Econometric Theory, **40**, 2024, 419–446.
doi:10.1017/S0266466622000366

TWO-STEP ESTIMATION OF QUANTILE PANEL DATA MODELS WITH INTERACTIVE FIXED EFFECTS

Liang Chen
Peking University HSBC Business School

This paper considers the estimation of panel data models with interactive fixed effects where the idiosyncratic errors are subject to conditional quantile restrictions. An easy-to-implement two-step estimator is proposed for the coefficients of the observed regressors. In the first step, the principal component analysis is applied to the cross-sectional averages of the regressors to estimate the latent factors. In the second step, the smoothed quantile regression is used to estimate the coefficients of the observed regressors and the factor loadings jointly. The consistency and asymptotic normality of the estimator are established under large N,T asymptotics. It is found that the asymptotic distribution of the estimator suffers from asymptotic biases, and this paper shows how to correct the biases using both analytical and split-panel jackknife bias corrections. Simulation studies confirm that the proposed estimator performs well with moderate sample sizes.

1. INTRODUCTION

This paper considers panel data models with interactive fixed effects, where the unobserved errors have a latent factor model structure. The assumption of interactive fixed effects has been adopted in a lot of recent studies—see Pesaran (2006), Bai (2009), Moon and Weidner (2015), and Lu and Su (2016), among many others. This assumption is general enough to nest the standard panel data models with only individual fixed effects and models with additive individual and time effects. It also allows the unobserved factors (or common shocks) to affect the dependent variables with different intensities that are measured by the individual-specific factor loadings. Moreover, the latent factor structure has become an important tool to characterize cross-sectional dependence in panel data models—see Chudik and Pesaran (2015) for an excellent review. Yet most of the existing studies focus on linear models where the idiosyncratic errors are subject to conditional mean restrictions, and the main object of interest is the coefficients that represent the partial effect of the regressors on the conditional mean of the dependent variable. This paper focuses on panel data models with interactive

I am indebted to Liangjun Su and three anonymous referees for their constructive inputs which have greatly improved the paper. Financial support from the National Natural Science Foundation of China (Grant No. 71703089) is gratefully acknowledged. Address correspondence to Liang Chen, Peking University HSBC Business School, No. 2199 Lishui Road, Shenzhen, Guangdong 518055, China; e-mail: chenliang@phbs.pku.edu.cn.

effects where the conditional mean restrictions are replaced by conditional quantile restrictions. In such models, the coefficients of the regressors measure the partial quantile effect, providing a more complete picture of how the regressors affect the distributions of the dependent variables.

This paper adopts the popular common correlated effects (CCEs) framework pioneered by Pesaran (2006). In this framework, the regressors are assumed to be driven by the same latent factors that affect the dependent variables, allowing the space of the common factors to be approximated by the cross-sectional averages of the observed variables. Compared with the approach that estimates the coefficients and fixed effects jointly, the CCE approach has two main advantages that are particular valuable for the quantile panel models: first, the computation of the CCE estimator is easy, because given the estimated factors, the coefficients of the regressors and the factor loadings can be simply estimated by treating the estimated factors as known. Second, the asymptotic properties of the estimators are much easier to derive since the estimated factors have a relatively simple expansion.

Like the CCE estimator, the proposed estimation method in this paper contains two steps. However, both of the steps differ from the standard CCE method that is widely used for linear and quantile panel data models in existing studies. In the first step, to avoid the *degenerated-regressors* problem of the standard CCE method (see Karabiyik, Reese, and Westerlund, 2017 and Remark 1), the principal component analysis (PCA) is applied to the cross-sectional averages of the regressors to estimate the common factors. In the second step, inspired by Galvao and Kato (2016), the smoothed quantile regression (SQR) instead of the standard quantile regression is used to estimate the coefficients of the regressors and the factor loadings jointly, treating the estimated factors from the first step as given. The main motivation of making such modifications in both steps of the standard CCE estimator is to facilitate the asymptotic analysis of the proposed estimator.

In the "large N, small T" framework,[1] the identification and estimation of quantile panel data models are very challenging even when there are only individual effects (see Arellano and Bonhomme, 2016 and Graham et al., 2018 for examples). When there are interactive effects in quantile panel models, there remains the open question of whether the parameter of interest can be point identified (see Chen, 2015 for a result of set identification). Thus, this paper follows Fernández-Val and Weidner (2016) and Chen, Fernández-Val, and Weidner (2021b) and considers the "large N, large T" framework where the realizations of the factors and factor loadings are treated as nonrandom fixed parameters, and the main contribution of this paper is that I establish the asymptotic properties of the proposed estimator in the context of quantile panel models. In particular, under some regularity conditions, it is shown that the proposed two-step estimator for the coefficients of the regressors is $\sqrt{NT}$-consistent and asymptotically normal with a leading

[1] Throughout this paper, N and T denote the numbers of cross-sectional and time series observations, respectively.

bias term of order $T^{-1}+N^{-1}$. More importantly, the analytical expression of the leading bias is derived, providing the basis of analytical bias correction and a heuristic justification for the use of the split-panel jackknife (SPJ) bias correction in practice. The Bahadur representation of my two-step estimator extends the similar representations of the estimators for linear panel data models (see Bai, 2009) and nonlinear panel data models with smooth objective functions (see Chen, 2016; Chen et al., 2021b) to quantile panel models. To the best of my knowledge, this is the first result of this kind in the literature.

1.1. Related Literature

This paper is related to the large and growing literature on quantile regressions for panel models. Abrevaya and Dahl (2008), Rosen (2012), Arellano and Bonhomme (2016), Cai, Chen, and Fang (2018), and Graham et al. (2018) considered identification and estimation of quantile effects with fixed T. In the large T framework, Canay (2011) and Chen and Huo (2021) proposed two-step estimation methods, Koenker (2004), Galvao and Montes-Rojas (2010), and Lamarche (2010) proposed penalized quantile regressions for panel models, Galvao (2011) considered quantile regressions of dynamic panels, Kato, Galvao, and Montes-Rojas (2012), Galvao and Kato (2016), and Galvao, Gu, and Volgushev (2020) focused on the asymptotic distributions of quantile regressions and SQRs, Galvao, Lamarche, and Lima (2013) studied censored quantile regressions for panel data, Yoon and Galvao (2020) considered the robust estimation of the covariance matrix, and Chen (2019) studied the nonparametric estimation of quantile panel models.

All the studies mentioned above only considered models with individual effects. Quantile panel models with interactive fixed effects were first studied by Harding and Lamarche (2014), and more recently by Belloni, Chen, and Padilla (2019), Feng (2019), Ando and Bai (2020), Harding, Lamarche, and Pesaran (2020), Chen, Dolado, and Gonzalo (2021a), and Ma, Linton, and Gao (2021).

As in this paper, Harding and Lamarche (2014) and Harding et al. (2020) also adopted the CCE framework. However, unlike my two-step estimator, they proposed to use the standard CCE estimator where the cross-sectional averages of the regressors and the dependent variables are used as the proxies of the unobserved factors, and in the second step, they use standard quantile regressions instead of SQR to estimate the coefficients of the regressors. More importantly, their asymptotic results are quite different in nature from the main conclusion of this paper. In particular, Harding and Lamarche (2014) showed that the CCE estimator has no asymptotic bias, whereas Harding et al. (2020) proved that the CCE estimator of the common slope parameter is $\sqrt{NT}$-consistent, with a leading bias term of approximate order $T^{-3/4}$, but they did not give the analytical expression of the bias.[2]

[2]Harding et al. (2020) also considered heterogenous slopes and showed that the CCE estimators are $\sqrt{N}$-consistent.

Chen et al. (2021a) and Ando and Bai (2020) both proposed an iterative procedure to estimate the quantile factors and factor loadings jointly. Chen et al.'s (2021a) model has no observed regressors, and they mainly focused on the asymptotic properties of the estimated factors and factor loadings. The model of Ando and Bai (2020) contains observed regressors, but they assumed the coefficients to be heterogenous across individuals. As a consequence, their estimators of the heterogenous coefficients converge at the rate of $\sqrt{N}$ and are free of asymptotic biases. Moreover, Ando and Bai's (2020) asymptotic analysis requires all the finite moments of the idiosyncratic errors to be bounded, whereas in the current paper, I only need the density functions of the idiosyncratic errors to exist and to be sufficiently smooth (i.e., continuously differentiable). Ma et al. (2021) considered a model that is similar to the quantile factor models of Chen et al. (2021a) except that they assumed the factor loadings to be smooth functions of observed (and time-invariant) individual characteristics.

One potential problem of the methods proposed by Chen et al. (2021a) and Ando and Bai (2020) is that their computational algorithm does not necessarily converge to the global minimum because their objective function is not convex. To solve this problem, Belloni et al. (2019) and Feng (2019) added to the objective function a nuclear-norm penalty term that is widely used in the matrix completion literature, resulting in a new objective function that is convex in the parameters. However, the convergence rates of their estimators are much slower than $\sqrt{NT}$ in general due to the regularization bias, and the asymptotic distributions of their estimators are not derived.

1.2. Structure of the Paper

The rest of the paper is organized as follows. Section 2 introduces the model and the new two-step estimator. Second 3 establishes the consistency and the asymptotic distribution of the estimator, and discusses how to correct the asymptotic bias, how to estimate the asymptotic covariance matrix, and how to choose the tuning parameters in practice. In Section 4, Monte Carlo simulations are used to evaluate the finite sample performance of the proposed estimator and the effectiveness of the bias correction methods. Finally, Section 5 concludes. To save space, the proofs of all the theorems are relegated to the Supplementary Material.

1.3. Notations

Throughout the paper, $\mathsf{Q}_Y[\tau|X=x]$ denotes the conditional τ-quantile of Y given $X=x$, $\|A\|$ denotes the Frobenius norm of matrix A, and $\mathsf{Tr}(\cdot)$ denotes the trace of a square matrix. For two sequences of nondecreasing real numbers $\{a_j\}$ and $\{b_j\}$, $a_j \asymp b_j$ means that there exists $0 < c_1 < c_2 < \infty$ such that $c_1 < a_j/b_j < c_2$ for all large j.

2. THE MODEL AND THE ESTIMATOR

2.1. The Model

For some $\tau \in (0,1)$, consider the model:

$$Y_{it} = \beta_0(\tau)'X_{it} + \lambda_i(\tau)'f_t + u_{it}, \quad \text{for } i = 1,\ldots,N; t = 1,\ldots,T, \tag{1}$$

where $(Y_{it}, X_{it}) \in \mathbb{R} \times \mathbb{R}^p$ is the vector of observed variables for individual i at time t, and $\lambda_i(\tau) \in \mathbb{R}^r$ and $f_t \in \mathbb{R}^r$ are the unobserved factor loadings (or individual effects) and common factors (or time effects), respectively. The idiosyncratic error u_{it} is assumed to satisfy the following conditional quantile restriction almost surely:

$$\mathsf{Q}_{u_{it}}[\tau | X_{it}, \lambda_i(\tau), f_t] = 0. \tag{2}$$

Given the above restriction, we have $\mathsf{Q}_{Y_{it}}[\tau | X_{it}, \lambda_i(\tau), f_t] = \beta_0(\tau)'X_{it} + \lambda_i(\tau)'f_t$. Thus, the main object of interest of this paper is $\beta_0(\tau)$, i.e., the marginal quantile effect of the regressors X_{it} conditional on the factors and factor loadings.

