i}\mathrm{tr}(U_i^{(1\diamond)}U_i^{(1\diamond)\prime}), c_{jU}^{(2)} = \frac{1}{TN_j}\mathrm{tr}(U_j^{(2\diamond)}U_j^{(2\diamond)\prime})$, and $\underline{c} \leq \mathbb{E}(c_{iU}^{(1)}), \mathbb{E}(c_{jU}^{(2)}) \leq \bar{c}$.

Assumption 7(i)–(iv) parallels Assumption 4(i)–(v). The first part of Assumption 7(v) imposes weak serial dependence in the process $\{(f_t^{(0)}, f_{it}^{(1)}, f_{jt}^{(2)}, u_{ijt}), t \geq 1\}$ and the other two parts require that $\{u_{ijt}\}$ be weakly correlated along the i and j crosssection dimensions. Assumption 8(i)–(iii) imposes some moment conditions that are satisfied under weak serial or cross-sectional dependence. Assumption 8(iv)–(vi) imposes some uniform probability orders that can be verified under primitive conditions.

Given the uniform consistency of the first-step estimators $\tilde{r}^{(0)}$, $\tilde{r}_i^{(1)}$, and $\tilde{r}_j^{(2)}$, we can obtain preliminary consistent estimates of the local factors and factor loadings. But because these are the intermediate results, we relegate them to Section S4 of the Supplementary Material to save space. The following theorem studies the (uniform) consistency of the second-step estimators $\hat{r}^{(0)}$, $\hat{r}_i^{(1)}$ and $\hat{r}_j^{(2)}$.

THEOREM 3.3. *Suppose that Assumption I holds. Suppose Assumptions 1–8 hold. Let* $\hat{r}^{(0)}, \hat{r}_i^{(1)}$ *and* $\hat{r}_j^{(2)}$ *be the second-step ER or GR estimator of* $r^{(0)}, r_i^{(1)}$ *and* $r_j^{(2)}$, *respectively. Then, as* $\underline{m} \to \infty$,

(i) $P\left(\hat{r}^{(0)} = r^{(0)}\right) \to 1$,

(ii) $P(\hat{r}_i^{(1)} = r_i^{(1)} \forall i = 1, ..., N) \to 1$,

(iii) $P(\hat{r}_j^{(2)} = r_j^{(2)} \forall j = 1, ..., M) \to 1$.

Theorem 3.3 implies that the second-step estimators $\hat{r}^{(0)}, \hat{r}_i^{(1)}$, and $\hat{r}_j^{(2)}$ are consistent uniformly in (i,j). Simulations demonstrate that these estimators typically

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

outperform the first-step estimators $\tilde{r}^{(0)}, \tilde{r}_i^{(1)}$, and $\tilde{r}_j^{(2)}$ and thus we recommend them for practical use.

Remark 8. We have shown the consistency of our estimators of the number of factors. Given this result, we can further test the hypothesis about the number of factors. For example, we may be interested in testing the null hypothesis: $r_j^{(2)} = 0, \forall j = 1, ..., M$. That is, there are only global factors and one type of local factors as in model (1.3). One simple approach is compare the difference between the estimated factor components under the null and under the alternative. We can also borrow the idea of testing rank of matrices, as in Kleibergen and Paap (2006) and Chen and Fang (2019). We leave the details for future research.

4. ASYMPTOTIC PROPERTIES OF THE SECOND-STEP ESTIMATORS OF THE FACTORS AND FACTOR LOADINGS

In this section, we study the asymptotic distributions of the estimators of the global and local factors and factor loadings. Given the fact that we can consistently estimate $r^{(0)}, r_i^{(1)}$, and $r_j^{(2)}$, we assume that they are known in the following analyses. We assume that $r^{(0)} > 0$ but do allow $r_i^{(1)}$ and $r_j^{(2)}$ to be 0 for some i, j. We comment on the case of $r^{(0)} = 0$ in Remark 10, which follows.

4.1. Asymptotic Properties of the Second-Step Estimators of Global Factors and Factor Loadings

To state the asymptotic properties of the second-step estimators of global factors, we introduce some new notations. Let $U_{ij\cdot} = \left(u_{ij1}, ..., u_{ijT}\right)'$, $U_{i\cdot t} = \{u_{ijt}\}_{j \in \mathcal{M}_i}, U_{\cdot jt} = \{u_{ijt}\}_{i \in \mathcal{N}_j}$, and $U_{\cdot\cdot t} = \{u_{ijt}\}_{i \in [N],\, j \in \mathcal{M}_i}$. Let $\lambda_{ij} = (\lambda_{ij}^{(0)\prime}, \lambda_{ij}^{(1)\prime}, \lambda_{ij}^{(2)\prime})'$, $c_{1i} = \frac{1}{\bar{M}} \Lambda_i^{(0)\prime} \Lambda_i^{(0)} (\frac{1}{N\bar{M}} \Lambda^{(0)\prime} \Lambda^{(0)})^{-1}$, and $c_{2j} = \frac{1}{\bar{N}} \Lambda_{\cdot j}^{(0)\prime} \Lambda_{\cdot j}^{(0)} (\frac{1}{N\bar{M}} \Lambda^{(0)\prime} \Lambda^{(0)})^{-1}$. Let $\mathbb{E}_{\mathcal{F}}(\cdot) = \mathbb{E}(\cdot|\mathcal{F})$ and $\mathcal{F}$ denotes the minimal sigma-field generated by $\{f_t^{(0)}\}, \{f_{it}^{(1)}\}$, and $\{f_{jt}^{(2)}\}$. Finally,

$$
\begin{aligned}
\boldsymbol{V}_t &= \frac{1}{\sqrt{N\bar{M}}} \sum_{i,j} (\lambda_{ij}^{(0)} - \chi_{1i} \lambda_{ij}^{(1)} - \chi_{2j} \lambda_{ij}^{(2)}) u_{ijt}, \\
\boldsymbol{B}_{1t} &= \frac{1}{N\bar{M}T} \sum_{i,j,s} \tilde{H}^{(0)\prime} f_s^{(0)} \mathbb{E}(u_{ijs} u_{ijt}), \\
\boldsymbol{B}_{2t} &= \frac{1}{N\bar{M}T} \sum_{i,j,s} [\tilde{H}^{(0)}]^{-1} \chi_{1i} H_i^{(1)} H_i^{(1)\prime} \mathbb{E}(f_{is}^{(1)} u_{ijs} u_{ijt}), \\
\boldsymbol{B}_{3t} &= \frac{1}{N\bar{M}T} \sum_{i,j,s} [\tilde{H}^{(0)}]^{-1} \chi_{2j} H_j^{(2)} H_j^{(2)\prime} \mathbb{E}(f_{js}^{(2)} u_{ijs} u_{ijt}), \text{ and} \\
\boldsymbol{B}_t &= \boldsymbol{B}_{1t} - \boldsymbol{B}_{2t} - \boldsymbol{B}_{3t},
\end{aligned}
\tag{4.1}
$$

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

where χ_{1i} and χ_{2j} are defined in (S3) in Section S1 of the Supplementary Material, $H_i^{(1)}$ and $H_j^{(2)}$ are, respectively, the probability limits of $\tilde{H}_i^{(1)}$ and $\tilde{H}_j^{(2)}$ defined in (S5) in Section S1 of the Supplementary Material (see also the statement of Theorem 4.1, which follows). Let $\breve{H}^{(0)}$ and Ω_0 be as defined, respectively in (S4) and (S2) in Section S1 of the Supplementary Material. Let $\omega_1^{ii_1}$ and $\omega_2^{jj_1}$ be as defined below (S2).

To study the asymptotic properties of the second-step estimators of the global and local factors in Theorem 4.1, we add the following two assumptions.

Assumption 9.

(i) $\max_{i,j,t}\mathbb{E}\left\|f_{ijt}\right\|^{2q}\le C$ for some $q>2$.
(ii) $\frac{1}{T}\sum_r\mathbb{E}||\frac{1}{\sqrt{N}}\sum_i c_{1i}[\xi_{1i,rt}-\mathbb{E}(\xi_{1i,rt})]||^2+\frac{1}{T}\sum_r\mathbb{E}||\frac{1}{\sqrt{M}}\sum_j c_{2j}[\xi_{2j,rt}-\mathbb{E}(\xi_{2j,rt})]||^2\le C,$ where $\xi_{1i,rt}=f_{ir}^{(1)\prime}H_i^{(1)}H_i^{(1)\prime}f_{it}^{(1)}$ and $\xi_{2j,rt}=f_{jr}^{(2)\prime}H_j^{(2)}H_j^{(2)\prime}f_{jt}^{(2)}$.
(iii) $\underline{c}\le\mu_{\min}(\boldsymbol{Q})\le\bar{c},$ $\max_i\sum_{i_1=1}^N||\omega_1^{ii_1}||\le\bar{c},$ and $\max_j\sum_{j_1=1}^M||\omega_2^{jj_1}||\le\bar{c}.$
(iv) $\max_{i,j}||\frac{1}{\sqrt{T}}F^{(0)\prime}U_{ij\cdot}||+\max_t||\sqrt{\frac{N\bar{M}}{T}}\sum_{s=1}^T f_s^{(0)}\zeta_{st}||=O_p((\ln T)^{1/2}),$ and $\max_{i,t}\frac{1}{\sqrt{M_i}}||\Lambda_i^{(l)\prime}U_{i\cdot t}||+\max_{j,t}\frac{1}{\sqrt{N_j}}||\Lambda_j^{(l)\prime}U_{\cdot jt}||+\max_t||\frac{1}{\sqrt{N\bar{M}}}\Lambda^{(0)\prime}U_{\cdot\cdot t}||=O_p((\ln T)^{1/2})$ for $l=0,1,2$.
(v) $\max_i||\frac{1}{\sqrt{TM_i}}\sum_{j\in\mathcal{M}_i}\sum_{t=1}^T\xi_{ijt}||+\max_j||\frac{1}{\sqrt{TN_j}}\sum_{i\in\mathcal{N}_j}\sum_{t=1}^T\xi_{ijt}||\}=O_p((\ln T)^{1/2})$ for $\xi_{ijt}\in\{f_t^{(0)}u_{ijt}\lambda_{ij}^{\prime},H_i^{(1)\prime}f_{it}^{(1)}u_{ijt}\lambda_{ij}^{\prime},H_j^{(2)\prime}f_{jt}^{(2)}u_{ijt}\lambda_{ij}^{\prime}\}$.
(vi) $\max_i\frac{1}{M_iT}\sum_{j,j_1,t}\left|\mathbb{E}\left(u_{ij_1t}u_{ijt}\right)\right|+\max_j\frac{1}{N_jT}\sum_{i,i_1,t}\left|\mathbb{E}\left(u_{ijt}u_{i_1jt}\right)\right|+\max_{i,j}\frac{1}{T}\sum_{r,s}\{||\mathbb{E}[f_{is}^{(1)}f_{ir}^{(1)\prime}u_{ijs}u_{ijr}]||+||\mathbb{E}[f_{js}^{(2)}f_{jr}^{(2)\prime}u_{ijs}u_{ijr}]||\}\le C,$ and $\frac{1}{N\bar{M}T}\sum_{i,j,r}\sum_{i_1,j_1,r_1}\left|\mathbb{E}_{\mathcal{F}}(u_{ijr}u_{i_1j_1r_1})\right|=O_p(1)$.
(vii) $(N+M)(T^{-q/2}+T^{-2+2/q})=o(1)$.

Assumption 10.

(i) For each t, $\frac{1}{(N\bar{M})^{1/2}}\sum_{i,j}(\lambda_{ij}^{(0)}-\chi_{1i}\lambda_{ij}^{(1)}-\chi_{2j}\lambda_{ij}^{(2)})u_{ijt}\xrightarrow{d}\mathcal{N}(0,\Gamma_t^{(0)})$ for some $\Gamma_t^{(0)}>0$.
(ii) For each (i,t) with $r_i^{(1)}\ge 1$, $\sum_{i_1=1}^N\omega_1^{ii_1}\frac{1}{\sqrt{M_{i_1}}}\sum_{j\in\mathcal{M}_{i_1}}\lambda_{i_1j}^{(1)}u_{i_1jt}\xrightarrow{d}\mathcal{N}(0,\Gamma_{it}^{(1)})$ for some $\Gamma_{it}^{(1)}>0$.
(iii) For each (j,t) with $r_j^{(2)}\ge 1$, $\sum_{j_1=1}^M\omega_2^{jj_1}\frac{1}{\sqrt{N_{j_1}}}\sum_{i\in\mathcal{N}_{j_1}}\lambda_{ij_1}^{(2)}u_{ij_1t}\xrightarrow{d}\mathcal{N}(0,\Gamma_{jt}^{(2)})$ for some $\Gamma_{jt}^{(2)}>0$.

Assumption 9(i)–(ii) imposes some additional moment conditions. Assumption 9(iii) ensures the large dimensional matrix $\boldsymbol{Q}$ and its inverse to be well behaved. Assumption 9 (iv)–(vi) imposes some uniform convergence conditions. Like Assumptions F.3 and F.4 in Bai (2003), Assumption 10(i)–(iii) requires that the normalized sample mean objects obey some versions of the central limit theorem.

Theorem 4.1 reports the asymptotic distributions of the second-step estimators of the factors.

THEOREM 4.1. *Suppose that Assumption I holds. Suppose Assumptions 1–10 hold. Then, we have the following.*

(i) $\frac{1}{T}\left\|\breve{F}^{(0)}-F^{(0)}\breve{H}^{(0)}\right\|^{2}=O_p((NM)^{-1}+T^{-2})$,

(ii) $\sqrt{N\bar{M}}(\breve{f}_t^{(0)}-\breve{H}^{(0)\prime}f_t^{(0)}-\tilde{H}^{(0)\prime}\Omega_0^{-1}\tilde{H}^{(0)}\boldsymbol{B}_t)=H^{(0)\prime}\Omega_0^{-1}\boldsymbol{V}_t+o_p(1)$ $\xrightarrow{d}\mathcal{N}(0,\lim_{(N,M)\to\infty}[H^{(0)\prime}]\Omega_0^{-1}\times\Gamma_t^{(0)}\Omega_0^{-1}H^{(0)})$,

(iii) $\sqrt{M_i}(\breve{f}_{it}^{(1)}-\tilde{H}_i^{(1)\prime}f_{it}^{(1)})=\tilde{H}_i^{(1)\prime}\sum_{i_1=1}^{N}\omega_1^{ii_1}\frac{1}{\sqrt{M_{i_1}}}\sum_{j\in\mathcal{M}_{i_1}}\lambda_{i_1j}^{(1)}u_{i_1jt}+o_p(1)\xrightarrow{d}$ $\mathcal{N}(0,H_i^{(1)\prime}\Gamma_{it}^{(1)}H_i^{(1)})$,

(iv) $\sqrt{N_j}(\breve{f}_{jt}^{(2)}-\tilde{H}_j^{(2)\prime}f_{jt}^{(2)})=\tilde{H}_j^{(2)\prime}\sum_{j_1=1}^{M}\omega_2^{jj_1}\frac{1}{\sqrt{N_{j_1}}}\sum_{i\in\mathcal{N}_{j_1}}\lambda_{ij_1}^{(2)}u_{ij_1t}+o_p(1)\xrightarrow{d}\mathcal{N}(0,$ $H_j^{(2)\prime}\Gamma_{jt}^{(2)}H_j^{(2)})$,

where $H^{(0)}=\Sigma_{\Lambda^{(0)}}^{1/2}\Upsilon^{(0)}(W^{(0)})^{-1/2}, W^{(0)}$ *denotes the diagonal matrix consisting of the eigenvalues of* $\Sigma_{\Lambda^{(0)}}^{1/2}\Sigma_{F^{(0)}}\Sigma_{\Lambda^{(0)}}^{1/2}$ *in descending order with the corresponding eigenvector matrix denoted as* $\Upsilon^{(0)}$ *such that* $\Upsilon^{(0)\prime}\Upsilon^{(0)}=\mathbb{I}_{r^{(0)}}; H_i^{(1)}=\Sigma_{\Lambda_i^{(1)}}^{1/2}\Upsilon_i^{(1)}(W_i^{(1)})^{-1/2}, W_i^{(1)}$ *denotes the diagonal matrix consisting of the eigenvalues of* $\Sigma_{\Lambda_i^{(1)}}\Sigma_{F_i^{(1)}}$ *in descending order with the corresponding eigenvector matrix denoted as* $\Upsilon_i^{(1)}$ *such that* $\Upsilon_i^{(1)\prime}\Upsilon_i^{(1)}=\mathbb{I}_{r_i^{(1)}}$; *and* $H_j^{(2)}$ *is analogously defined.*

Remark 9. Theorem 4.1(i) reports the MS convergence rate of $\breve{F}^{(0)}$, and Theorem 4.1(ii) reports the asymptotic distribution of $\breve{f}_t^{(0)}$. Let $\ddot{f}_t^{(0)}$ denote the PCA estimator of $f_t^{(0)}$ in (1.1) by assuming the absence of the local factor components under the same normalization rules as used to obtain the first-step estimator $\tilde{f}_t^{(0)}$. Let $\ddot{F}^{(0)}=(\ddot{f}_1^{(0)},...,\ddot{f}_T^{(0)})'$. Then, we can readily show that $\frac{1}{T}\left\|\ddot{F}^{(0)}-F^{(0)}\ddot{H}^{(0)}\right\|^2=O_p((NM)^{-1}+T^{-2})$, and

$$\sqrt{N\bar{M}}(\ddot{f}_t^{(0)}-\ddot{H}^{(0)\prime}f_t^{(0)}-\ddot{H}^{(0)\prime}\boldsymbol{Q}_{00}^{-1}\ddot{H}^{(0)}\boldsymbol{B}_{1t})=[W^{(0)}]^{-1}\frac{1}{T}\ddot{F}^{(0)\prime}F^{(0)}\ddot{\boldsymbol{V}}_t+o_p(1)$$
$$=H^{(0)\prime}\boldsymbol{Q}_{00}^{-1}\ddot{\boldsymbol{V}}_t+o_p(1)\xrightarrow{d}\mathcal{N}(0,\lim_{(N,M)\to\infty}H^{(0)\prime}\boldsymbol{Q}_{00}^{-1}\ddot{\Gamma}_t^{(0)}\boldsymbol{Q}_{00}^{-1}H^{(0)}),$$

where $\boldsymbol{Q}_{00}=\frac{1}{N\bar{M}}\Lambda^{(0)\prime}\Lambda^{(0)},\ddot{H}^{(0)}$ is a rotational matrix, $\ddot{\boldsymbol{V}}_t=\frac{1}{\sqrt{N\bar{M}}}\sum_{i,j}\lambda_{ij}^{(0)}u_{ijt}$, and $\ddot{\Gamma}_t^{(0)}$ denotes the asymptotic variance of $\ddot{\boldsymbol{V}}_t$. Note that $\boldsymbol{B}_{1t}$ is present even in the absence of the local factors.[9] Obviously, the MS rate $O_p(T^{-2}+(NM)^{-1})$ is the

[9] In Bai (2003)'s 2D factor model with N cross-sectional units and T time series observations, the term that corresponds to our $\boldsymbol{B}_{1t}$ is also of $O_p(T^{-1})$, which is $o_p(N^{-1/2})$ under the usual condition $N/T^2=o(1)$ and thus asymptotically vanishing. In contrast, the cross-sectional dimension in our case is $N\bar{M}$, which explains why we need $NM/T^2=o(1)$ in order for $\boldsymbol{B}_{1t}$ to vanish here.

optimal rate that is achievable in a 3D factor model in the absence of the local factors. But our second-step estimator $\breve{f}_t^{(0)}$ of the global factors have three bias terms associated with $\boldsymbol{B}_{1t}, \boldsymbol{B}_{2t},$ and $\boldsymbol{B}_{3t},$ all of which can be shown to be $O_p\left(T^{-1}\right)$ and would be vanishing if $NM/T^2 \to 0$. In general, $\breve{f}_t^{(0)}$ is not as asymptotically efficient as $\ddot{f}_t^{(0)}$. This reflects the cost of estimating the local factors and factor loadings whose slow convergence rates generally affect the asymptotic distribution of $\breve{f}_t^{(0)}$. To see the sufficient conditions to ensure the asymptotic equivalence of $\breve{f}_t^{(0)}$ and $\ddot{f}_t^{(0)}$, we focus on a special case where the local factor loadings are nearly orthogonal to the global ones in the sense

$$\frac{1}{M_i}\Lambda_i^{(0)\prime}\Lambda_i^{(1)} = o(1) \text{ uniformly in } i \text{ and } \frac{1}{N_j}\Lambda_{\cdot j}^{(0)\prime}\Lambda_{\cdot j}^{(2)} = o(1) \text{ uniformly in } j. \tag{4.2}$$

In this case, we can readily show that $\sqrt{N\bar{M}}\tilde{H}^{(0)\prime}\Omega_0^{-1}\tilde{H}^{(0)}\boldsymbol{B}_t = \sqrt{N\bar{M}}\ddot{H}^{(0)\prime}\boldsymbol{Q}_{00}^{-1}\ddot{H}^{(0)}\boldsymbol{B}_{1t} + o_p(1)$ and $\Omega_0^{-1}\boldsymbol{V}_t = \boldsymbol{Q}_{00}^{-1}\ddot{\boldsymbol{V}}_t + o_p(1)$. As a result, $\breve{f}_t^{(0)}$ is as asymptotically efficient as $\ddot{f}_t^{(0)}$ and we say it is **oracle efficient** in this case.