The above model encompasses the location-scale-shift model and random coefficients model as special cases. First, consider a location-scale-shift model: $Y_{it} = \beta_0' X_{it} + \lambda_i' f_t + (\gamma' X_{it} + \theta' f_t)\epsilon_{it}$. If $\gamma' X_{it} + \theta' f_t > 0$ and ϵ_{it} is independent of (X_{it}, λ_i, f_t), then $\beta_0(\tau) = \beta_0 + \gamma \mathsf{Q}_{\epsilon_{it}}(\tau)$, $\lambda_i(\tau) = \lambda_i + \theta \mathsf{Q}_{\epsilon_{it}}(\tau)$, and $u_{it} = (\gamma' X_{it} + \theta' f_t)(\epsilon_{it} - \mathsf{Q}_{\epsilon_{it}}(\tau))$. Second, consider a random coefficients model: $Y_{it} = \beta_0(\epsilon_{it})' X_{it} + \lambda_i(\epsilon_{it})' f_t$, where $\epsilon_{it} | (X_{it}, f_t) \sim \mathsf{U}[0,1]$ and the mapping $\tau \mapsto \beta_0(\tau)' X_{it} + \lambda_i(\tau)' f_t$ is strictly increasing, then (1) and (2) hold with $u_{it} = (\beta_0(\epsilon_{it}) - \beta_0(\tau))' X_{it} + (\lambda_i(\epsilon_{it}) - \lambda_i(\tau))' f_t$.

In addition, following the literature on CCE estimation of panel data models (see Pesaran, 2006; Karabiyik et al., 2017), the regressors are assumed to be driven by the common factors f_t, i.e., the dynamics of X_{it} is captured by the following factor model structure:

$$X_{it} = \Gamma_i f_t + e_{it}, \quad \text{for } i = 1,\ldots,N; t = 1,\ldots,T, \tag{3}$$

where $\Gamma_i \in \mathbb{R}^{p \times r}$ is a matrix of constants, and $e_{it} \in \mathbb{R}^p$ is a vector of random errors.

The main reason for adopting the CCE framework in this paper is that it allows us to estimate $\beta_0(\tau)$ in a simple two-step procedure that will be defined below. The benefits of employing the two-step estimation approach are twofold: first, under some standard assumptions, the factors can be consistently estimated in the first step using the regressors, which greatly simplifies the asymptotic analysis of the estimator in the second step; second, the low computational cost of the two-step estimator makes it appealing to empirical researchers.

In comparison, in an alternative framework where the relationship between the regressors and the factors is left unspecified (such as Bai, 2009 and Ando and Bai, 2020), the coefficients for the regressors, the factors, and the factor loadings are usually estimated jointly. On the one hand, such "joint estimators" are computationally intensive since they involve iterations between the factors and the factor loadings. On the other hand, the asymptotic properties of such "joint estimators" are much more difficult to establish in the context of quantile regressions. Ando and

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

Bai (2020) consider heterogeneous panels where the estimator of each individual's coefficient converges at $\sqrt{N}$ rate. In a homogeneous panel, the estimator for the coefficients converges at the much faster $\sqrt{NT}$ rate, making it much more challenging to derive the asymptotic distribution of the estimator because many higher-order terms that have been ignored in Ando and Bai's (2020) analysis will become relevant. Moreover, note that the asymptotic analysis of Chen et al. (2021b) for nonlinear panel data models with interactive fixed effects, which is already very involved, does not apply to these "joint estimators" since the parameters in the quantile models are defined through nonsmooth moment conditions.[3]

2.2. The Two-Step Estimator

For the moment, assume that the number of factors r is known (Section 3.1 discusses how to consistently estimate r) and that $p \geq r$. Define $\bar{X}_t = N^{-1}\sum_{i=1}^{N} X_{it}$, and $\hat{\Sigma}_{\bar{X}} = T^{-1}\sum_{t=1}^{T}\bar{X}_t\bar{X}_t'$. Moreover, let $K(z) = 1 - \int_{-1}^{z} k(x)dx$, where $k(\cdot)$ is a symmetric continuous kernel function with support $[-1,1]$ and h is a bandwidth parameter. Then the two-step estimator of $\beta_0(\tau)$ is defined as follows:

Step 1: $\hat{f}_t = \hat{\Psi}'\bar{X}_t$, where $\hat{\Psi} \in \mathbb{R}^{p\times r}$ is the matrix of eigenvectors associated with largest r eigenvalues of $\hat{\Sigma}_{\bar{X}}$.

Step 2: $\hat{\beta}(\tau)$ is defined as

$$[\hat{\beta}(\tau), \hat{\Lambda}(\tau)] = \underset{\beta\in\mathcal{B},\lambda_i\in\mathcal{A}}{\arg\min} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\tau - K\left(\frac{Y_{it}-\beta' X_{it}-\lambda_i'\hat{f}_t}{h}\right)\right](Y_{it}-\beta' X_{it}-\lambda_i'\hat{f}_t), \tag{4}$$

where $\hat{\Lambda}(\tau) = [\hat{\lambda}_1(\tau),\ldots,\hat{\lambda}_N(\tau)]'$.

Define $l(u) = (\tau - K(u/h))u$ and $L(\beta,\Lambda) = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} l(Y_{it}-\beta' X_{it} - \lambda_i'\hat{f}_t)$. Step 2 of the estimation procedure can be effectively solved by the gradient descent algorithm as follows:

Step 2.1: Choose the initial value of the parameter: $(\beta^{(0)}, \Lambda^{(0)})$.

Step 2.2: For $j = 0$, set $s_j = 1$; for $j \geq 1$, define $L_j^{\beta} = -(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} l^{(1)}(Y_{it} - \beta^{(j)\prime}X_{it} - \lambda_i^{(j)\prime}\hat{f}_t)X_{it}$, $L_j^{\lambda_i} = -(NT)^{-1}\sum_{t=1}^{T} l^{(1)}(Y_{it} - \beta^{(j)\prime}X_{it} - \lambda_i^{(j)\prime}\hat{f}_t)\hat{f}_t$, where $l^{(1)}(u) = \partial l(u)/\partial u$. Set[4]

$$s_j = \frac{\left|(\beta^{(j)}-\beta^{(j-1)})'(L_j^{\beta}-L_{j-1}^{\beta}) + \sum_{i=1}^{N}(\lambda_i^{(j)}-\lambda_i^{(j-1)})'(L_j^{\lambda_i}-L_{j-1}^{\lambda_i})\right|}{\left\|L_j^{\beta}-L_{j-1}^{\beta}\right\|^2 + \sum_{i=1}^{N}\left\|L_j^{\lambda_i}-L_{j-1}^{\lambda_i}\right\|^2}.$$

[3] One can smooth the objective function in quantile regressions as I do in this paper, but some important assumptions of Chen et al. (2021b) (such as the boundness of the derivatives of the objective function) cannot be satisfied by the smoothed check function.

[4] This method of choosing the step size is known as the Barzilai–Borwein method.

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

Step 2.3: Update the parameters by

$$\beta^{(j+1)} = \beta^{(j)} - s_j \cdot L_j^{\beta} \quad \text{and} \quad \lambda_i^{(j+1)} = \lambda_i^{(j)} - s_j \cdot L_j^{\lambda_i}.$$

Step 2.4: Iterate Steps 2.2 and 2.3 until the objective function converges.

Since the objective function $L(\beta, \Lambda)$ is not convex in (β, Λ), there is no guarantee that the gradient descent algorithm above is able to find the global minimum. Thus, choosing a good initial value for the parameter is essential. In practice, I recommend using the following estimator as the initial value of the parameter:

$$[\tilde{\beta}(\tau), \tilde{\Lambda}(\tau)] = \underset{\beta \in \mathcal{B}, \lambda_i \in \mathcal{A}}{\arg\min} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau (Y_{it} - \beta' X_{it} - \lambda_i' \hat{f}_t), \tag{5}$$

where $\rho_\tau = (\tau - \mathbf{1}\{u \leq 0\})u$ is the check function. Under Assumptions 1 and 2 of the next section, it can be shown that $\tilde{\beta}(\tau)$ is a consistent estimator of $\beta_0(\tau)$.[5]

Remark 1. The way I estimate the unknown factors in Step 1 is different from the standard CCE method that uses $\bar{X}_t$ and $\bar{Y}_t = N^{-1} \sum_{i=1}^{N} Y_{it}$ as the proxies of f_t. A problem with the CCE approach, as pointed out by Karabiyik et al. (2017), is that the second moment matrix of the estimated factors is asymptotically singular when $p+1 > r$, known as the problem of "degenerated regressors." This problem results in two possible complications for nonlinear panel data models: first, the asymptotic property of the CCE estimator is more challenging to establish and there might be extra biases due to the degenerated regressors (see Theorem 3 of Karabiyik et al., 2017); second, since the nonlinear models usually require nonlinear optimization algorithms to obtain the estimator, it is difficult to find the (local) minimum with degenerated regressors. My approach avoids this problem because it will be shown in the next section that the second moment matrix of the estimated factors is asymptotically full rank as long as $p \geq r$.

Remark 2. A natural question that arises is why not just use the estimator given in (5) in Step 2. The main reason is that it is difficult to work out the analytical expression of the asymptotic bias of $\tilde{\beta}(\tau)$ due to the nonsmoothness of the check function—see Kato et al. (2012) for a detailed discussion. The use of SQR in Step 2 is inspired by Galvao and Kato (2016), who derived the asymptotic bias of the fixed effects estimator for quantile panel data models with only individual effects. Similar ideas have been explored by Amemiya (1982) and Horowitz (1998), but for different objectives.

Remark 3. At $\tau = 0.5$, models (1)–(3) can be viewed as a variant of the model of Pesaran (2006) where the assumption that u_{it} has conditional mean 0 is replaced by the assumption that u_{it} has conditional median 0. Accordingly, the proposed

[5]The proof of this claim is essentially the same as the proof of Theorem 1, and it is therefore omitted. In fact, the consistency of $\tilde{\beta}(\tau)$ does not require $\mathcal{B}$ to be compact thanks to the convexity of the check function—see Kato et al. (2012).

two-step estimator at $\tau = 0.5$ can be viewed as the least absolute deviation (LAD) counterpart of the CCE estimator. As will be shown below, the advantage of the LAD estimator is that only restrictions on the conditional density of u_{it} are needed to establish its consistency and asymptotic normality, making it more robust to outliers and heavy-tailed distributions. The robustness of the proposed estimator against heavy-tailed distributions is examined through Monte Carlo simulations in Section 4.2.

3. ASYMPTOTIC RESULTS

Suppose that we have a panel of observations $\{(Y_{it}, X_{it}), i = 1, \ldots, N, t = 1, \ldots, T\}$ generated from (1) and (3), where the realized values of the factors and factor loadings are $F_0 = [f_{01}, \ldots, f_{0T}]'$ and $\Lambda_0(\tau) = [\lambda_{01}(\tau), \ldots, \lambda_{0N}(\tau)]'$. In this section, following the literature on nonlinear panel data models, $\Lambda_0(\tau)$ and F_0 are treated as fixed parameters. Thus, given $\Lambda_0(\tau)$ and F_0, my model can be written as

$$Y_{it} = \beta_0(\tau)' X_{it} + \lambda_{0i}(\tau)' f_{0t} + u_{it}, \quad \mathsf{Q}_{u_{it}}[\tau | X_{it}] = 0, \quad \text{and} \quad X_{it} = \Gamma_i f_{0t} + e_{it}.$$

Alternatively, all the assumptions and asymptotic results in this section can be understood as being conditional on $\Lambda(\tau) = \Lambda_0(\tau)$ and $F = F_0$ (see Remark 6). Moreover, to simplify the notations, the dependence of $\lambda_{0i}(\tau)$ on τ is suppressed throughout this section.

3.1. The Number of Factors

In the previous section, r is assumed to be known, which is rarely the case in most empirical applications. Thus, this subsection considers the estimation of r.

Note that

$$\hat{\Sigma}_{\bar{X}} = \frac{1}{T}\sum_{t=1}^{T} \bar{e}_t \bar{e}_t' + \bar{\Gamma} \cdot \frac{1}{T}\sum_{t=1}^{T} f_{0t} \bar{e}_t' + \frac{1}{T}\sum_{t=1}^{T} \bar{e}_t f_{0t}' \bar{\Gamma}' + \bar{\Gamma} \hat{\Sigma}_{f_0} \bar{\Gamma}',$$

where $\hat{\Sigma}_{f_0} = T^{-1} \sum_{t=1}^{T} f_{0t} f_{0t}'$, $\bar{\Gamma} = N^{-1} \sum_{i=1}^{N} \Gamma_i$, and $\bar{e}_t = N^{-1} \sum_{i=1}^{N} e_{it}$. If $\{e_{it}, i = 1, \ldots, N\}$ are weakly dependent for each t, the first three terms on the right-hand side of the above equation can be shown to be $o_P(1)$. Moreover, if both $\bar{\Gamma}$ and $\hat{\Sigma}_{f_0}$ have full rank, then $\hat{\Sigma}_{\bar{X}}$ converges in probability to a matrix with rank r. This observation motivates the following estimator of r.