Remark 10. Theorem 4.1(iii) and (iv) reports the asymptotic distributions of $\breve{f}_{it}^{(1)}$, and $\breve{f}_{jt}^{(2)}$, respectively. Let $\ddot{f}_{it}^{(1)}$ denote the PCA estimator of $f_{it}^{(1)}$ in (1.1) by assuming the absence of $f_t^{(0)}$ and $f_{jt}^{(2)}$ under the same normalization rules as used to obtain the first-step estimator $\tilde{f}_{it}^{(1)}$. Define $\ddot{f}_{jt}^{(2)}$ analogously. Then, we can show that

$$\sqrt{M_i}(\ddot{f}_{it}^{(1)} - \ddot{H}_i^{(1)\prime}f_{it}^{(1)}) \xrightarrow{d} \mathcal{N}(0, [W_i^{(1)}]^{-1}Q_i^{(1)}\ddot{\Gamma}_{it}^{(1)}Q_i^{(1)\prime}[W_i^{(1)}]^{-1}), \text{ and}$$
$$\sqrt{N_j}(\ddot{f}_{jt}^{(2)} - \ddot{H}_j^{(2)\prime}f_{jt}^{(2)}) \xrightarrow{d} \mathcal{N}(0, [W_j^{(2)}]^{-1}Q_j^{(2)}\ddot{\Gamma}_{jt}^{(2)}Q_j^{(2)\prime}[W_j^{(2)}]^{-1}),$$

where $\ddot{H}_i^{(1)}$ and $\ddot{H}_j^{(2)}$ are certain rotational matrices; $W_i^{(1)}$ is as defined in Theorem 4.1 and $Q_i^{(1)} = [W_i^{(1)}]^{1/2}\Upsilon_i^{(1)\prime}\Sigma_{\Lambda_i^{(1)}}^{-1/2}$; and $W_j^{(2)}$ and $Q_j^{(2)}$ are analogously defined; $\ddot{\Gamma}_{it}^{(1)}$ and $\ddot{\Gamma}_{jt}^{(2)}$ are the asymptotic variances of $\frac{1}{\sqrt{M_i}}\sum_{j\in\mathcal{M}_i}\lambda_{ij}^{(1)}u_{ijt}$ and $\frac{1}{\sqrt{N_j}}\sum_{i\in\mathcal{N}_j}\lambda_{ij}^{(2)}u_{ijt}$, respectively. Like $\ddot{f}_{it}^{(1)}$ and $\ddot{f}_{jt}^{(2)}$, the estimators $\breve{f}_{it}^{(1)}$ and $\breve{f}_{jt}^{(2)}$ do not have asymptotic biases. But $\breve{f}_{it}^{(1)}$ (resp. $\breve{f}_{jt}^{(2)}$) is generally not asymptotically equivalent to the infeasible estimator $\ddot{f}_{it}^{(1)}$ (resp. $\ddot{f}_{jt}^{(2)}$). Exceptions occur when $\max_i ||\frac{1}{M_i}\sum_{j\in\mathcal{M}_i}\lambda_{ij}^{(1)}\lambda_{ij}^{(2)\prime}|| = o(1)$, $\max_j ||\frac{1}{N_j}\sum_{i\in\mathcal{N}_j}\lambda_{ij}^{(1)}\lambda_{ij}^{(2)\prime}|| = o(1)$, and N and M pass to infinity at the same rate. In this case, we can readily show that

$$\sum_{i_1=1}^{N}\omega_1^{ii_1}\frac{1}{\sqrt{M_{i_1}}}\sum_{j\in\mathcal{M}_{i_1}}\lambda_{i_1j}^{(1)}u_{i_1jt} = (\frac{1}{M_i}\Lambda_i^{(1)\prime}\Lambda_i^{(1)})^{-1}\frac{1}{\sqrt{M_i}}\sum_{j\in\mathcal{M}_i}\lambda_{ij}^{(1)}u_{ijt} + o_p(1),$$

$$\sum_{j_1=1}^{M}\omega_2^{jj_1}\frac{1}{\sqrt{N_{j_1}}}\sum_{i\in\mathcal{N}_{j_1}}\lambda_{ij_1}^{(2)}u_{ij_1t} = (\frac{1}{N_j}\Lambda_{\cdot j}^{(2)\prime}\Lambda_{\cdot j}^{(2)})^{-1}\frac{1}{\sqrt{N_j}}\sum_{i\in\mathcal{N}_j}\lambda_{ij}^{(2)}u_{ijt} + o_p(1),$$

$\tilde{H}_i^{(1)} = \ddot{H}_i^{(1)} + o_p(M_i^{-1/2})$, and $\tilde{H}_j^{(2)} = \ddot{H}_j^{(2)} + o_p(N_j^{-1/2})$. Then, $\breve{f}_{it}^{(1)}$ and $\breve{f}_{jt}^{(2)}$ are asymptotically equivalent to their infeasible versions, respectively. In addition, it is easy to see that even if $r^{(0)} = 0$, the results in Theorem 4.1(iii)–(iv) continue to hold.

To make inferences on the global factors, one needs to consistently estimate the asymptotic variance and biases. See Section 4.3, which follows and Section S6 of the Supplementary Material for the discussion on the estimates.

4.2. Asymptotic Properties of the Estimators of the Factor Loadings

In this section, we study the asymptotic properties of the second-step estimators of the global and local factor loadings.

Let $\breve{f}_{ijt} = (\breve{f}_t^{(0)\prime}, \breve{f}_{it}^{(1)\prime}, \breve{f}_{jt}^{(2)\prime})'$ and $\breve{F}_{ij} = (\breve{f}_{ij1}, ..., \breve{f}_{ijT})'$. For each (i,j) pair, consider the time series OLS regression of y_{ij} on $\breve{F}_{ij}$

$$y_{ij} = \breve{F}_{ij}\lambda_{ij} + \breve{v}_{ij},$$

where $\lambda_{ij}^{\ddagger} = (([\breve{H}^{(0)}]^{-1}\lambda_{ij}^{(0)})', ([\tilde{H}_i^{(1)}]^{-1}\lambda_{ij}^{(1)})', ([\tilde{H}_j^{(2)}]^{-1}\lambda_{ij}^{(2)})')'$ denotes the "true" value of λ_{ij} in the above regression and $\breve{v}_{ij} \equiv y_{ij} - \breve{F}_{ij}\lambda_{ij}^{\ddagger}$. Let $\breve{\lambda}_{ij} = (\breve{\lambda}_{ij}^{(0)\prime}, \breve{\lambda}_{ij}^{(1)\prime}, \breve{\lambda}_{ij}^{(2)\prime})'$ denote the OLS estimator of λ_{ij} in the above regression. Let $\tilde{H}_{ij} =$ bdiag$(\tilde{H}^{(0)}, \tilde{H}_i^{(1)}, \tilde{H}_j^{(1)})$ and $H_{ij} =$bdiag$(H^{(0)}, H_i^{(1)}, H_j^{(1)})$. The following theorem reports the asymptotic distribution of $\breve{\lambda}_{ij}$.

THEOREM 4.2. *Suppose that Assumption I holds. Suppose Assumptions 1–10 hold. If* $\frac{1}{T^{1/2}}\sum_t f_{ijt}u_{ijt} \xrightarrow{d} \mathcal{N}(0, \Gamma_{ij})$ *for some* $\Gamma_{ij} > 0$, *then we have*

$$\sqrt{T}(\breve{\lambda}_{ij} - \lambda_{ij}^{\ddagger}) \xrightarrow{d} \mathcal{N}(0,\ H_{ij}'\Gamma_{ij}H_{ij}).$$

Remark 11. Theorem 4.2, in conjunction with the result in Theorem 2 of Bai (2003), implies that for $\ell = 0, 1, 2$, $\breve{\lambda}_{ij}^{(\ell)}$ is asymptotically equivalent to $\ddot{\lambda}_{ij}^{(\ell)}$, which is obtained by assuming the absence of the other two factor components in (1.1). In this sense, we can say these estimators enjoy the *oracle efficiency* property. In particular, for the estimator of the global factor loadings, we have

$$\sqrt{T}(\breve{\lambda}_{ij}^{(0)} - [\breve{H}^{(0)}]^{-1}\lambda_{ij}^{(0)}) \xrightarrow{d} \mathcal{N}(0,\ H^{(0)\prime}\Gamma_{ij}^{(0)}H^{(0)}),$$

where $\Gamma_{ij}^{(0)} \equiv \lim_{T\to\infty}\mathrm{Var}(\frac{1}{T^{1/2}}\sum_t f_t^{(0)}u_{ijt})$. Nevertheless, $\breve{\lambda}_{ij}^{(0)}, \breve{\lambda}_{ij}^{(1)}$, and $\breve{\lambda}_{ij}^{(2)}$ are not asymptotically independent because Γ_{ij} is generally not a block diagonal matrix.

4.3. Inference on the Global Factors

In this section, we consider inferences on the global factors. We relegate the inferences on the global factor loadings to the Section S6.2of the Supplementary

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

Material. The inferences on the local factors and factor loadings are analogous to those in the 2D case and thus omitted.