Let $\hat{\rho}_1 \geq \hat{\rho}_2 \geq \cdots \geq \hat{\rho}_p$ be the eigenvalues of $\hat{\Sigma}_{\bar{X}}$, and let $\mathbb{P}_{NT}$ be a sequence of nonnegative constants. Then the estimator of r is defined as

$$\hat{r} = \sum_{j=1}^{p} \mathbf{1}\{\hat{\rho}_j > \mathbb{P}_{NT}\}.$$

In order to prove the consistency of $\hat{r}$, the following conditions are imposed.

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

Assumption 1. Let $M > 0$ be a generic bounded constant.

(i) $p \geq r$.
(ii) $\|f_{0t}\| \leq M$ for all t. There exists $\Sigma_{f_0} \in \mathbb{R}^{r\times r}$ and $\Gamma_0 \in \mathbb{R}^{p\times r}$ such that $\|\hat{\Sigma}_{f_0} - \Sigma_{f_0}\| = O(T^{-1/2})$, $\|\bar{\Gamma} - \Gamma_0\| = O(N^{-1/2})$, and $\text{rank}(\Sigma_{f_0}) = \text{rank}(\Gamma_0) = r$.
(iii) $\mathbb{E}[e_{it}] = 0$ for i,t, and $\mathbb{E}\|N^{-1/2}\sum_{i=1}^{N} e_{it}\|^2 \leq M$ for all t.

The conditions that $p \geq r$ and $\text{rank}(\Gamma_0) = r$ are standard in the literature of CCE estimation to ensure that the space of the common factors can be approximated by the cross-sectional averages of the regressors. Condition (ii) implies that $\Gamma_0\Sigma_{f_0}\Gamma_0'$ has full rank. It is also worth noting that only weak cross-sectional correlations of e_{it} are required through condition (iii), and the serial correlations of e_{it} are left unrestricted.

Then it can be shown that:

Proposition 1. *Under Assumption 1, $P[\hat{r} = r] \to 1$ as $N,T \to \infty$ if $\mathbb{P}_{NT} \to 0$ and $\mathbb{P}_{NT} \cdot \min(\sqrt{N},\sqrt{T}) \to \infty$.*

Given the above result, the number of factors r can be treated as known in the subsequent analysis regarding the asymptotic properties of $\hat{\beta}(\tau)$ (see footnote 5 of Bai, 2003).

3.2. Consistency

Let $\Psi_0 \in \mathbb{R}^{p\times r}$ be the matrix of eigenvectors associated with the r distinct positive eigenvalues of $\Gamma_0\Sigma_{f_0}\Gamma_0'$, and define $\mathbf{H}_0 = \Psi_0'\Gamma_0$, $\tilde{f}_{0t} = \mathbf{H}_0 f_{0t}$, and $\tilde{\lambda}_{0i} = (\mathbf{H}_0')^{-1}\lambda_{0i}$. Note that $\mathbf{H}_0$ is a full rank matrix.[6] In addition, define $V_{it} = [X_{it}', \tilde{f}_{0t}']'$, let $\mathsf{f}_{it}(\cdot|x)$ denote the conditional density of u_{it} given $X_{it} = x$, and let $\varrho_{i,T}$ denote the smallest eigenvalue of $T^{-1}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})V_{it}V_{it}']$.

To derive the consistency of the estimator, the following conditions are imposed in addition to Assumption 1.

Assumption 2. Let $M > 0$ be a generic bounded constant, and let $m \geq 1$ be a positive integer.

(i) The r positive eigenvalues of $\Gamma_0\Sigma_{f_0}\Gamma_0'$ are distinct.
(ii) $\beta_0(\tau) \in \mathcal{B}$, $\tilde{\lambda}_{0i} \in \mathcal{A}$ for all i, and $\mathcal{A}, \mathcal{B}$ are compact.
(iii) There exists $\underline{\varrho} > 0$ such that $N^{-1}\sum_{i=1}^{N}\varrho_{i,T} > \underline{\varrho}$ for all large N and T. $\mathsf{f}_{it}^{(1)}(c|x) = \partial \mathsf{f}_{it}(c|x)/\partial c$ exists and $\max_{i,t}|\mathsf{f}_{it}^{(1)}(c|x)| < M$ uniformly over (c,x).
(iv) For each i, the sequence $\{(X_{it},u_{it}), i = 1,\ldots,N\}$ is α-mixing with coefficients $\alpha_i(j)$ satisfying that $\max_{1\leq i\leq N}\alpha_i(j) \leq M\cdot\alpha^j$ for some $0 < \alpha < 1$.
(v) There exists $\gamma > 0$ such that $\mathbb{E}\|X_{it}\|^{2m+\gamma} < M$ for all i,t.
(vi) As $N,T \to \infty$, $h \to 0$ and $N/T^m \to 0$.

[6] See the proof of Lemma 1.

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

Before presenting the consistency result, I briefly comment on the conditions in Assumption 2.

Condition (i) allows the use of perturbation theory for the eigenvectors of $\Gamma_0 \Sigma_{f_0} \Gamma_0'$, which is important for the result that $\hat{f}_t$ converges to $\mathbf{H}_0 f_{0t}$. A similar condition has been imposed in the study of PCA estimators for approximate factor models (see Assumption G of Bai, 2003).

Condition (ii) requires the parameter spaces to be compact. The compactness of $\mathcal{B}$ is needed because the smoothed check function is no longer convex in (β, λ_i) given (X_{it}, f_t), and the compactness of $\mathcal{A}$ helps to bound the impact of the estimation errors of $\hat{f}_t$ on the objective function.

Condition (iii) is similar to the standard identification condition in quantile regressions. The main difference here is that the common factors need to be taken into account. Note that it allows $\varrho_{i,T}$, the smallest eigenvalue of $T^{-1} \sum_{t=1}^{T} \mathbb{E}[\mathsf{f}_{it}(0|X_{it}) V_{it} V_{it}']$, to be 0 for some i as long as $N^{-1} \sum_{i=1}^{N} \varrho_{i,T}$ is bounded below by a positive constant. But it will fail if $X_i = [X_{i1}, \ldots, X_{iT}]'$ and $F_0 = [f_{01}, \ldots, f_{0t}]'$ span the same space for all i, e.g., $e_{it} = 0$ for all i, t.

Condition (iv) is also standard in the literature (see Assumption D.1 of Kato et al., 2012). The strong mixing condition is used to derive moments bounds in order to apply law of large numbers and central limit theorems. It is commonly employed in nonlinear panel data models because the mixing property is nicely preserved by nonlinear transformations. However, stationarity is not imposed here because for the factor loadings to be quantile dependent, u_{it} should be allowed to depend on the factors. Therefore, conditional on f_{0t}, it becomes necessary to allow the distribution of u_{it} to change across t. Moreover, conditional on f_{0t}, the mean of X_{it} is given by $\Gamma_i f_{0t}$, and thus the distribution of X_{it} should also be time-dependent.

Conditions (v) and (vi) reflect a trade-off between the moments of X_{it} and the required relative size of T compared to N. The existence of higher moments of X_{it} allows for less restrictive conditions on the size of T. In particular, if $m = 1$ and $\mathbb{E}\|X_{it}\|^{2+\gamma} < M$, we need $N/T \to 0$—a very strong condition that is hard to satisfy in most empirical applications. However, if $m = 2$ and $\mathbb{E}\|X_{it}\|^{4+\gamma} < M$, only $N/T^2 \to 0$ is needed. Moreover, if it is assumed that $\|X_{it}\| \leq M$ for all i, t almost surely, condition (vi) can be relaxed to $\log N/\sqrt{T} \to 0$ (see Proposition 3.1 of Galvao and Kato, 2016).

In addition to the above conditions imposed in Assumption 2, it is worth mentioning that the cross-sectional dependence of (X_{it}, u_{it}) is not explicitly restricted. While the cross-sectional dependence of X_{it} is implicitly controlled by Assumption 1(iii), no such restriction is needed for u_{it}. The intuition is that in the large-T asymptotic framework, given $\{f_{0t}\}$, $\beta_0(\tau)$ can be consistently estimated from the observations of any individual i. Thus, for consistency, I only need weak dependence of u_{it} on the time dimension, and the weak cross-sectional dependence of X_{it} is only needed to ensure that the space of $\{f_{0t}\}$ can be well approximated by $\{\hat{f}_t\}$.

Last but not least, note that unlike Ando and Bai (2020), there is no moment restrictions on u_{it}, making the proposed estimation procedure robust to outliers and heavy-tailed distributions. Moreover, compared with the procedures that estimate the factors and factor loadings jointly (such as Chen et al., 2021a), rank condition on the factor loading matrix is not needed, which means that some of the factor loadings in model (1) can be 0. In other words, there can be some factors that affect X_{it} but not Y_{it}.

The following theorem establishes the consistency of $\hat{\beta}(\tau)$ for any given $\tau \in (0,1)$.

THEOREM 1. *Under Assumptions 1 and 2, $\hat{\beta}(\tau)$ is weakly consistent, i.e., $\|\hat{\beta}(\tau) - \beta_0(\tau)\| = o_P(1)$ for any $\tau \in (0,1)$.*

3.3. Asymptotic Distribution

Let $\mathsf{f}_{it}(\cdot)$ be the density function of u_{it}, and $\mathsf{f}_{i,ts}(\cdot,\cdot|x_{it},x_{is})$ be the joint density of (u_{it},u_{is}) given $(X_{it},X_{is}) = (x_{it},x_{is})$. Moreover, let $\mathsf{f}_{it}^{(j)}(c) = \partial^j \mathsf{f}_{it}(c)/\partial c^j$, $\mathsf{f}_{it}^{(j)}(c|x_{it}) = \partial^j \mathsf{f}_{it}(c|x_{it})/\partial c^j$, and $\mathsf{f}_{i,ts}^{(j,k)}(c_1,c_2|x_{it},x_{is}) = \partial^{j+k}\mathsf{f}_{i,ts}(c_1,c_2|x_{it},x_{is})/\partial c_1^j \partial c_2^k$. In particular, let $\mathsf{f}_{it}^{(0)}(c) = \mathsf{f}_{it}(c)$ and $\mathsf{f}_{it}^{(0)}(c|x_{it}) = \mathsf{f}_{it}(c|x_{it})$.

In addition, define

$$\underbrace{\Xi_i}_{p\times r} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})X_{it}]f_{0t}', \quad \underbrace{\Omega_i}_{r\times r} = \frac{1}{T}\sum_{t=1}^{T}\mathsf{f}_{it}(0)f_{0t}f_{0t}', \quad \underbrace{\Phi_i}_{p\times r} = \Xi_i\Omega_i^{-1},$$

$$\underbrace{Z_{it}}_{p\times 1} = X_{it} - \Xi_i\Omega_i^{-1}f_{0t}, \quad \underbrace{\Delta}_{p\times p} = \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}Z_{it}'].$$

To derive the asymptotic distribution of $\hat{\beta}(\tau)$, the following conditions are imposed.

Assumption 3. Let $M > 0$ a generic bounded constant, and let $q \geq 8$ be an even integer.