Let $\Psi_t^{(0)} = [H^{(0)\prime}]\Omega_0^{-1}\Gamma_t^{(0)}\Omega_0^{-1}H^{(0)}$, the asymptotic variance of $\breve{f}_t^{(0)}$. To make inferences on $f_t^{(0)}$, we need to estimate both the asymptotic variance $\Psi_t^{(0)}$ and the asymptotic bias $H^{(0)\prime}\Omega_0^{-1}H^{(0)}\boldsymbol{B}_t$. Let $U_{\cdot\cdot t} = \{u_{ijt}\}_{i\in[N],\, j\in\mathcal{M}_i}$. We assume that the large-dimensional process $\{U_{\cdot\cdot t},\ t\geq 1\}$ is covariance-stationary with variance-covariance matrix $\Sigma \equiv \mathbb{E}(U_{\cdot\cdot t}U_{\cdot\cdot t}')$, an $N\bar{M}\times N\bar{M}$ matrix. Let $\Lambda = \{\lambda_{ij}^{(0)} - \chi_{1i}\lambda_{ij}^{(1)} - \chi_{2j}\lambda_{ij}^{(2)}\}$, an $N\bar{M}\times r^{(0)}$ matrix, such that $\boldsymbol{V}_t = \frac{1}{\sqrt{N\bar{M}}}\Lambda' U_{\cdot\cdot t}$. Let $\breve{\Lambda} = \{\breve{\lambda}_{ij}^{(0)} - \breve{\chi}_{1i}\breve{\lambda}_{ij}^{(1)} - \breve{\chi}_{2j}\breve{\lambda}_{ij}^{(2)}\}$, an $N\bar{M}\times r^{(0)}$ matrix, where $\breve{\chi}_{1i}$ and $\breve{\chi}_{2j}$are defined in (S7) in the Supplementary Material.

Note that

$$H^{(0)\prime}\Omega_0^{-1}\frac{1}{\sqrt{N\bar{M}}}\Lambda' U_{\cdot\cdot t} = (H^{(0)\prime}\Omega_0^{-1}H^{(0)})[H^{(0)}]^{-1}\frac{1}{\sqrt{N\bar{M}}}\Lambda' U_{\cdot\cdot t},$$

where $\mathrm{Var}(\Lambda' U_{\cdot\cdot t}) = \Lambda'\Sigma\Lambda$. We propose to estimate $H^{(0)\prime}\Omega_0^{-1}H^{(0)}$ by $\breve{\Omega}_0^{-1}$, where $\breve{\Omega}_0$ is as defined below (S6) in the Supplementary Material. Suppose that $\breve{\Sigma}$ is a consistent estimator of Σ in the sense $||\breve{\Sigma}-\Sigma||_{\mathrm{sp}} = o_p(1)$. Then, we can estimate $\Psi_t^{(0)}$ consistently by

$$\breve{\Psi}_t^{(0)} \equiv \breve{\Omega}_0^{-1}\breve{\Gamma}_t\breve{\Omega}_0^{-1},$$

where $\breve{\Gamma}_t \equiv \breve{\Gamma} \equiv \frac{1}{N\bar{M}}\breve{\Lambda}'\breve{\Sigma}\breve{\Lambda}$ is a consistent estimator of $[H^{(0)}]^{-1}\Gamma_t^{(0)}[H^{(0)\prime}]^{-1}$. So the key is to find an estimator $\breve{\Sigma}$ such that $||\breve{\Sigma}-\Sigma||_{\mathrm{sp}} = o_p(1)$.

Let $\breve{u}_{ijt} = y_{ijt} - \breve{\lambda}_{ij}^{(0)\prime}\breve{f}_t^{(0)} - \breve{\lambda}_{ij}^{(1)\prime}\breve{f}_{it}^{(1)} - \breve{\lambda}_{ij}^{(2)\prime}\breve{f}_{jt}^{(2)}$. Define

$$\breve{\sigma}_{ij,i_1j_1} = \frac{1}{T}\sum_{t=1}^{T}\breve{u}_{ijt}\breve{u}_{i_1j_1t} \text{ and } \breve{\theta}_{ij,i_1j_1} = \frac{1}{T}\sum_{t=1}^{T}\left(\breve{u}_{ijt}\breve{u}_{i_1j_1t} - \breve{\sigma}_{ij,i_1j_1}\right)^2.$$

We follow the lead of Fan, Liao, and Mincheva (2013)'s POET estimator and propose to estimate Σ by $\breve{\Sigma} = \{\breve{\sigma}_{ij,i_1j_1}^{\mathcal{T}}\}$, where

$$\breve{\sigma}_{ij,i_1j_1}^{\mathcal{T}} = \begin{cases} \breve{\sigma}_{ij,i_1j_1} & \text{if } (i,j) = (i_1,j_1) \\ s_{ij,i_1j_1}\left(\breve{\sigma}_{ij,i_1j_1}\right) & \text{if } (i,j) \neq (i_1,j_1) \end{cases},$$

where $s_{ij,i_1j_1}(\cdot)$ is the soft thresholding function: $s_{ij,i_1j_1}(z) \equiv sgn(z)\left(|z| - \tau_{ij,i_1j_1}\right)_+$, $\tau_{ij,i_1j_1} = C_1(\underline{m}^{-1}\ln T)^{1/2}(\breve{\theta}_{ij,i_1j_1})^{1/2}$, and C_1 is a positive constant. We will show that $||\breve{\Sigma}-\Sigma||_{\mathrm{sp}} = o_p(1)$ under some additional conditions.

Let ι_l be the lth column of $\mathbb{I}_{r^{(0)}}$. Let $\breve{\nu}_{ts} = (\breve{\nu}_{ts,1},\ldots,\breve{\nu}_{ts,r^{(0)}})'$, where $\breve{\nu}_{ts,l} = \frac{1}{N\bar{M}}\sum_{i,j}\iota_l'(\breve{f}_s^{(0)} - \breve{\chi}_{1i}\breve{f}_{is}^{(1)} - \breve{\chi}_{2j}\breve{f}_{js}^{(2)})\breve{u}_{ijs}\breve{u}_{ijt}$. Let

$$\breve{\theta}_{ts,l} = \frac{1}{N\bar{M}}\sum_{i,j}\left[\iota_l'(\breve{f}_s^{(0)} - \breve{\chi}_{1i}\breve{f}_{is}^{(1)} - \breve{\chi}_{2j}\breve{f}_{js}^{(2)})\breve{u}_{ijs}\breve{u}_{ijt} - \breve{\nu}_{ts,l}\right]^2 \text{ and } \breve{\nu}_{ts,l}^{\mathcal{T}} = s_{ts,l}\left(\hat{\nu}_{ts,l}\right),$$

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

where $s_{ts,l}(z) = sgn(z)\left(|z| - \tau_{ts,l}\right)_+$, $\tau_{ts,l} = C_{2,l}(\underline{m}^{-1}\ln T)^{1/2}(\breve{\theta}_{ts,l})^{1/2}$, and $C_{2,l}$ is a positive constant. We propose to estimate the bias term $\boldsymbol{B}_t$ by $\breve{\boldsymbol{B}}_t = \frac{1}{T}\sum_{s=1}^{T}\breve{\nu}_{ts}^{\mathcal{T}}$.

To study the consistency of the above estimators, we add the following assumption.

Assumption 11.

(i) The process $\{U_{\cdot\cdot t},\ t \geq 1\}$ is covariance-stationary with covariance matrix $\Sigma = \mathbb{E}\left(U_{\cdot\cdot t}U_{\cdot\cdot t}'\right) = \left\{\sigma_{ij,i_1j_1}\right\}$.

(ii) There exists $\gamma \in [0,1)$ such that $\max_{i,j}\sum_{i_1,j_1}\left|\sigma_{ij,i_1j_1}\right|^{\gamma} \leq C$ for some $C > 0$.

(iii) Let $\omega_T = T^{1/(2q)}\underline{m}^{-1/2}\ln T.T^{-1/2+1/(2q)}(N \vee M)^{1/(2q)}(\ln T)^{1/2} \to 0$ and $T^{-1}\omega_T^{1-\gamma}(NM)^{1/2} \to 0$ as $\underline{m} \to \infty$.

Assumption 11(i) is typically assumed in the literature. Assumption 11(ii) strengthens the typical weak cross-sectional dependence condition $\max_{i,j}\sum_{i_1,j_1}\left|\sigma_{ij,i_1j_1}\right| = O(1)$. It is satisfied if u_{ijt}'s satisfy certain m-dependence condition cross-sectionally or the correlation between u_{ijt} and $u_{i_1j_1t}$ vanishes sufficiently fast as the "distance" between (i,j) and (i_1,j_1) increases.

The following theorem reports the consistency of $\breve{\Psi}_t^{(0)}, \breve{\boldsymbol{B}}_t$, and $\breve{\Omega}_0^{-1}\breve{\boldsymbol{B}}_t$.

THEOREM 4.3. *Suppose that Assumption I holds. Suppose Assumptions 1–11 hold. Then,*

(i) $\breve{\Psi}_t^{(0)} = \Psi_t^{(0)} + o_p(1)$,

(ii) $\breve{\boldsymbol{B}}_t = \boldsymbol{B}_t + o_p\left((NM)^{-1/2}\right)$

(iii) $\breve{\Omega}_0^{-1}\breve{\boldsymbol{B}}_t = H^{(0)\prime}\Omega_0^{-1}H^{(0)}\boldsymbol{B}_t + o_p\left((NM)^{-1/2}\right)$.

Given the above results, we can make inferences on the global factors. The procedure is standard and omitted for brevity.

5. MONTE CARLO SIMULATIONS

5.1. DGPs and Implementation

We consider four DGPs, where the true numbers of factors are all specified as $r^{(0)} = 1, r_i^{(1)} = 2$, and $r_j^{(2)} = 1$. DGPs 1 and 2 are ideal cases where all the factors and error terms are i.i.d. random variables. In DGP 1, all factor loadings are generated from $N(0,1)$, while in DGP 2, all factor loadings are drawn from $N(1,1)$. As shown in our theory, whether factor loadings have zero means has important implications for the estimator of the global factors. We purposefully use these two ideal DGPs to illustrate the different impacts of local factor components on the estimators of global factors.

DGPs 3 and 4 are more realistic. In DGP 3, the error terms are both serially-correlated and cross-sectionally dependent. DGP 4 is the most complicated case where local factors are cross-sectionally dependent and the error terms are serially

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

correlated and cross-sectionally dependent. All four DGPs are generated according to (1.1). Below are the details of the four DGPs.