(i) $\beta_0(\tau)$ is an interior point of $\mathcal{B}$, and $\{\tilde{\lambda}_{01},\ldots,\tilde{\lambda}_{0N}\}$ are interior points of $\mathcal{A}$.
(ii) $\{\Omega_i, i = 1,\ldots,N\}$ are all invertible for large T, and Δ is invertible.
(iii) $\max_{i,t}\mathbb{E}\|X_{it}\|^{2m+\gamma} < M$ for some $\gamma > 0$ and $m > \max\{3q(p+r)/4, 3/(1-6c)\}$ where c is defined below in (vii).
(iv) Define $X_i^T = (X_{i1},\ldots,X_{iT})$ and $u_i^T = (u_{i1},\ldots,u_{iT})$. $\{(X_i^T,u_i^T), i = 1,\ldots,T\}$ are independent across i.
(v) $\mathsf{f}_{it}(c|x_{it})$ is q times continuously differentiable with respect to c, and $\mathsf{f}_{i,ts}(c_1,c_2|x_{it},x_{is})$ is q times continuously differentiable with respect to (c_1,c_2); $\left|\mathsf{f}_{it}^{(j)}(c|x_{it})\right| \leq M$ uniformly over (c,x_{it}) for all $j = 0,\ldots,q$; $\left|\mathsf{f}_{i,ts}^{(j,0)}(c_1,c_2|x_{it},x_{is})\right| \leq M$ and $\left|\mathsf{f}_{i,ts}^{(0,j)}(c_1,c_2|x_{it},x_{is})\right| \leq M$ uniformly over (c_1,c_2,x_{it},x_{is}), for all $j = 0,\ldots,q$.

(vi) $\int_{-1}^{1} k(u)du = 1$, $\int_{-1}^{1} k(u)u^j du = 0$, for $j = 1, \ldots, q-1$, and $\int_{-1}^{1} k(u)u^q du \neq 0$.
(vii) $N/T \to \kappa^2 > 0$ as $N, T \to \infty$. $h \asymp T^{-c}$ and $1/q < c < 1/6$.

Remark 4. The conditions of Assumption 3 are very similar to the assumptions imposed in Galvao and Kato (2016). Thus, the readers are referred to Galvao and Kato (2016) for the details of these conditions. The main difference is that Galvao and Kato (2016) require $q \geq 4$ and $1/q < c < 1/3$, whereas I need the stronger conditions that $q \geq 8$ and $1/q < c < 1/6$. More specifically, due to the presence of the interactive effects, Lemma B.2 of Galvao and Kato (2016) cannot be used to show that the remaining term in the expansion of $\hat{\beta}(\tau) - \beta_0(\tau)$ is $o_P(T^{-1})$. Instead, to bound the higher-order terms, I combine the uniform convergence rates of $\hat{\lambda}_i$ and $\hat{f}_t$ and the fact that the third-order derivative of the objective function is uniformly bounded (up to a positive constant) by $1/h^2$—this is why a much larger h and therefore a much smaller c are needed in the current paper.

Next, define

$$\underbrace{\mathbf{A}_t}_{p\times r} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}]\lambda_{0i}', \qquad \underbrace{\mathbf{B}_{t,k}}_{r\times r} = \frac{1}{N}\sum_{i=1}^{N} \mathsf{f}_{it}(0)\lambda_{0i}\Phi_{i,k},$$

$$\underbrace{\mathbf{C}_{i,k}}_{r\times r} = -\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[\mathsf{f}_{it}^{(1)}(0|X_{it})Z_{it,k}]f_{0t}f_{0t}', \qquad \underbrace{\mathbf{D}_{t,k}}_{r\times r} = -\frac{1}{N}\sum_{i=1}^{N} \mathbb{E}[\mathsf{f}_{it}^{(1)}(0|X_{it})Z_{it,k}]\lambda_{0i}\lambda_{0i}'.$$

Some extra conditions are needed to make sure that the asymptotic biases of $\hat{\beta}(\tau)$ are well defined.

Assumption 4. Define

$$\omega_{T,i}^{(1)} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\big[\mathsf{f}_{it}(0|X_{it})Z_{it}\big] f_{0t}'\Omega_i^{-1} f_{0t},$$

$$\omega_{T,i}^{(2)} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s\neq t}^{T} \left(\tau\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}] - \mathbb{E}\left[\int_{\infty}^{0} \mathsf{f}_{i,ts}(0,v|X_{it},X_{is})dv \cdot Z_{it}\right]\right) f_{0t}'\Omega_i^{-1} f_{0s},$$

$$\omega_{T,i,k}^{(3)} = \tau(1-\tau)\cdot\frac{1}{T}\sum_{t=1}^{T} f_{0t}'\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1} f_{0t},$$

$$\omega_{T,i,k}^{(4)} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s\neq t}^{T} \left\{\mathbb{E}[\mathbf{1}\{u_{it}\leq 0, u_{is}\leq 0\}] - \tau^2\right\} f_{0t}'\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1} f_{0s},$$

and assume that the following limits exist:

$$b_1 = -(\tau - 0.5)\cdot \lim_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\omega_{T,i}^{(1)} - \lim_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\omega_{T,i}^{(2)},$$

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

$$b_{2,k} = 0.5 \lim_{N,T\to\infty} \frac{1}{N}\sum_{i=1}^{N} \omega_{T,i,k}^{(3)} + 0.5 \lim_{N,T\to\infty} \frac{1}{N}\sum_{i=1}^{N} \omega_{T,i,k}^{(4)},$$

$$d_1 = -\lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}\lambda'_{0i}(\mathbf{H}_0)^{-1}\Psi'_0 e_{it}],$$

$$d_{2,k} = 0.5 \lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \mathsf{Tr}\left\{\mathbb{E}[e_{it}e'_{it}]\cdot \Psi_0(\mathbf{H}'_0)^{-1}\left(2\mathbf{B}_{t,k}+\mathbf{D}_{t,k}\right)(\mathbf{H}_0)^{-1}\Psi'_0\right\}.$$

The following theorem gives the asymptotic distribution of $\hat{\beta}(\tau)$.

THEOREM 2. *Suppose that Assumptions 1– 4 hold, then as $N,T\to\infty$,*

$$\sqrt{NT}\left[\hat{\beta}(\tau)-\beta_0(\tau)\right] \xrightarrow{d} \mathcal{N}(\Delta^{-1}(\kappa b+\kappa^{-1}d), \Delta^{-1}\mathbf{V}\Delta^{-1}),$$

where $b = b_1 + b_2$, $d = d_1 + d_2$, $b_2 = [b_{2,1},\dots b_{2,p}]'$, $d_2 = [d_{2,1},\dots d_{2,p}]'$, $\mathbf{V} = \mathbf{V}_1+\mathbf{V}_2$,

$$\mathbf{V}_1 = \lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[W_{it}W'_{it}], \quad \mathbf{V}_2 = \lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\mathbb{E}[W_{it}W'_{is}],$$

and

$$W_{it} = [\tau - \mathbf{1}\{u_{it}\leq 0\}]Z_{it} - \mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi'_0 e_{it}.$$

Remark 5. In the proof of Theorem 2, I show that $\hat{\beta}(\tau)-\beta_0(\tau)$ has the following Bahadur representation:

$$\Delta(\hat{\beta}(\tau)-\beta_0(\tau)) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} W^*_{it} + \frac{b}{T} + \frac{d}{N} + o_P(T^{-1}),$$

where $W^*_{it} = l^{(1)}(u_{it})Z_{it} - \mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi'_0 e_{it}$, and $l^{(1)}(u) = \tau - K(u/h) + k(u/h)u/h$. First, note that $l^{(1)}(u_{it})$ is a smooth approximation of $\tau - \mathbf{1}\{u_{it}\leq 0\}$, and their differences disappear in the limit when calculating the variance–covariance matrix of W^*_{it}. This is why W_{it} instead of W^*_{it} appears in the definition of $\mathbf{V}$. Second, when deriving the limiting distribution of the estimator, I use a strategy that is similar to the one used by Galvao and Kato (2016). In particular, Lyapunov's central limit theorem is applied to $N^{-1/2}\sum_{i=1}^{N}\bar{W}^*_i$, where $\bar{W}^*_i = T^{-1/2}\sum_{t=1}^{T} W^*_{it}$ and $\bar{W}^*_1,\dots,\bar{W}^*_N$ are independent by Assumption 3(iv). Third, the first bias term b/T is caused by the estimation of Λ_0 and the second bias term d/N originates from the estimation of F_0. In nonlinear (probit, logit) panel data models with interactive effects, Chen et al. (2021b) is the first to establish a similar Bahadur representation for the fixed-effects estimator. Similar to Theorem 2, the biases of their estimator, which are generally nonzero except for some special cases, arise from the estimation of the fixed effects. This is in contrast to linear

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

panel data models with interactive effects, where the fixed-effects estimator of the slope parameter has a similar Bahadur representation (see Theorem 3 of Bai, 2009), but the bias term b/T is due to cross-sectional correlation and heteroskedasticity and the bias term d/T is caused by serial correlation and heteroskedasticity.

Remark 6. Suppose that $\Lambda_0(\tau)$ and F_0 are the realizations of the random fixed effects: $\Lambda(\tau) = (\lambda_1(\tau),\ldots,\lambda_N(\tau))'$ and $F = (f_1,\ldots,f_T)'$, then the analysis and asymptotic results in this paper can be viewed as being conditional on $(\Lambda(\tau),F) = (\Lambda_0(\tau),F_0)$. Alternatively, one can condition on the sigma-algebra generated by $\Lambda(\tau)$ and F (denoted by $\mathcal{D}$), and use the conditional central limit theorem (see Rao, 2009) to derive the asymptotic distribution of the estimator. In this case, the α-mixing condition in Assumption 2(iv) should be replaced by a conditional α-mixing condition with coefficients that depend on $\mathcal{D}$ (see Assumption A.2 of Su and Chen, 2013 for example), and Assumption 3(iv) should be replaced by conditional independence.

3.3.1. *Some Special Cases.*

(a) Observed factors

First, in some applications, the common factors are observed (e.g., inflation rate). In this case, there is no need to estimate F_0 from the first step. As a consequence, the asymptotic distribution of $\hat{\beta}(\tau)$ will not be affected by the estimation errors of the factors. In particular, it can be shown that $d_1 = d_2 = 0$ and that $W_{it} = [\tau - \mathbf{1}\{u_{it} \leq 0\}]Z_{it}$. Thus,

$$\sqrt{NT}\left[\hat{\beta}(\tau) - \beta_0(\tau)\right] \xrightarrow{d} \mathcal{N}(\kappa\Delta^{-1}b, \Delta^{-1}\mathbf{V}\Delta^{-1}),$$

where

$$\mathbf{V} = \tau(1-\tau)\cdot \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[Z_{it}Z_{it}'] + \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T} \mathbb{E}\left[\left(\tau^2 - \mathsf{F}_{it}(0|X_{it},X_{is}) - \mathsf{F}_{is}(0|X_{it},X_{is}) + \mathsf{F}_{i,ts}(0,0|X_{it},X_{is})\right)Z_{it}Z_{is}'\right],$$

and $\mathsf{F}_{it}(0|X_{it},X_{is}) = \mathbb{E}[\mathbf{1}\{u_{it} \leq 0\}|X_{it},X_{is}]$, $\mathsf{F}_{i,ts}(0,0|X_{it},X_{is}) = \mathbb{E}[\mathbf{1}\{u_{it} \leq 0, u_{is} \leq 0\}|X_{it},X_{is}]$.

(b) Only individual effects

If we further assume that $r = 1$, $f_{0t} = 1$ for all t, and $\{(X_{it},u_{it}), t = 1,\ldots,T\}$ is stationary for each i, then $\Xi_i = \mathbb{E}[\mathsf{f}_i(0|X_{it})X_{it}]$, $\Omega_i = \mathsf{f}_i(0)$, $Z_{it} = X_{it} -$

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

$\mathbb{E}[\mathsf{f}_i(0|X_{it})X_{it}]/\mathsf{f}_i(0)$,

$$\omega_{T,i}^{(1)} = \mathbb{E}\big[\mathsf{f}_{it}(0|X_{it})Z_{it}\big]/\mathsf{f}_i(0) = 0,$$

$$\omega_{T,i}^{(2)} = \sum_{1\le|k|\le T-1}\left(1-\frac{|k|}{T}\right)\left(\tau\mathbb{E}[\mathsf{f}_i(0|X_{it})Z_{it}] - \mathbb{E}\left[\int_{\infty}^{0}\mathsf{f}_{i,t,t+k}(0,v|X_{it},X_{i,t+k})dv\cdot Z_{it}\right]\right)/\mathsf{f}_i(0),$$

$$\omega_{T,i}^{(3)} = -\frac{\tau(1-\tau)}{\mathsf{f}_i(0)^2}\mathbb{E}[\mathsf{f}_i^{(1)}(0|X_{it})Z_{it}],$$

$$\omega_{T,i}^{(4)} = -\sum_{1\le|k|\le T-1}\left(1-\frac{|k|}{T}\right)\left\{\mathbb{E}[\mathbf{1}\{u_{it}\le 0, u_{i,t+k}\le 0\}]-\tau^2\right\}\cdot\mathbb{E}[\mathsf{f}_i^{(1)}(0|X_{it})Z_{it}]/\mathsf{f}_i(0)^2.$$

Therefore, the asymptotic distribution of $\hat{\beta}(\tau)-\beta_0(\tau)$ is identical to the one given by Theorem 3.2 of Galvao and Kato (2016).