Elements of $(\lambda_{ij}^{(0)\prime},\lambda_{ij}^{(1)\prime},\lambda_{ij}^{(2)\prime})'$ are all i.i.d. $N(0,1)$ random variables in DGP 1. Elements of $(\lambda_{ij}^{(0)\prime},\lambda_{ij}^{(1)\prime},\lambda_{ij}^{(2)\prime})'$ are all i.i.d. $N(1,1)$ random variables in DGPs 2–4. In DGPs 1 and 2, elements in $(f_t^{(0)\prime}, f_{it}^{(1)\prime}, f_{jt}^{(2)\prime}, u_{ijt})$ are all i.i.d. $N(0,1)$ random variables.

In DGP 3, the error terms u_{ijt} is generated as

$$u_{ijt} = \left(u_{1ijt} + u_{2ijt} + u_{3ijt}\right)/\sqrt{3}, \tag{5.1}$$

where $u_{1ijt} = 0.5u_{1i-1,jt} + \sqrt{0.75}e_{1ijt}, u_{2ijt} = 0.5u_{2i,j-1,t} + \sqrt{0.75}e_{2ijt}, u_{3ijt} = 0.5u_{3ij,t-1} + \sqrt{0.75}e_{3ijt}$ and elements of $(e_{1jit}, e_{2jit}, e_{3ijt})$ are all i.i.d. $N(0,1)$ random variables. Here, u_{1ijt}, u_{2ijt}, and u_{3ijt} all follow an AR(1) structure and generate dependence along the i, j, and t dimensions respectively, and the error term u_{ijt} is normalized to have variance 1. Elements in $(f_t^{(0)\prime}, f_{it}^{(1)\prime}, f_{jt}^{(2)\prime})'$ are all i.i.d. $N(0,1)$ random variables.

In DGP 4, $f_{it}^{(1)}$ and $f_{jt}^{(2)}$ are cross-sectionally dependent. Specifically, the two i-specific factors, denoted as $f_{it}^{(1)} \equiv (f_{it,1}^{(1)}, f_{it,2}^{(1)})'$, and the one j-specific factor, $f_{jt}^{(2)}$, are generated, respectively, as

$$f_{it,1}^{(1)} = \left(\sum_{k=i}^{i+1} f_{kt}^{*} + e_{it}^{*}\right)/\sqrt{3},\ f_{it,2}^{(1)} \sim N(0,1) \text{ and } f_{jt}^{(2)} = \left(\sum_{k=j}^{j+1} f_{kt}^{**} + e_{jt}^{**}\right)/\sqrt{3}, \tag{5.2}$$

where $f_{kt}^{*}, f_{kt}^{**}, e_{it}^{*}$, and e_{jt}^{**} are i.i.d. $N(0,1)$ random variables. All the factors are normalized to have variance 1. Here, $f_{it,1}^{(1)}$ and $f_{jt}^{(2)}$ are cross-sectionally dependent.[10] For example, $\mathrm{Cov}(f_{i-1,t,1}^{(1)}, f_{i,t,1}^{(1)}) = \frac{1}{3}$ and $\mathrm{Cov}(f_{j-1,t}^{(2)}, f_{j,t}^{(2)}) = \frac{1}{3}$. Elements in $f_t^{(0)}$ are all i.i.d. $N(0,1)$ random variables. In DGP 4, u_{ijt} is also generated as (5.1).

We consider six combinations of sample sizes: $(N,M,T) = (50,50,50)$, $(50,50,100)$, $(50,100,50)$, $(100,100,50)$, $(50,100,100)$, and $(100,100,100)$. The number of replications is 250. We find that in our simulations, (AH)'s GR estimator performs slightly better than their ER estimator. Therefore, we focus on the GR estimator with $r_{\max} = 8$. We set $C_1 = 0.25$ in the POET estimator and $C_{2,l} = 0.25$ in the bias estimator, both of which are discussed in Section 4.3.

5.2. Simulation Results

Table 1 compares the first-step (initial), second-step (final), and oracle estimators of global factors for DGPs 1 and 2.[11] The oracle estimators are obtained as in the 2D case where the local factor components are absent and the true number

[10] We have also tried serially correlated factors, and find similar results. The detailed results are available upon request.

[11] The results for DGPs 3 and 4 are similar to those for DGP 2 and are omitted to save space.

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

of global factors is known. For convenience, for these two DGPs $f_t^{(0)}$'s are fixed and normalized such that $F^{(0)\prime}F^{(0)}/T = 1$. Therefore, $f_t^{(0)}$ is treated as the true parameter. Table 1 reports the bias, variance and MSEs of the three types of estimators. Note that the bias, variance, and MSEs all depend on t. For ease of reporting, we take the average values over t. We use these two ideal DGPs to confirm some key theoretical findings. First, in DGP 1, our final estimator of the global factors should be *oracle efficient* in the sense that it is asymptotically equivalent to the oracle estimator. In DGP 2, our final estimator is not oracle efficient, and the presence of local factor components should inflate the asymptotic variance of the final estimators of the global factors. Specifically, the "nearly orthogonal" condition of (4.2) in Remark 9 is satisfied in DGP 1, and violated in DGP 2. This theoretical result is supported by our simulations. Although for DGP 1, our final estimator has a larger variance and MSE than the oracle estimator when the sample size is small, the differences become smaller when the sample size increases. For example, when $(N,M,T) = (100,100,100)$, the MSEs of our final estimator and the oracle estimators are 1.08×10^{-4} and 0.99×10^{-4}, respectively. In contrast, for DGP 2, the MSE of our final estimator (1.17×10^{-4}) is larger than that of the oracle estimator (0.50×10^{-4}) even when $(N,M,T) = (100,100,100)$.

Second, although in both DGPs, our final estimators of global factors are $\sqrt{N\bar{M}}$-consistent, the initial estimators behave differently for these two DGPs. Consider the simple case where N, M, and T pass to infinity at the same rate. In DGP 1, the initial estimator of the global factor is $\sqrt{N\bar{M}}$-consistent, while in DGP 2, the initial estimator is only $\min(\sqrt{N},\sqrt{M})$-consistent. In DGP 2, the asymptotic variance of the initial estimator is only of the order of $\left(N^{-1}+M^{-1}\right)$. Our simulation results confirm those theoretical implications. For both DGPs, the MSEs of our final estimator decline with N and M. For DGP 1, the MSE of the initial estimator appears to decrease at the $N\bar{M}$-rate. For example, when $(N,M,T) = (50,50,50)$ and $(N,M,T) = (100,100,100)$, the MSEs of the initial estimators are 31.59×10^{-4} and 7.51×10^{-4}, respectively. In contrast, for DGP 2, the MSE of the initial estimator declines slowly with N and M. For example, when $(N,M,T) = (50,50,50)$ and $(N,M,T) = (100,100,100)$, the MSEs of the initial estimators are 175.36×10^{-4} and 83.88×10^{-4}, respectively.

Third, for both DGPs, we find that the second-step estimators are much more efficient than the first-step estimators. For example, for DGP 1, when $(N,M,T) = (100,100,100)$, the MSEs of the first-step and second-step estimators are 7.51×10^{-4} and 1.08×10^{-4}, respectively. For DGP 2, when $(N,M,T) = (100,100,100)$, the MSEs of the first-step and second-step estimators are vastly different, being 83.88×10^{-4} and 1.17×10^{-4}, respectively.

Table 2 shows the proportions of replications in which the number of factors determined by our method is less than, equal to, and greater than the true number out of total 250 replications. Our method can determine the true number of factors with a correct rate close to 100% for all the sample sizes considered. Table 3 presents the performance of our estimators. In Panel A, we report the coefficient

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

of determination (R^2) for the regression of true factors on the estimated factors. Since the estimated factors should estimate the linear span of the true factors, we expect the R^2 to be high. Indeed, for the global factors, we find that the R^2 is almost 1.00 for all sample sizes under scrutiny. For the two i-specific local factors and one j-specific local factor, the R^2's are all equal to or above 0.95. The general pattern is that when the sample size increases, the performance of our estimators improves. Panel B reports the correlations between the true factor components and the estimated factor components, which are all above 0.95 for all sample sizes.

Table 4 reports the empirical coverage of 95% confidence intervals (CIs) for the global factors based on the theory in Section 4.3. In general, when the sample size is small, we have under-coverage. For example, when $(N,M,T) = (50,50,50)$, the actual coverages for DGPs 1–4 are 90.6%, 88.8%, 88.1%, and 86.5%, respectively. In general, the performance deteriorates when the DGP becomes more and more complicated. The worst performance occurs in DGP 4 where both local factors are cross-sectionally dependent and the error terms are serially correlated and cross-sectionally dependent. However, the performance improves when the sample size increases. For example, when $(N,M,T) = (100,100,100)$, the actual coverages are 96.6%, 95.7%, 94.9%, and 94.2% for DGPs 1–4, respectively.

6. AN EMPIRICAL APPLICATION TO INTERNATIONAL TRADE

We apply our new method to study the international trade flows. Let y_{ijt} denote the log change of the trade flows: $y_{ijt} = \ln\left(Export_{ijt}\right) - \ln\left(Export_{ij,t-1}\right)$, where $Export_{ijt}$ is the trade flow from source country i to destination country j at year t. Here, we consider the growth of trade volume. How to understand trade growth has been a fundamental and long-standing issue in the international trade literature (see, e.g., Baier and Bergstrand, 2001).[12]

6.1. Data

The sample includes 49 source countries and 58 destination countries over 34 years $(1973-2006)$. Thus, $N = 49, M = 58,$ and $T = 34$. With missing values, the total sample size is 46,852. The data are obtained from the companion website of Head and Mayer (2014). The details of the construction of the sample can be found in Section S10 of the Supplementary Material. All the data are demeaned as $y_{ijt} - \frac{1}{T}\sum_{t=1}^{T} y_{ijt}$ for all (i,j) pairs before the estimation.

[12] Here, we consider a simple log-linearized equation for the trade growth. There is some recent literature on nonlinear models for trade data (see, e.g., Santos Silva and Tenreyro, 2006). It will be interesting to extend our method to nonlinear models, but it is beyond the scope of this article.