(c) When there is no time series dependence

When there is no time series dependence, i.e., $\{(X_{it},u_{it}), t=1,\ldots,T\}$ is independent across t for each i, it is easy to see that $\omega_{T,i}^{(2)}=0$, $\omega_{T,i,k}^{(4)}=0$, and $\mathbf{V}_2=0$. Thus, we have

$$b_1 = -(\tau-0.5)\cdot\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\big[\mathsf{f}_{it}(0|X_{it})Z_{it}\big]f_{0t}'\Omega_i^{-1}f_{0t},$$

$$b_{2,k} = \frac{\tau(1-\tau)}{2}\cdot\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}f_{0t}'\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1}f_{0t},$$

and

$$\mathbf{V} = \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[W_{it}W_{it}'] = \tau(1-\tau)\cdot\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[Z_{it}Z_{it}'] + \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi_0'\mathbb{E}[e_{it}e_{it}']\Psi_0(\mathbf{H}_0')^{-1}\mathbf{A}_t'.$$

3.4. Bias Correction

3.4.1. *Analytical Bias Correction.* Theorem 2 provides the basis of analytical bias correction for $\hat{\beta}(\tau)$. Suppose that $\hat{\Delta}$, $\hat{b}$, and $\hat{d}$ are consistent estimators of

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

Δ, b, d, respectively, and define

$$\hat{\beta}_{abc}(\tau) = \hat{\beta}(\tau) - \hat{\Delta}^{-1}\left(\frac{\hat{b}}{T} + \frac{\hat{d}}{N}\right).$$

Then it follows easily from Theorem 2 that the bias corrected estimator $\hat{\beta}_{abc}(\tau)$ will have an asymptotic normal distribution that is centered around 0, i.e.,

$$\sqrt{NT}\left[\hat{\beta}_{abc}(\tau) - \beta_0(\tau)\right] \xrightarrow{d} \mathcal{N}(0, \Delta^{-1}\mathbf{V}\Delta^{-1}). \tag{6}$$

To construct consistent estimators of Δ, b, d, let $\{\hat{e}_{it}\}$ be the ordinary least squares residuals of regressing $\{X_{it}\}$ on $\{\hat{f}_t\}$, define $l(u) = (\tau - K(u/h))u$, $l^{(j)}(u) = \partial^j l(u)/\partial u^j$, $\hat{u}_{it} = Y_{it} - \hat{\beta}(\tau)' X_{it} - \hat{\lambda}_i' \hat{f}_t$, and

$$\hat{\Xi}_i = \frac{1}{T}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it}) X_{it}\hat{f}_t', \quad \hat{\Omega}_i = \frac{1}{T}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})\hat{f}_t\hat{f}_t', \quad \hat{\Phi}_i = \hat{\Xi}_i\hat{\Omega}_i^{-1},$$

$$\hat{Z}_{it} = X_{it} - \hat{\Xi}_i\hat{\Omega}_i^{-1}\hat{f}_t, \quad \hat{\Delta} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\hat{Z}_{it}',$$

$$\hat{\mathbf{B}}_{t,k} = \frac{1}{N}\sum_{i=1}^{N} l^{(2)}(\hat{u}_{it})\hat{\lambda}_i\hat{\Phi}_{i,k}, \quad \hat{\mathbf{C}}_{i,k} = \frac{1}{T}\sum_{t=1}^{T} l^{(3)}(\hat{u}_{it})\hat{Z}_{it,k}\hat{f}_t\hat{f}_t',$$

$$\hat{\mathbf{D}}_{t,k} = \frac{1}{N}\sum_{i=1}^{N} l^{(3)}(\hat{u}_{it})\hat{Z}_{it,k}\hat{\lambda}_i\hat{\lambda}_i'.$$

$$\hat{\omega}_{T,i}^{(1)} = \frac{1}{T}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\cdot\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_t, \quad \hat{\omega}_{T,i,k}^{(3)} = \tau(1-\tau)\cdot\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t'\hat{\Omega}_i^{-1}\hat{\mathbf{C}}_{i,k}\hat{\Omega}_i^{-1}\hat{f}_t,$$

$$\hat{\omega}_{T,i}^{(2)} = \frac{1}{T}\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L} l^{(2)}(\hat{u}_{it})\hat{Z}_{it} l^{(1)}(\hat{u}_{is})\cdot\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_s + \frac{1}{T}\sum_{t=L+1}^{T}\sum_{s=t-L}^{t-1} l^{(2)}(\hat{u}_{it})\hat{Z}_{it} l^{(1)}(\hat{u}_{is})\cdot\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_s,$$

$$\hat{\omega}_{T,i,k}^{(4)} = \frac{1}{T}\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L} l^{(1)}(\hat{u}_{it}) l^{(1)}(\hat{u}_{is})\hat{f}_t'\hat{\Omega}_i^{-1}\hat{\mathbf{C}}_{i,k}\hat{\Omega}_i^{-1}\hat{f}_s + \frac{1}{T}\sum_{t=L+1}^{T}\sum_{s=t-L}^{t-1} l^{(1)}(\hat{u}_{it}) l^{(1)}(\hat{u}_{is})\hat{f}_t'\hat{\Omega}_i^{-1}\hat{\mathbf{C}}_{i,k}\hat{\Omega}_i^{-1}\hat{f}_s,$$

$$\hat{b}_1 = -(\tau - 0.5)\cdot\frac{1}{N}\sum_{i=1}^{N}\hat{\omega}_{T,i}^{(1)} - \frac{1}{N}\sum_{i=1}^{N}\hat{\omega}_{T,i}^{(2)}, \quad \hat{b}_{2,k} = 0.5\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\omega}_{T,i,k}^{(3)} + \hat{\omega}_{T,i,k}^{(4)}\right),$$

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

$$\hat{d}_1 = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\hat{\lambda}_i'\hat{\Psi}'\hat{e}_{it},$$

$$\hat{d}_{2,k} = 0.5\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}'\hat{\Psi}\left(2\hat{\mathbf{B}}_{t,k}+\hat{\mathbf{D}}_{t,k}\right)\hat{\Psi}'\hat{e}_{it}.$$

Given the above definitions, the estimators for b and d are given by $\hat{b} = \hat{b}_1 + \hat{b}_2$ and $\hat{d} = \hat{d}_1 + \hat{d}_2$, respectively, where $\hat{b}_2 = [\hat{b}_{2,1},\ldots,\hat{b}_{2,p}]'$ and $\hat{d}_2 = [\hat{d}_{2,1},\ldots,\hat{d}_{2,p}]'$. The following result confirms the validity of the proposed analytical bias correction.

THEOREM 3. *Let c be the constant defined in Assumption 3(vii). Then, under Assumptions 1–4, $\hat{\Delta} = \Delta + o_P(1)$, $\hat{b} = b + o_P(1)$, $\hat{d} = d + o_P(1)$, and therefore (6) holds if $L \to \infty$ and $L \cdot T^{\frac{1}{2m}-0.5+3c} \to 0$ as $N,T \to \infty$.*

3.4.2. *Jackknife Bias Correction.* Following Dhaene and Jochmans (2015), Fernández-Val and Weidner (2016), and Chen et al. (2021b), an alternative method to correct the leading bias of $\hat{\beta}(\tau)$ is the SPJ.

For a given τ, let $\hat{\beta}^{(1)}_{N,T/2}(\tau)$ be the two-step estimator, defined as in (4), using the subsample $i = 1,\ldots,N; t = 1,\ldots,T/2$, and let $\hat{\beta}^{(2)}_{N,T/2}(\tau)$ be the two-step estimator using the subsample $i = 1,\ldots,N; t = T/2+1,\ldots,T$. Similarly, define $\hat{\beta}^{(1)}_{N/2,T}(\tau)$ as the two-step estimator using the subsample $i = 1,\ldots,N/2; t = 1,\ldots,T$, and $\hat{\beta}^{(2)}_{N/2,T}(\tau)$ as the two-step estimator using the subsample $i = N/2+1,\ldots,N$ and $t = 1,\ldots,T$. Then the bias-corrected estimator using the SPJ is defined as

$$\hat{\beta}_{spj}(\tau) = 3\hat{\beta}(\tau) - \frac{1}{2}\left[\hat{\beta}^{(1)}_{N,T/2}(\tau)+\hat{\beta}^{(2)}_{N,T/2}(\tau)\right] - \frac{1}{2}\left[\hat{\beta}^{(1)}_{N/2,T}(\tau)+\hat{\beta}^{(2)}_{N/2,T}(\tau)\right]. \quad (7)$$

The computation of this estimator is almost as easy as the original two-step estimator $\hat{\beta}(\tau)$.

The main intuition of the SPJ estimator is that if the underlying distributions of the data are stable across i and t, the term $0.5(\hat{\beta}^{(1)}_{N,T/2}(\tau)+\hat{\beta}^{(2)}_{N,T/2}(\tau)) - \hat{\beta}(\tau)$ is a good estimate of b/T, and the term $0.5(\hat{\beta}^{(1)}_{N/2,T}(\tau)+\hat{\beta}^{(2)}_{N/2,T}(\tau)) - \hat{\beta}(\tau)$ is a good estimate of d/N. In models with only individual effects, the asymptotic bias of the fixed-effects estimator is determined by the distribution of (X_{it}, u_{it}). Thus, the formal justification of the SPJ only requires the sequence $\{(X_{it}, u_{it}), t = 1,2,\ldots\}$ to be stationary for each i (see Dhaene and Jochmans, 2015; Galvao and Kato, 2016). However, in models with interactive effects, the asymptotic biases are also affected by Λ_0 and F_0. Thus, to justify the use of the SPJ, we also need some kinds of conditions to ensure that the distributions of $f_1,\ldots,f_T$ are stable across t and that the distributions of $\lambda_1,\ldots,\lambda_N$ are stable across i. On the one hand, such assumptions involve the unconditional distributions of Λ and F; on the other hand, the asymptotic theory of this paper is established conditional on Λ_0 and F_0 (realizations of Λ and F)—this gap makes it difficult to rigorously prove the validity of the SPJ estimator. This important but challenging question is left for future research, and the finite sample performance of the SPJ estimator is evaluated in the next section using Monte Carlo simulations.

3.5. Estimating the Variance

The previous subsection gives a consistent estimator of Δ. Thus, it remains to construct a consistent estimator of $\mathbf{V}$. Define

$$\hat{\mathbf{A}}_t = \frac{1}{N}\sum_{i=1}^{N} l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\hat{\lambda}_i', \quad \hat{W}_{it} = l^{(1)}(\hat{u}_{it})\hat{Z}_{it} - \hat{\mathbf{A}}_t\hat{\Psi}'\hat{e}_{it}, \quad \hat{\mathbf{V}}_1 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{W}_{it}\hat{W}_{it}',$$

$$\hat{\mathbf{V}}_2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L}\hat{W}_{it}\hat{W}_{is}' + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1+L}^{T}\sum_{s=t-L}^{t-1}\hat{W}_{it}\hat{W}_{is}',$$

and $\hat{\mathbf{V}} = \hat{\mathbf{V}}_1 + \hat{\mathbf{V}}_2$. The following result establishes the consistency of $\hat{\mathbf{V}}$.

THEOREM 4. *Let L satisfy the condition of Theorem 3. Then, under Assumptions 1– 4,* $\hat{\mathbf{V}} = \mathbf{V} + o_P(1)$.

3.6. The Choice of Tuning Parameters in Practice

The implementation of the proposed estimation procedure in practice involves choosing the kernel function $k(\cdot)$, the bandwidth parameter h, and the truncation parameter L in the estimators of the biases and variance.

First, Assumption 3 requires $k(\cdot)$ to be (at least) an eighth-order kernel function. Thus, the following kernel function of Muller (1984) is recommended[7]:

$$k(z) = \mathbf{1}\{|z| \leq 1\} \cdot \frac{3,465}{8,192}\left(7 - 105z^2 + 462z^4 - 858z^6 + 715z^8 - 221z^{10}\right).$$

Second, if one chooses the eighth-order kernel function above, Assumption 3 requires that $h \asymp T^{-c}$ and $1/8 < c < 1/6$. Thus, when N is about the size of T in practice, a possible choice is $h = 1.5(NT)^{-1/14}$, which is the one used in all the simulations in the next section. Note that even when there are only individual effects, the optimal bandwidth choice in SQR still remains an open question (see Galvao and Kato, 2016). Thus, I would like to leave this important but challenging question for future research.

Finally, the choice of L in finite samples is a more delicate issue, and it seems that there is no consensus in the literature regarding this choice. For example, Hahn and Kuersteiner (2011) and Galvao and Kato (2016) recommended $L = 1$ as a rule of thumb, whereas Fernández-Val and Weidner (2016) suggested to conduct a sensitivity analysis starting from $L = 0$. Moreover, as pointed out by Galvao and Kato (2016), the standard theory for the HAC estimator of covariance matrix in

[7] Both Horowitz (1998) and Galvao and Kato (2016) used the fourth-order kernel of Muller (1984). Higher-order Gaussian kernels are not recommended because they have unbounded support.

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

models with smooth objective functions does not apply to the quantile panel data models. In the spirit of Fernández-Val and Weidner (2016), Section 4.3 reports simulation results with $L = 1, \ldots, 4$ for models with moderate serial correlations. Based on the simulation results, I suggest the practitioners report estimation results with different choices of L, but values of L greater than 4 are not recommended for datasets with moderate number of time series observations ($T \leq 200$).

4. FINITE SAMPLE PERFORMANCE

To evaluate the finite sample performance of the proposed estimators, the following data generating process (DGP) is employed:

$$Y_{it} = \beta_1 X_{it,1} + \beta_2 X_{it,2} + \beta_3 X_{it,3} + \alpha_i + \gamma_i f_t + X_{it,1} \cdot \epsilon_{it},$$

where $[\beta_1, \beta_2, \beta_3] = [1,1,1]$, $\alpha_i \sim$ i.i.d. $\mathcal{N}(0,1)$, $\gamma_i \sim$ i.i.d. $\mathcal{N}(0,1)$, $f_t \sim$ i.i.d. $\mathcal{N}(0,1)$, $X_{it,1} \sim$ i.i.d. $\chi^2(1)+1$, and $X_{it,2} = \theta_{2i} + \eta_{2i} f_t + e_{2,it}$, $X_{it,3} = \theta_{3i} + \eta_{3i} f_t + e_{3,it}$, where $\theta_{2i}, \theta_{3i}, \eta_{2i}, \eta_{3i} \sim$ i.i.d. $\mathcal{N}(1,1)$. Since the asymptotic results are conditional on the fixed effects, only $X_{it,1}, e_{2,it}, e_{3,it}, \epsilon_{it}$ vary across repetitions. The distributions of $e_{2,it}, e_{3,it}, \epsilon_{it}$ are specified in each subsection below. Throughout this section, the kernel function $k(\cdot)$ and the bandwidth parameter h are chosen as mentioned in Section 3.6.

In this DGP, there are two common factors: 1 and f_t and the factor loading is given by $\lambda_i = [\alpha_i, \gamma_i]'$. Section 4.1 examines the estimation of r, whereas Sections 4.2 and 4.3 focus on the estimation of the coefficient of $X_{it,1}$, which varies across different quantiles.

4.1. Estimating the Number of Factors

The performance of the estimator for the number of factor depends crucially on the properties of $e_{j,it}$. Following Bai and Ng (2002), the following DGP for $e_{j,it}$ is used:

$$e_{j,it} = \gamma e_{j,it-1} + \nu_{j,it} + \zeta \cdot \sum_{l=i-m, l\neq i}^{i+m} \nu_{j,lt},$$

where $\nu_{j,it} \sim$ i.i.d. $\mathcal{N}(0,1)$, for $j = 2,3$. The parameter γ controls the serial dependence, and the parameters ζ, m determine the cross-sectional dependence. The following models are considered in the simulations:

Q1: i.i.d. errors: $\gamma = \zeta = 0$.

Q2: serial dependence: $\gamma = 0.8$ and $\zeta = 0$.

Q3: cross-sectional dependence: $\gamma = 0$, $\zeta = 0.2$, and $m = 5$.

Q4: serial and cross-sectional dependence: $\gamma = 0.8$, $\zeta = 0.2$, and $m = 5$.

Recall that the estimator for the number of factors is defined as the number of eigenvalues of $\hat{\Sigma}_{\bar{X}}$ that is larger than $\mathbb{P}_{NT}$. Note that Proposition 1 requires that $\mathbb{P}_{NT} = (\min\{N,T\})^{-c}$ for some $0 < c < 1/2$. Thus, in the simulations, I choose

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

$c = 1/3$. The upper panel of Table 1 reports the frequencies of choosing the right number of factors (denoted as $\hat{P}[\hat{r} = 2]$) and the mean number of estimated factors (denoted as mean$[\hat{r}]$) from 1,000 repetitions for $N, T \in \{20, 50, 100, 200\}$. It can be seen that the proposed method chooses the right number of factors in all models with very high precision as long as $\min[N, T] \geq 50$.

For comparison purposes, the lower panel of Table 1 also reports the results using the eigen-ratio estimator of Ahn and Horenstein (2013) (denoted as $\tilde{r}$). In particular, the eigen-ratio estimator is obtained from a panel consisting of $\bar{X}_{it} = \sum_{j=1}^{3} X_{it,j}$.[8] Generally speaking, $\hat{r}$ performs much better than $\tilde{r}$, especially when N, T is not large, or when there are considerable amounts of serial correlations in the errors (DGPs **Q2** and **Q4**).

4.2. Estimators with Independent Errors

In this subsection, $e_{2,it}, e_{3,it}$ are generated as i.i.d. standard normal random variables, and two different specifications for the distribution of ϵ_{it} are considered:

M1: $\epsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$.

M2: $\epsilon_{it} \sim$ i.i.d. $\mathcal{T}(3)$, where $\mathcal{T}(3)$ denotes the Student's t distribution with three degrees of freedom.

The main object of interest is the quantile coefficients of $X_{1,it}$ at $\tau = 0.25, 0.9$, and the following three estimators are considered:

$\hat{\beta}(\tau)$: the two-step estimator using SQR.

$\hat{\beta}_{abc}(\tau)$: the bias-corrected two-step estimator using analytical bias correction.

$\hat{\beta}_{spj}(\tau)$: the bias-corrected two-step estimator using the SPJ.

The kernel function and the bandwidth parameter are chosen as mentioned in Section 3.6. Given the excellent performance of the estimated number of factors in the previous subsection, the true number of factors is treated as known. The simulation results from 500 repetitions are reported in Table 2, where columns 3–5 report the biases of the estimators, columns 6–8 report the standard deviations, and the last three columns report the coverage rates of the confidence intervals with 95% nominal levels. Note that the DGP in this subsection has no serial correlations. Thus, when constructing the analytical-bias-correction estimators and the confidence intervals, the biases are estimated by setting $\hat{\omega}_{T,i}^{(2)} = \hat{\omega}_{T,i}^{(4)} = 0$, and the covariance matrices are estimated using the formula given in Section 3.5 with $\hat{\mathbf{V}}_2 = 0$.

There are four main takeaways from the simulation results. First, the biases and the standard deviations of the estimators are larger when the distributions of the idiosyncratic errors have heavier tails (normal vs. Student's t distributions) and when the quantile of interest is further away from the median ($\tau = 0.9$ vs. $\tau = 0.25$). This is true for both the original two-step estimators and the bias-corrected estimators.

[8] If one obtains the eigen-ratio estimator for each of the regressors, it is likely that some of the regressors contain less than r factors, and therefore the eigen-ratio estimator will underestimate the number of factors for some regressors. For example, for the DGP considered here, f_t is not a common factor for $X_{it,1}$.

Table 1. The number of factors

	Q1		Q2		Q3		Q4	
(N,T)	$\hat{P}[\hat{r}=2]$	Mean$[\hat{r}]$	$\hat{P}[\hat{r}=2]$	Mean$[\hat{r}]$	$\hat{P}[\hat{r}=2]$	Mean$[\hat{r}]$	$\hat{P}[\hat{r}=2]$	Mean$[\hat{r}]$
(20,20)	0.997	1.997	0.992	1.992	0.996	1.996	0.946	2.028
(20,50)	1.000	2.000	1.000	2.000	1.000	2.000	0.936	2.064
(20,100)	1.000	2.000	1.000	2.000	1.000	2.000	0.887	2.113
(20,200)	1.000	2.000	1.000	2.000	1.000	2.000	0.903	2.097
(50,20)	0.002	1.002	0.021	1.021	0.064	1.064	0.331	1.331
(50,50)	1.000	2.000	0.994	1.994	1.000	2.000	0.978	1.988
(50,100)	1.000	2.000	1.000	2.000	1.000	2.000	0.996	2.004
(50,200)	1.000	2.000	1.000	2.000	1.000	2.000	0.999	2.001
(100,20)	1.000	2.000	1.000	2.000	1.000	2.000	0.987	1.987
(100,50)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(100,100)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(100,200)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(200,20)	1.000	2.000	1.000	2.000	1.000	2.000	0.999	1.999
(200,50)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(200,100)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(200,200)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
	$\hat{P}[\tilde{r}=2]$	Mean$[\tilde{r}]$	$\hat{P}[\tilde{r}=2]$	Mean$[\tilde{r}]$	$\hat{P}[\tilde{r}=2]$	Mean$[\tilde{r}]$	$\hat{P}[\tilde{r}=2]$	Mean$[\tilde{r}]$
(20,20)	0.000	1.000	0.009	1.011	0.000	1.000	0.037	1.059
(20,50)	0.000	1.000	0.000	1.000	0.000	1.000	0.001	1.001
(20,100)	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
(20,200)	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
(50,20)	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
(50,50)	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
(50,100)	0.341	1.341	0.000	1.000	0.000	1.000	0.000	1.000
(50,200)	0.998	1.998	0.000	1.000	0.002	1.002	0.000	1.000
(100,20)	1.000	2.000	0.420	1.454	1.000	2.000	0.342	1.505
(100,50)	1.000	2.000	0.000	1.000	0.988	1.988	0.003	1.003
(100,100)	1.000	2.000	0.000	1.000	0.920	1.920	0.000	1.000
(100,200)	1.000	2.000	0.000	1.000	0.995	1.995	0.000	1.000
(200,20)	1.000	2.000	0.000	1.000	1.000	2.000	0.000	1.008
(200,50)	1.000	2.000	0.000	1.000	0.999	1.999	0.000	1.000
(200,100)	1.000	2.000	0.000	1.000	1.000	2.000	0.000	1.000
(200,200)	1.000	2.000	0.000	1.000	1.000	2.000	0.000	1.000

Notes: 1,000 repetitions. DGP: $f_t \sim$ i.i.d. $\mathcal{N}(0,1)$, $X_{it,1} \sim$ i.i.d. $\chi^2(1)+1$, and $X_{it,2} = \theta_{2i} + \eta_{2i} f_t + e_{2,it}$, $X_{it,3} = \theta_{3i} + \eta_{3i} f_t + e_{3,it}$, where $e_{j,it} = \gamma e_{j,it-1} + \nu_{j,it} + \zeta \cdot \sum_{l=i-m, l \neq i}^{i+m} \nu_{j,lt}$, $\theta_{2i}, \theta_{3i}, \eta_{2i}, \eta_{3i} \sim$ i.i.d. $\mathcal{N}(1,1)$, $\nu_{2,it}, \nu_{3,it} \sim$ i.i.d. $\mathcal{N}(0,1)$. **Q1:** $\gamma = \zeta = 0$; **Q2:** $\gamma = 0.8$, $\zeta = 0$; **Q3:** $\gamma = 0$, $\zeta = 0.2$, $m = 5$; **Q4:** $\gamma = 0.8$, $\zeta = 0.2$, $m = 5$. This table reports the frequencies of choosing the right number of factors, denoted as $\hat{P}[\hat{r}=2]$ and $\hat{P}[\tilde{r}=2]$, and the mean number of estimated factors, denoted as mean$[\hat{r}]$ and mean$[\tilde{r}]$, where $\hat{r}$ is the estimator proposed in Section 3.1, and $\tilde{r}$ is the estimator of Ahn and Horenstein (2013).