TABLE 2. Determination of the correct numbers of factors

		Global factors $\hat{r}^{(0)}$			Local i-factors $\hat{r}_i^{(1)}$			Local j-factors $\hat{r}_j^{(2)}$		
DGP	(N, M, T)	< 1	$= 1$	> 1	< 2	$= 2$	> 2	< 1	$= 1$	> 1
	$(50, 50, 50)$	0	1	0	0	.99	.01	0	1	0
1	$(50, 100, 50)$	0	1	0	0	1	0	0	1	0
	$(100, 100, 50)$	0	1	0	0	1	0	0	1	0
	$(50, 50, 100)$	0	1	0	0	.98	.02	0	1	0
	$(50, 100, 100)$	0	1	0	0	1	0	0	1	0
	$(100, 100, 100)$	0	1	0	0	1	0	0	1	0
	$(50, 50, 50)$	0	1	0	0	1	0	0	1	0
2	$(50, 100, 50)$	0	1	0	0	1	0	0	1	0
	$(100, 100, 50)$	0	1	0	0	1	0	0	1	0
	$(50, 50, 100)$	0	1	0	0	1	0	0	1	0
	$(50, 100, 100)$	0	1	0	0	1	0	0	1	0
	$(100, 100, 100)$	0	1	0	0	1	0	0	1	0
	$(50, 50, 50)$	0	1	0	0	.99	0	0	1	0
3	$(50, 100, 50)$	0	1	0	0	1	0	0	1	0
	$(100, 100, 50)$	0	1	0	0	1	0	0	1	0
	$(50, 50, 100)$	0	1	0	0	1	0	0	1	0
	$(50, 100, 100)$	0	1	0	0	1	0	0	1	0
	$(100, 100, 100)$	0	1	0	0	1	0	0	1	0
	$(50, 50, 50)$	0	1	0	0	1	0	0	1	0
4	$(50, 100, 50)$	0	1	0	0	1	0	0	1	0
	$(100, 100, 50)$	0	1	0	0	1	0	0	1	0
	$(50, 50, 100)$	0	1	0	0	1	0	0	1	0
	$(50, 100, 100)$	0	1	0	0	1	0	0	1	0
	$(100, 100, 100)$	0	1	0	0	1	0	0	1	0

Note: Numbers in the main entries are the proportions of replications in which the selected number of factors is less than, equal to, or greater than the true number of factors out of total 250 replications. For $\hat{r}_i^{(1)}$ and $\hat{r}_j^{(2)}$, the numbers are also averaged over i and j, respectively.

6.2. The Number of Factors and Variance Decomposition

Using our method with $r_{\max} = 8$, we find that there is one global factor, i.e., $\hat{r}^{(0)} = 1$. The numbers of estimated source country factors ($\hat{r}_i^{(1)}$) and destination country factors ($\hat{r}_j^{(2)}$) are presented in Figure 1. The median numbers of these two local factors are both 1. Most of the source country factor numbers are either 0 or 1, whereas most of the destination country factor numbers are 1.

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

TABLE 3. Performance of estimated factors and estimated components

		Panel A: coefficient of determination (R^2) for the regression of true factors on the estimated factors				Panel B: correlation between the estimated components and true components (cmpts)		
DGP	(N,M,T)	Global factors	Local i-factors		Local j-factors	Global cmpts	Local i-cmpts	Local j-cmpts
			1st factor	2nd factor				
	(50, 50, 50)	1.00	0.96	0.96	0.95	0.96	0.97	0.97
1	(50, 100, 50)	1.00	0.97	0.97	0.96	0.96	0.97	0.97
	(100, 100, 50)	1.00	0.97	0.97	0.97	0.96	0.97	0.97
	(50, 50, 100)	1.00	0.97	0.97	0.97	0.98	0.98	0.98
	(50, 100, 100)	1.00	0.98	0.98	0.97	0.98	0.98	0.98
	(100, 100, 100)	1.00	0.98	0.98	0.98	0.98	0.98	0.98
	(50, 50, 50)	1.00	0.96	0.96	0.97	0.97	0.98	0.98
2	(50, 100, 50)	1.00	0.97	0.97	0.97	0.97	0.98	0.98
	(100, 100, 50)	1.00	0.97	0.97	0.97	0.97	0.98	0.98
	(50, 50, 100)	1.00	0.97	0.97	0.98	0.98	0.99	0.99
	(50, 100, 100)	1.00	0.98	0.98	0.98	0.98	0.99	0.99
	(100, 100, 100)	1.00	0.98	0.98	0.98	0.98	0.99	0.99
	(50, 50, 50)	1.00	0.96	0.96	0.96	0.97	0.98	0.98
3	(50, 100, 50)	1.00	0.97	0.97	0.96	0.97	0.98	0.98
	(100, 100, 50)	1.00	0.97	0.97	0.97	0.97	0.98	0.98
	(50, 50, 100)	1.00	0.97	0.97	0.97	0.98	0.99	0.99
	(50, 100, 100)	1.00	0.98	0.98	0.98	0.98	0.99	0.99
	(100, 100, 100)	1.00	0.98	0.98	0.98	0.98	0.99	0.99
	(50, 50, 50)	1.00	0.96	0.96	0.96	0.97	0.98	0.98
4	(50, 100, 50)	1.00	0.97	0.97	0.96	0.97	0.98	0.98
	(100, 100, 50)	1.00	0.97	0.97	0.97	0.97	0.98	0.98
	(50, 50, 100)	1.00	0.97	0.97	0.97	0.98	0.99	0.99
	(50, 100, 100)	1.00	0.98	0.98	0.98	0.98	0.99	0.99
	(100, 100, 100)	1.00	0.98	0.98	0.98	0.98	0.99	0.99

Note: Numbers in the main entries of Panel A are the R^2 for the regression of the true factors on the estimated factors averaging over 250 replications. For local i-factors and local j-factors, the numbers are also averaged over i and j, respectively. Numbers in the main entries of Panel B are the correlations between estimated and true factor components averaging over 250 replications. For local i-components and local j-components, the numbers are also averaged over i and j, respectively.

TABLE 4. Empirical coverage of 95% CIs for global factors

(N,M,T)	DGP 1	DGP 2	DGP 3	DGP 4
$(50,50,50)$	90.6%	88.8%	88.1%	86.5%
$(50,100,50)$	91.7%	89.2%	89.1%	87.3%
$(100,100,50)$	94.4%	92.7%	89.9%	88.7%
$(50,50,100)$	90.9%	90.2%	90.0%	88.8%
$(50,100,100)$	94.2%	93.2%	92.5%	91.4%
$(100,100,100)$	96.6%	95.7%	94.9%	94.2%

Note: Numbers in the main entries are the empirical coverage of 95% CIs for global factors based on the inference methods in Section 4.3.

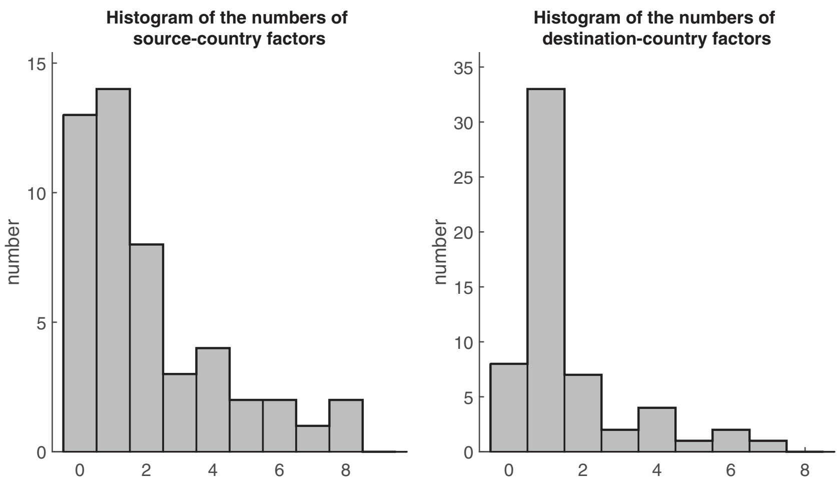

FIGURE 1. The numbers of local factors for the trade application.

We examine how much the sample variance of the trade growth can be explained by global factors and local factors. Table 5 presents the variance decomposition for the whole sample and selected countries. The decomposition for all countries can be found in Section S10 of the Supplementary Material. For the whole sample, we find that global factors can explain about 14.4% of the total sample variance, while the source country factors and destination country factors can explain about 21.0% and 28.1%, respectively. The sample covariance between the two local components is quite small. In total, the factors can explain about 64.2% of the sample variance. We can also conduct the variance decomposition country by country, say, for a given source country (a subsample with the same i index) or for a given destination country (a subsample with the same j index). For example, when China is a source country, the global factors, source country (China) factors, and destination country factors can explain 7.4%, 12.5%, and 43.0% of its sample

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

TABLE 5. Variance decomposition for whole sample and selected countries

		Variance of global components	Variance source country components	Variance destination country components	Total variance explained
Whole sample		14.4%	21.0%	28.1%	64.2%
Source countries	China	7.4%	12.5%	43.0%	63.4%
	Germany	22.3%	12.4%	35.1%	72.5%
	Japan	13.8%	0%	46.2%	60.0%
	U.K.	14.5%	0%	39.1%	53.6%
	U.S.	19.7%	31.6%	18.0%	75.2%
Destination countries	China	6.4%	54.4%	15.0%	77.2%
	Germany	32.8%	17.2%	10.1%	62.4%
	Japan	25.0%	21.1%	15.2%	60.5%
	U.K.	14.8%	17.8%	0%	32.6%
	U.S.	21.9%	14.7%	19.5%	56.2%

variance, respectively. This suggests that the variation of China's export is mainly affected by its destination countries. When China is a destination country, the global factors, source country factors, and destination country (China) factors can explain 6.4%, 54.4%, and 15.0% of its sample variance, respectively. This shows that China's import is mainly driven by its source countries. This decomposition exercise for China seems to suggest that China's trading partners have relatively large impacts on China's international trade flows, while the global factors and its own factors play a relatively small role. For the U.S., we find that the global factors and its own factors are more important than its trading partners' factors. Specifically, when the U.S. is a source country, the global factors, source country (U.S.) factors and destination country factors can explain 19.7%, 31.6%, and 18.4% of its sample variance, respectively. When the U.S. is a destination country, the global factors, source country factors, and destination country (U.S.) factors can explain 21.9%, 14.7%, and 19.5% of its sample variance, respectively.

6.3. Global Factors

The upper panel of Figure 2 plots the estimated global factors (bias-corrected) and its 95% CIs. The bias-corrected and non-bias-corrected estimates are similar in this application. The CIs are narrow, which suggests our estimates are quite precise with small standard errors. The shaded areas in the figure indicate five global recessions during the sample period (1973–2006): 1974–75, 1980–83, 1990–93, 1998, and 2001–02. In general, the estimated factors reflect the global business

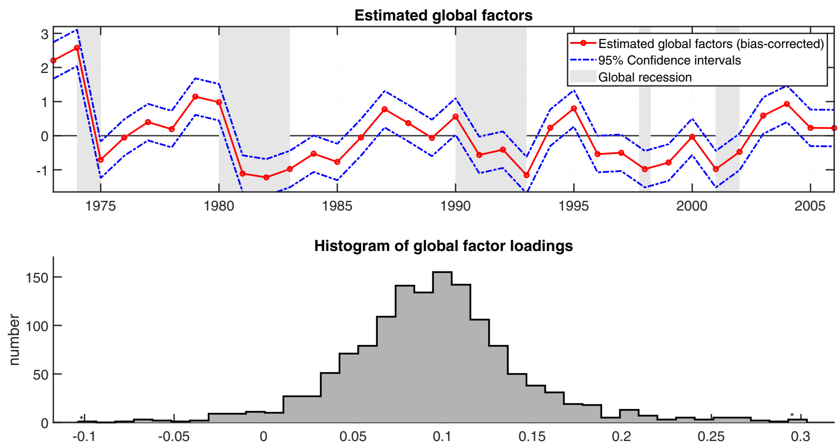

FIGURE 2. Global factor and factor loadings for the trade application.

cycles and have relatively low values during the recession periods. The lowest point of the global factor occurs during the early 1980s recession. The lower panel of Figure 2 shows the histograms of the estimates of global factor loadings. The majority of global factor loadings are positive with a median value of 0.10, which means that approximately a one-standard-deviation increase of the global factor leads to about 10% absolute increase in the growth rate of export.

One natural question is how the estimated global factor is linked to other global economic variables. In Figure 3, we plot the estimated global factors against four economic variables: (*i*) the world economic growth, (*ii*) the lagged world economic growth, (*iii*) the log change of the world openness index, and (*iv*) the log change of the crude oil price (constant price), where openness index is defined as exports and imports of goods and services measured as a share of gross domestic product (GDP). The sources of these data are described in the Section S10.1of the Supplementary Material. For ease of comparison, all the series are normalized to have mean zero and variance one. Apparently, the estimated factors reflect those fundamental economic variables. The correlations between the estimated global factors and these four economic variables are 0.46, 0.59, 0.72, and 0.64, respectively.

6.4. Local Factors

Since there are a large number of local factors, we select two representative countries: China and the U.S. Figure 4 shows the source country factors and destination

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

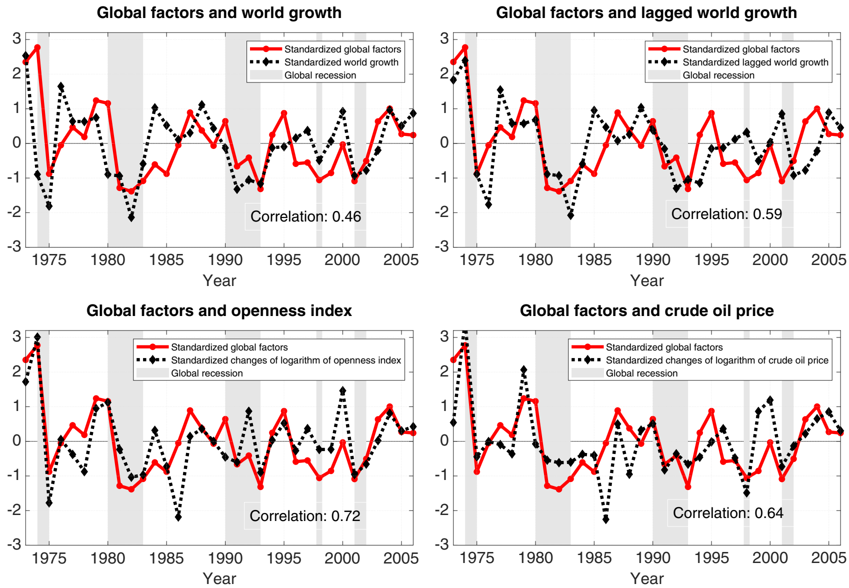

FIGURE 3. Comovement between the (standardized) estimated global factors and some economic variables for the trade application.

country factors for China and the U.S. against their economic growth.[13] We find that for both China and the U.S., the correlations between the source country factors and their economic growth are not high (0.28 and 0.12, respectively). The destination country factors are more closely related to their economic growth, with correlations being 0.44 and 0.58 for China and the U.S., respectively.

As our method allows general pattern of correlations in the local factors, in Figure 5, we plot the histogram of the following four types of sample correlations: (*i*) the cross-sectional correlation between $f_{it}^{(1)}$'s, (*ii*) the cross-sectional correlation between $f_{jt}^{(2)}$'s, (*iii*) the time-series correlation between $f_{it}^{(1)}$'s, and (*iv*) the time-series correlation between $f_{jt}^{(2)}$'s, which are defined, respectively, as,[14]

$$(i)\ \text{the correlation between}\ \left\{f_{i_1,t}^{(1)}\right\}_{t=1}^{T}\ \text{and}\ \left\{f_{i_2,t}^{(1)}\right\}_{t=1}^{T}\ \text{for all pairs of}\ (i_1,i_2), \tag{6.1}$$

[13]For both countries, the numbers of local factors are all determined to be 1 except that there are four source country factors for the U.S., in which case we plot the first factor (the factor corresponding to the largest eigenvalue).

[14]If the number of local factors is larger than one, we use the first factor (the factor corresponding to the largest eigenvalue) to calculate the sample correlations.

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

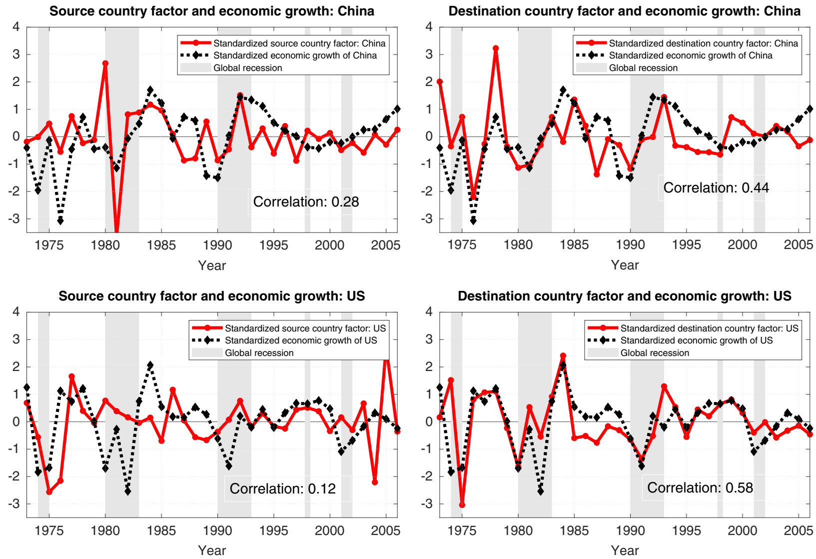

FIGURE 4. (Standardized) local factors and economic growth for the trade application.

$$(ii)\ \text{the correlation between}\ \left\{f^{(2)}_{j_1,t}\right\}_{t=1}^{T}\ \text{and}\ \left\{f^{(2)}_{j_2,t}\right\}_{t=1}^{T}\ \text{for all pairs of}\ (j_1,j_2), \tag{6.2}$$

$$(iv)\ \text{the correlation between}\ \left\{f^{(1)}_{i,t_1}\right\}_{i=1}^{N}\ \text{and}\ \left\{f^{(1)}_{i,t_2}\right\}_{i=1}^{N}\ \text{for all pairs of}\ (t_1,t_2), \tag{6.3}$$

$$(iv)\ \text{the correlation between}\ \left\{f^{(2)}_{j,t_1}\right\}_{j=1}^{M}\ \text{and}\ \left\{f^{(2)}_{j,t_2}\right\}_{j=1}^{M}\ \text{for all pairs of}\ (t_1,t_2). \tag{6.4}$$

In this application, we find that most of the correlations are below 0.5. This suggests that the correlations in the local factors are small.

In sum, we find that global factors, source country factors, and destination country factors are all important for international trade flows, which account for 14.4%, 21.0%, and 28.1% of its sample variance, respectively. We find that there is one global factor and the extracted global factor is closely related to some fundamental global economic variables, such as the lagged world economic growth, the openness index, and the crude oil price. The median numbers of source country factors and destination country factors are both one. The magnitudes of variance decomposition for individual countries vary across countries.

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

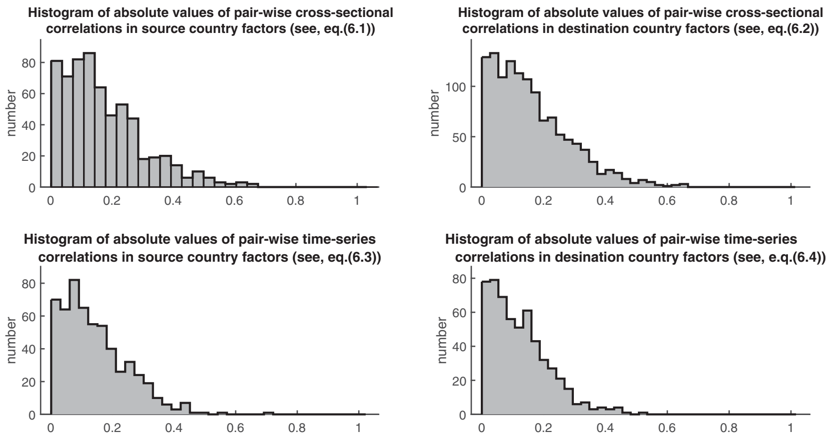

FIGURE 5. Correlations in the estimated local factors for the trade application.