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

Table 2. Estimation results without serial dependence

M1		Bias			Std			Coverage rate (95%)		
τ	(N,T)	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$
0.25	(50,50)	0.024	0.004	0.005	0.086	0.484	0.111	0.908	0.596	0.852
	(50,100)	0.016	0.005	−0.011	0.062	0.081	0.076	0.910	0.860	0.854
	(50,200)	0.007	−0.002	−0.001	0.043	0.049	0.050	0.934	0.910	0.906
	(100,50)	0.019	−0.014	0.001	0.062	0.413	0.074	0.908	0.582	0.898
	(100,100)	0.009	0.005	0.000	0.043	0.063	0.051	0.924	0.814	0.890
	(100,200)	0.003	0.001	−0.001	0.031	0.035	0.034	0.946	0.926	0.920
	(200,50)	0.016	0.017	−0.001	0.043	0.484	0.050	0.924	0.484	0.906
	(200,100)	0.007	0.005	−0.001	0.030	0.050	0.035	0.926	0.820	0.908
	(200,200)	0.004	0.002	0.001	0.022	0.024	0.024	0.946	0.920	0.916
0.9	(50,50)	−0.051	0.004	0.017	0.110	0.812	0.151	0.874	0.676	0.812
	(50,100)	−0.030	−0.022	0.001	0.076	0.097	0.100	0.874	0.808	0.798
	(50,200)	−0.014	−0.010	−0.001	0.054	0.061	0.067	0.914	0.874	0.834
	(100,50)	−0.049	−0.039	0.019	0.074	0.218	0.091	0.864	0.664	0.860
	(100,100)	−0.026	−0.021	0.003	0.055	0.066	0.065	0.884	0.834	0.850
	(100,200)	−0.010	−0.003	0.004	0.038	0.043	0.045	0.898	0.868	0.846
	(200,50)	−0.048	−0.046	−0.013	0.057	0.128	0.069	0.796	0.640	0.866
	(200,100)	−0.021	−0.017	0.004	0.037	0.045	0.044	0.864	0.814	0.862
	(200,200)	−0.012	−0.007	0.001	0.026	0.028	0.030	0.884	0.882	0.880
M2		Bias			Std			Coverage rate (95%)		
τ	(N,T)	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$
0.25	(50,50)	0.033	0.077	0.004	0.110	1.502	0.140	0.874	0.604	0.814
	(50,100)	0.020	0.010	0.000	0.075	0.095	0.090	0.884	0.844	0.842
	(50,200)	0.010	−0.000	−0.000	0.053	0.062	0.061	0.928	0.902	0.898
	(100,50)	0.028	0.022	−0.002	0.077	0.341	0.092	0.896	0.606	0.882
	(100,100)	0.012	0.005	−0.004	0.053	0.078	0.061	0.934	0.818	0.902
	(100,200)	0.005	0.001	−0.003	0.036	0.042	0.040	0.950	0.902	0.928
	(200,50)	0.027	0.023	−0.001	0.055	0.214	0.064	0.904	0.522	0.904
	(200,100)	0.010	0.007	−0.004	0.037	0.056	0.042	0.924	0.788	0.902
	(200,200)	0.007	0.002	−0.000	0.026	0.029	0.029	0.946	0.916	0.918
0.9	(50,50)	−0.113	−0.259	−0.006	0.180	3.739	0.250	0.806	0.668	0.782
	(50,100)	−0.061	−0.070	−0.004	0.122	0.183	0.151	0.806	0.750	0.782
	(50,200)	−0.028	−0.027	0.001	0.097	0.111	0.117	0.814	0.784	0.760
	(100,50)	−0.115	−0.113	0.001	0.121	0.266	0.161	0.778	0.656	0.812
	(100,100)	−0.055	−0.050	−0.000	0.095	0.109	0.116	0.768	0.742	0.778
	(100,200)	−0.029	−0.021	0.001	0.066	0.076	0.078	0.808	0.786	0.782
	(200,50)	−0.109	−0.111	0.012	0.094	0.182	0.117	0.688	0.594	0.810
	(200,100)	−0.055	−0.051	0.001	0.067	0.080	0.079	0.726	0.688	0.772
	(200,200)	−0.027	−0.021	0.001	0.044	0.048	0.050	0.786	0.782	0.802

Notes: 500 repetitions. DGP: $Y_{it} = \beta_1 X_{it,1} + \beta_2 X_{it,2} + \beta_3 X_{it,3} + \alpha_i + \gamma_i f_t + X_{it,1} \cdot \epsilon_{it}$, where $[\beta_1, \beta_2, \beta_3] = [1,1,1]$, $\alpha_i \sim$ i.i.d. $\mathcal{N}(0,1)$, $\gamma_i \sim$ i.i.d. $\mathcal{N}(0,1)$, $f_t \sim$ i.i.d. $\mathcal{N}(0,1)$, $X_{it,1} \sim$ i.i.d. $\chi^2(1) + 1$, and $X_{it,2} = \theta_{2i} + \eta_{2i} f_t + e_{2,it}$, $X_{it,3} = \theta_{3i} + \eta_{3i} f_t + e_{3,it}$, where $\theta_{2i}, \theta_{3i}, \eta_{2i}, \eta_{3i} \sim$ i.i.d. $\mathcal{N}(1,1)$, $e_{2,it}, e_{3,it} \sim$ i.i.d. $\mathcal{N}(0,1)$. **M1**: $\epsilon_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$; **M2**: $\epsilon_{it} \sim$ i.i.d. $\mathcal{T}(3)$.

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

Second, it is clear from the results that the biases of the estimators decrease either as N increases while T is fixed, or as T increases while N is fixed. This confirms the existence of a leading bias term whose size depends on both N and T, as established in Theorem 2. Such results are in contrast with the findings in quantile panel models with only individual effects, where the leading bias term is approximately of order T^{-1} and thus the biases decrease only when T increases.

Third, for the analytical bias correction to have good performance, the number of time series observations (T) needs to be at least 100. On the other hand, the SPJ performs much better when $T = 50$ because there is no need to estimate those complex objects (such as the inverse of the density functions) when constructing the estimators of the biases.

Last but not least, it can be seen that both analytical and the SPJ bias corrections can significantly reduce the biases of the two-step estimator, as predicted by the theoretical results. In particular, it can be observed that the SPJ generally does a better job at reducing the biases. However, the reduction of biases comes at the cost of inflating the standard deviations—this is especially noticeable for the analytical bias correction when $T = 50$. As a consequence, the coverage rates of the confidence intervals based on the bias-corrected estimators are in general lower than those based on the original two-step estimators. Therefore, different from the usual suggestion of applying bias correction technique to the fixed-effects estimator of nonlinear panel data models (including quantile panel data models) to improve finite sample performance, for the models considered in this paper, the important lesson we can learn is that bias correction can be harmful and it is actually better to use the original estimator (without bias correction) to achieve better finite sample performance.

4.3. Estimators with Serially Correlated Errors

In this subsection, I consider models where ϵ_{it} are generated as autoregressive processes:

$$\epsilon_{it} = \rho \cdot \epsilon_{i,t-1} + \sqrt{1-\rho^2} \cdot \nu_{it}, \text{ where } \nu_{it} \sim \text{i.i.d. } \mathcal{N}(0,1).$$

As in the previous subsection, $e_{2,it}$ and $e_{3,it}$ are i.i.d. standard normal variables. Now, $\hat{\omega}_{T,i}^{(2)}$, $\hat{\omega}_{T,i}^{(4)}$, and $\hat{\mathbf{V}}_2$ are estimated by the formulas given in Sections 3.4 and 3.5. As discussed in Section 3.6, I focus on the choice of $L = 1, \ldots, 4$. The results with moderate serial correlation ($\rho = 0.5$) are reported in Table 3 for $L = 1, 2$ and in Table 4 for $L = 3, 4$.

In general, except for a few cases where the standard deviation of $\hat{\beta}_{abc}$ is extremely large, which usually happens when $T = 50$, the results are very similar to those reported in Table 2 where the errors have no serial correlations. In particular, changing the truncation parameter L from 1 to 4 does not significantly improve the finite sample performance of the estimators. This is also true if L is allowed to increase with sample sizes (more simulation results are available upon request).

TABLE 3. Estimation results with serial dependence

$L=1$		Bias			Std			Coverage rate (95%)		
τ	(N,T)	$\hat\beta(\tau)$	$\hat\beta_{abc}(\tau)$	$\hat\beta_{spj}(\tau)$	$\hat\beta(\tau)$	$\hat\beta_{abc}(\tau)$	$\hat\beta_{spj}(\tau)$	$\hat\beta(\tau)$	$\hat\beta_{abc}(\tau)$	$\hat\beta_{spj}(\tau)$
0.25	(50,50)	0.026	−0.018	−0.007	0.087	1.060	0.112	0.910	0.564	0.856
	(50,100)	0.021	0.008	0.002	0.063	0.084	0.073	0.924	0.840	0.888
	(50,200)	0.012	0.005	0.004	0.047	0.056	0.052	0.922	0.880	0.898
	(100,50)	0.024	0.005	−0.001	0.062	0.419	0.076	0.930	0.542	0.894
	(100,100)	0.017	0.009	0.004	0.042	0.077	0.050	0.940	0.776	0.916
	(100,200)	0.005	0.001	−0.001	0.031	0.039	0.036	0.948	0.914	0.916
	(200,50)	0.024	0.013	−0.002	0.045	0.290	0.053	0.908	0.508	0.894
	(200,100)	0.012	0.006	−0.000	0.032	0.050	0.037	0.932	0.794	0.902
	(200,200)	0.006	0.003	0.000	0.023	0.027	0.025	0.934	0.880	0.914
0.9	(50,50)	−0.076	−0.095	0.013	0.110	0.475	0.154	0.824	0.632	0.800
	(50,100)	−0.040	−0.034	−0.001	0.074	0.095	0.092	0.848	0.806	0.820
	(50,200)	−0.012	−0.006	0.005	0.052	0.064	0.062	0.916	0.882	0.886
	(100,50)	−0.063	−0.050	0.015	0.082	0.414	0.107	0.814	0.612	0.814
	(100,100)	−0.035	−0.030	0.000	0.058	0.083	0.071	0.822	0.774	0.808
	(100,200)	−0.018	−0.009	0.002	0.039	0.044	0.045	0.876	0.868	0.870
	(200,50)	−0.065	−0.065	0.011	0.055	0.151	0.071	0.728	0.584	0.836
	(200,100)	−0.032	−0.024	0.004	0.040	0.048	0.049	0.786	0.792	0.840
	(200,200)	−0.015	−0.008	0.002	0.028	0.031	0.031	0.854	0.852	0.856
$L=2$		Bias			Std			Coverage rate (95%)		
τ	(N,T)	$\hat\beta(\tau)$	$\hat\beta_{abc}(\tau)$	$\hat\beta_{spj}(\tau)$	$\hat\beta(\tau)$	$\hat\beta_{abc}(\tau)$	$\hat\beta_{spj}(\tau)$	$\hat\beta(\tau)$	$\hat\beta_{abc}(\tau)$	$\hat\beta_{spj}(\tau)$
0.25	(50,50)	0.026	−0.046	−0.007	0.087	1.548	0.112	0.910	0.564	0.866
	(50,100)	0.021	0.008	0.002	0.063	0.086	0.073	0.930	0.834	0.896
	(50,200)	0.012	0.005	0.004	0.047	0.057	0.052	0.926	0.886	0.902
	(100,50)	0.024	0.006	−0.001	0.062	0.411	0.076	0.938	0.520	0.898
	(100,100)	0.017	0.008	0.004	0.042	0.083	0.050	0.940	0.766	0.918
	(100,200)	0.005	0.001	−0.001	0.031	0.040	0.036	0.952	0.914	0.916
	(200,50)	0.024	0.010	−0.002	0.045	0.312	0.053	0.914	0.508	0.898
	(200,100)	0.012	0.005	−0.000	0.032	0.052	0.037	0.934	0.776	0.912
	(200,200)	0.006	0.003	0.000	0.023	0.027	0.025	0.936	0.876	0.916
0.9	(50,50)	−0.076	−0.091	0.013	0.110	0.472	0.154	0.828	0.624	0.798
	(50,100)	−0.040	−0.034	−0.001	0.074	0.097	0.092	0.850	0.800	0.822
	(50,200)	−0.012	−0.005	0.005	0.052	0.064	0.062	0.877	0.916	0.901
	(100,50)	−0.063	−0.049	0.015	0.082	0.461	0.107	0.818	0.616	0.808
	(100,100)	−0.035	−0.030	0.000	0.058	0.084	0.071	0.824	0.762	0.816
	(100,200)	−0.018	−0.009	0.002	0.039	0.044	0.045	0.878	0.864	0.870
	(200,50)	−0.065	−0.063	0.011	0.055	0.161	0.071	0.736	0.562	0.844
	(200,100)	−0.032	−0.024	0.004	0.040	0.049	0.049	0.790	0.808	0.844
	(200,200)	−0.015	−0.008	0.002	0.028	0.031	0.031	0.860	0.856	0.858