7. CONCLUSION

This article studies the estimation and inference for 3D factor models where there are both global factors and two types of local factors. Such models have potential applications in many economic fields. We determine the number of factors based on the ER or GR statistics of Ahn and Horenstein (2013). Given the number of factors, we propose a two-step estimation method based on PCA, which is easy to implement. We derive the inferential theory for the estimators, including the convergence rates and limiting distributions under general conditions that allow both serial correlation and cross-sectional dependence. The new method is applied to study the international trade flow data.

There are many interesting topics for further research. First, in our framework, we allow correlations within the same type of local factors. In empirical research, we are often interested in the correlations and may want to test whether the correlations are equal to zero. Therefore, we can study the asymptotic properties for the estimators of the correlations and develop a testing procedure. Second, so far, we only consider a pure factor structure. We may extend our model to allow exogenous regressors in the same spirit as in two dimensional panel models with interactive fixed effect models (see, e.g., Bai, 2009). Alternatively, we may allow the factors or factor loadings to depend on observable covariates, as in Fan et al. (2016a) and Fan et al. (2016b). Third, we only consider stationary and strong factors in our model. It is possible to consider nonstationary factors with unit-root-type behavior (e.g., Bai and Ng, 2004), weak factors (e.g., Onatski, 2012) or structural changes (e.g., Cheng, Liao, and Schorfheide, 2016) and develop separate theories. We are exploring these topics in ongoing works.

https://doi.org/10.1017/S0266466625100091 Published online by Cambridge University Press

SUPPLEMENTARY MATERIAL

Sainan Jin, Xun Lu, and Liangjun Su (July 2025): Supplement to “Three-Dimensional Factor Models with Global and Local Factors”, Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466625100091.

REFERENCES

Ahn, S. C., & Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3), 1203–1227.

Anderson, J. E., & van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93(1), 170–192.

Ando, T., & Bai, J. (2017). Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. *Journal of the American Statistical Association*, 112(519), 1182–1198.

Andrade, P., & Zachariadis, M. (2016). Global versus local shocks in micro price dynamics. *Journal of International Economics*, 98, 78–92.

Andreou, E., Gagliardini, P., Ghysels, E., & Rubin, M. (2019). Inference in group factor models with an application to mixed-frequency data. *Econometrica*, 87(4), 1267–1305.

Babii, A., Ghysels, E., & Pan, J. (2024). Tensor principal component analysis. Working Paper.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1), 135–171.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4), 1229–1279.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221.

Bai, J., & Ng, S. (2004). A panic attack on unit roots and cointegration. *Econometrica*, 72(4), 1127–1177.

Bai, J., & Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4), 1133–1150.

Bai, J., & Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1), 18–29.

Bai, J., & Ng, S. (2023). Approximate factor models with weaker loadings. *Journal of Econometrics*, 235(2), 1893–1916.

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

Bai, Z., & Saranadasa, H. (1996). Effect of high dimension: An example of a two sample problem. *Statistica Sinica*, 6(2), 311–329.

Baier, S. L., & Bergstrand, J. H. (2001). The growth of world trade: Tariffs, transport costs, and income similarity. *Journal of International Economics*, 53(1), 1–27.

Bernard, A. B., Dhyne, E., Magerman, G., Manova, K., & Moxnes, A. (2022). The origins of firm heterogeneity: A production network approach. *Journal of Political Economy*, 130(7), 1765–1804.

Breitung, J., & Eickmeier, S. (2016). Analyzing international business and financial cycles using multi-level factor models: A comparison of alternative approaches. In E. Hillebrand & S. J. Koopman (Eds.), *Dynamic factor models* (vol. 35, pp. 177–214). Emerald Group Publishing Limited.

Chen, B., Han, Y., & Yu, Q. (2024). Estimation and inference for CP tensor factor models. arXiv preprint arXiv:2406.17278.

Chen, E. Y., Tsay, R. S., & Chen, R. (2020). Constrained factor models for high-dimensional matrix-variate time series. *Journal of the American Statistical Association*, 115(530), 775–793.

Chen, M. (2023). Circularly projected common factors for grouped data. *Journal of Business & Economic Statistics*, 41(2), 636–649.

Chen, Q., & Fang, Z. (2019). Improved inference on the rank of a matrix. *Quantitative Economics*, 10(4), 1787–1824.

Chen, R., Yang, D., & Zhang, C.-H. (2022). Factor models for high-dimensional tensor time series. *Journal of the American Statistical Association*, 117(537), 94–116.

Chen, S. X., & Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2), 808–835.

Cheng, X., & Hansen, B. E. (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics*, 186(2), 280–293.

Cheng, X., Liao, Z., & Schorfheide, F. (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies*, 83(4), 1511–1543.

Chiang, H. D., Rodrigue, J., & Sasaki, Y. (2023). Post-selection inference in three-dimensional panel data. *Econometric Theory*, 39(3), 623–658.

Choi, I., Kim, D., Kim, Y. J., & Kwark, N.-S. (2018). A multilevel factor model: Identification, asymptotic theory and applications. *Journal of Applied Econometrics*, 33(3), 355–377.

Choi, I., Lin, R., & Shin, Y. (2023). Canonical correlation-based model selection for the multilevel factors. *Journal of Econometrics*, 233(1), 22–44.

Chudik, A., & Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, 188(2), 393–420.

Dhyne, E., Kikkawa, A. K., Mogstad, M., & Tintelnot, F. (2021). Trade and domestic production networks. *The Review of Economic Studies*, 88(2), 643–668.

Dias, F., Pinheiro, M., & Rua, A. (2013). Determining the number of global and country-specific factors in the euro area. *Studies in Nonlinear Dynamics and Econometrics*, 17(5), 573–617.

Fan, J., Ke, Y., & Liao, Y. (2016a). Robust factor models with explanatory proxies. arXiv preprint arXiv:1603.07041.

Fan, J., Liao, Y., & Wang, W. (2016b). Projected principal component analysis in factor models. *The Annals of Statistics*, 44(1), 219–254.

Fan, J., Liao, Y., & Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 75(4), 603–680.

Feenstra, R. C. (2015). *Advanced international trade: Theory and evidence*. Princeton University Press.

Feng, G., Gao, J., Liu, F., & Peng, B. (2024). Estimation and inference for three-dimensional panel data models. arXiv preprint arXiv:2404.08365.

Freeman, H. (2022). Multidimensional interactive fixed-effects. arXiv preprint arXiv:2209.11691.

Gao, Z., & Tsay, R. S. (2023). Divide-and-conquer: A distributed hierarchical factor approach to modeling large-scale time series data. *Journal of the American Statistical Association*, 118(544), 2698–2711.

Giglio, S., & Xiu, D. (2021). Asset pricing with omitted factors. *Journal of Political Economy*, 129(7), 1947–1990.

Han, X. (2021). Shrinkage estimation of factor models with global and group-specific factors. *Journal of Business & Economic Statistics*, 39(1), 1–17.

Han, Y., Yang, D., Zhang, C.-H., & Chen, R. (2024). CP factor model for dynamic tensors. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(5), 1383–1413.

He, Y., Kong, X., Yu, L., Zhang, X., & Zhao, C. (2024). Matrix factor analysis: From least squares to iterative projection. *Journal of Business & Economic Statistics*, 42(1), 322–334.

Head, K., & Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman, & K. Rogoff (Eds.), *Handbook of international economics* (vol. 4, pp. 131–195). Elsevier.

Jin, S., Lu, X., & Su, L. (2025). Three-dimensional heterogeneous panel data models with multi-level interactive fixed effects. *Journal of Econometrics*, 249(2), 105957.

Kapetanios, G., Serlenga, L., & Shin, Y. (2021). Estimation and inference for multi-dimensional heterogeneous panel datasets with hierarchical multi-factor error structure. *Journal of Econometrics*, 220(2), 504–531.

Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3), 501–524.

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1), 97–126.

Koren, M., & Tenreyro, S. (2007). Volatility and development. *The Quarterly Journal of Economics*, 122(1), 243–287.

Latała, R. (2005). Some estimates of norms of random matrices. *Proceedings of the American Mathematical Society*, 133(5), 1273–1282.

Lettau, M. (2022). High-dimensional factor models with an application to mutual fund characteristics. Technical report, National Bureau of Economic Research.

Lu, X., Miao, K., & Su, L. (2021). Determination of different types of fixed effects in three-dimensional panels. *Econometric Reviews*, 40(9), 867–898.

Lu, X., & Su, L. (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics*, 190(1), 148–175.

Ma, S., Lan, W., Su, L., & Tsai, C.-L. (2020). Testing alphas in conditional time-varying factor models with high-dimensional assets. *Journal of Business & Economic Statistics*, 38(1), 214–227.

Matyas, L. (2017). *The econometrics of multi-dimensional panels*. Springer.

Moench, E., & Ng, S. (2011). A hierarchical factor analysis of us housing market dynamics. *Econometrics Journal*, 14(1), 1–24.

Moench, E., Ng, S., & Potter, S. (2013). Dynamic hierarchical factor models. *The Review of Economics and Statistics*, 95(5), 1811–1817.

Moon, H. R., & Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4), 1543–1579.

Moon, H. R., & Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1), 158–195.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4), 1004–1016.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2), 244–258.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4), 967–1012.

Santos Silva, J., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4), 641–658.

Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2), 147–162.

Su, L., & Ju, G. (2018). Identifying latent grouped patterns in panel data models with interactive fixed effects. *Journal of Econometrics*, 206(2), 554–573.

Vershynin, R. (2011). Spectral norm of products of random and deterministic matrices. *Probability Theory and Related Fields*, 150(3–4), 471–509.

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In Y. C. Eldar & G. Kutyniok (Eds.), *Compressed sensing, theory and applications* (pp. 210–268). Cambridge University Press.

Wang, D., Liu, X., & Chen, R. (2019). Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics*, 208(1), 231–248.

Wang, P. (2014). Large dimensional factor models with a multi-level factor structure: identification, estimation and inference. Technical report, Hong Kong University of Science and Technology.

Yang, Y., & Schmidt, P. (2021). An econometric approach to the estimation of multi-level models. *Journal of Econometrics*, 220(2), 532–543.

Yu, L., He, Y., Kong, X., & Zhang, X. (2022). Projected estimation for large-dimensional matrix factor models. *Journal of Econometrics*, 229(1), 201–217.