Notes: 500 repetitions. DGP: $Y_{it}=\beta_1X_{it,1}+\beta_2X_{it,2}+\beta_3X_{it,3}+\alpha_i+\gamma_if_t+X_{it,1}\cdot\epsilon_{it}$, where $[\beta_1,\beta_2,\beta_3]=[1,1,1]$, $\alpha_i\sim$ i.i.d. $\mathcal{N}(0,1)$, $\gamma_i\sim$ i.i.d. $\mathcal{N}(0,1)$, $f_t\sim$ i.i.d. $\mathcal{N}(0,1)$, $X_{it,1}\sim$ i.i.d. $\chi^2(1)+1$, and $X_{it,2}=\theta_{2i}+\eta_{2i}f_t+e_{2,it}$, $X_{it,3}=\theta_{3i}+\eta_{3i}f_t+e_{3,it}$, where $\theta_{2i},\theta_{3i},\eta_{2i},\eta_{3i}\sim$ i.i.d. $\mathcal{N}(1,1)$, $e_{2,it},e_{3,it}\sim$ i.i.d. $\mathcal{N}(0,1)$. $\epsilon_{it}=\rho\cdot\epsilon_{i,t-1}+\sqrt{1-\rho^2}\cdot\nu_{it}$, where $\nu_{it}\sim$ i.i.d. $\mathcal{N}(0,1)$ and $\rho=0.5$.

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

TABLE 4. Estimation results with serial dependence

$L=3$		Bias			Std			Coverage rate (95%)		
τ	(N,T)	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$
0.25	(50,50)	0.026	−0.037	−0.007	0.087	1.559	0.112	0.910	0.574	0.864
	(50,100)	0.021	0.007	0.002	0.063	0.087	0.073	0.928	0.838	0.898
	(50,200)	0.012	0.005	0.004	0.047	0.057	0.052	0.926	0.878	0.902
	(100,50)	0.024	0.002	−0.001	0.062	0.409	0.076	0.938	0.528	0.898
	(100,100)	0.017	0.008	0.004	0.042	0.086	0.050	0.940	0.766	0.922
	(100,200)	0.005	0.001	−0.001	0.031	0.040	0.036	0.952	0.916	0.920
	(200,50)	0.024	0.009	−0.002	0.045	0.381	0.054	0.916	0.462	0.904
	(200,100)	0.012	0.008	−0.000	0.032	0.055	0.037	0.938	0.774	0.914
	(200,200)	0.006	0.004	0.000	0.023	0.028	0.025	0.942	0.867	0.928
0.9	(50,50)	−0.076	−0.092	0.013	0.110	0.477	0.154	0.830	0.632	0.796
	(50,100)	−0.040	−0.034	−0.001	0.074	0.097	0.092	0.850	0.794	0.818
	(50,200)	−0.012	−0.005	0.005	0.052	0.065	0.062	0.920	0.878	0.892
	(100,50)	−0.063	−0.052	0.015	0.082	0.421	0.107	0.816	0.614	0.806
	(100,100)	−0.035	−0.030	0.000	0.058	0.085	0.071	0.822	0.756	0.812
	(100,200)	−0.018	−0.009	0.002	0.039	0.044	0.045	0.878	0.868	0.870
	(200,50)	−0.065	−0.064	0.011	0.055	0.148	0.071	0.730	0.566	0.854
	(200,100)	−0.032	−0.025	0.004	0.040	0.051	0.049	0.792	0.774	0.850
	(200,200)	−0.015	−0.008	0.002	0.028	0.035	0.030	0.881	0.880	0.867
$L=4$		Bias			Std			Coverage rate (95%)		
τ	(N,T)	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$
0.25	(50,50)	0.026	−0.047	−0.007	0.087	1.628	0.112	0.914	0.578	0.866
	(50,100)	0.021	0.007	0.002	0.063	0.087	0.073	0.930	0.840	0.898
	(50,200)	0.012	0.005	0.004	0.047	0.058	0.052	0.926	0.880	0.906
	(100,50)	0.024	0.006	−0.001	0.062	0.411	0.076	0.938	0.520	0.898
	(100,100)	0.017	0.009	0.004	0.042	0.086	0.050	0.942	0.772	0.920
	(100,200)	0.005	0.001	−0.001	0.031	0.040	0.036	0.956	0.922	0.918
	(200,50)	0.024	0.011	−0.002	0.045	0.408	0.054	0.916	0.476	0.900
	(200,100)	0.012	0.008	−0.000	0.032	0.055	0.037	0.938	0.762	0.918
	(200,200)	0.006	0.004	0.000	0.023	0.028	0.025	0.942	0.859	0.921
0.9	(50,50)	−0.076	−0.087	0.013	0.110	0.437	0.154	0.826	0.654	0.796
	(50,100)	−0.040	−0.034	−0.001	0.074	0.097	0.092	0.850	0.794	0.814
	(50,200)	−0.012	−0.005	0.005	0.052	0.065	0.062	0.922	0.880	0.890
	(100,50)	−0.063	−0.051	0.015	0.082	0.390	0.107	0.820	0.620	0.806
	(100,100)	−0.035	−0.030	0.000	0.058	0.086	0.071	0.824	0.762	0.814
	(100,200)	−0.018	−0.009	0.002	0.039	0.044	0.045	0.878	0.869	0.870
	(200,50)	−0.065	−0.065	0.011	0.055	0.135	0.071	0.734	0.568	0.864
	(200,100)	−0.032	−0.025	0.004	0.040	0.051	0.049	0.792	0.778	0.854
	(200,200)	−0.015	−0.008	0.002	0.028	0.032	0.031	0.863	0.841	0.868

Notes: 500 repetitions. DGP: $Y_{it} = \beta_1 X_{it,1} + \beta_2 X_{it,2} + \beta_3 X_{it,3} + \alpha_i + \gamma_i f_t + X_{it,1} \cdot \epsilon_{it}$, where $[\beta_1, \beta_2, \beta_3] = [1,1,1]$, $\alpha_i \sim$ i.i.d. $\mathcal{N}(0,1)$, $\gamma_i \sim$ i.i.d. $\mathcal{N}(0,1)$, $f_t \sim$ i.i.d. $\mathcal{N}(0,1)$, $X_{it,1} \sim$ i.i.d. $\chi^2(1)+1$, and $X_{it,2} = \theta_{2i} + \eta_{2i} f_t + e_{2,it}$, $X_{it,3} = \theta_{3i} + \eta_{3i} f_t + e_{3,it}$, where $\theta_{2i}, \theta_{3i}, \eta_{2i}, \eta_{3i} \sim$ i.i.d. $\mathcal{N}(1,1)$, $e_{2,it}, e_{3,it} \sim$ i.i.d. $\mathcal{N}(0,1)$. $\epsilon_{it} = \rho \cdot \epsilon_{i,t-1} + \sqrt{1-\rho^2} \cdot \nu_{it}$, where $\nu_{it} \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\rho = 0.5$.

https://doi.org/10.1017/S0266466622000366 Published online by Cambridge University Press

5. CONCLUSIONS

Estimating the coefficients of the regressors and the interactive fixed effects jointly in a quantile panel model is not only computationally difficult but also theoretically challenging to derive the asymptotic properties of the estimators, mainly due to the fact that the objective function is nonsmooth and nonconvex. In this paper, an easy-to-implement two-step estimator is proposed. Because the SQRs are used in the second step, the derivation of the asymptotic distribution and the asymptotic biases of the estimator is feasible. The asymptotic distribution provides a formal justification for the use of analytical bias correction and a heuristic argument for the use of the SPJ to correct the asymptotic biases, and the simulation results confirm that both bias correction methods can effectively reduce the biases with moderate sample sizes. However, it should be cautioned that the bias correction methods inevitably inflate the standard deviations of the estimators, and result in confidence intervals with lower coverage rate than the estimators without bias correction. Finally, even though this paper provides conditions with regard to the sizes of the bandwidth parameter in SQR and the truncation parameter in the HAC-type estimators of the bias and variance, there remains the important but challenging question of how to choose these parameters optimally in a data-dependent manner. This question is left for future research.

SUPPLEMENTARY MATERIAL

Chen, L. (2022): Supplement to "Two-Step Estimation of Quantile Panel Data Models with Interactive Fixed Effects", Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466622000366

REFERENCES

Abrevaya, J. & C.M. Dahl (2008) The effects of birth inputs on birthweight: Evidence from quantile estimation on panel data. *Journal of Business & Economic Statistics* 26(4), 379–397.

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

Amemiya, T. (1982) Two stage least absolute deviations estimators. *Econometrica* 50(3), 689–711.

Ando, T. & J. Bai (2020) Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity. *Journal of the American Statistical Association* 115(529), 266–279.

Arellano, M. & S. Bonhomme (2016) Nonlinear panel data estimation via quantile regressions. *The Econometrics Journal* 19(3), C61–C94.

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

Belloni, A., M. Chen, & O.H.M. Padilla (2019) High Dimensional Latent Panel Quantile Regression with an Application to Asset Pricing. WERPS working paper 1230, Department of Economics, University of Warwick.

Cai, Z., L. Chen, & Y. Fang (2018) A semiparametric quantile panel data model with an application to estimating the growth effect of FDI. *Journal of Econometrics* 206(2), 531–553.

Canay, I.A. (2011) A simple approach to quantile regression for panel data. *The Econometrics Journal* 14(3), 368–386.
Chen, L. (2015) Set identification of panel data models with interactive effects via quantile restrictions. *Economics Letters* 137, 36–40.
Chen, L. (2019) Nonparametric Quantile Regressions for Panel Data Models with Large T. PHBS working paper 20210104, Peking University HSBC Business School.
Chen, L., J.J. Dolado, & J. Gonzalo (2021a) Quantile factor models. *Econometrica* 89(2), 875–910.
Chen, L. & Y. Huo (2021) A simple estimator for quantile panel data models using smoothed quantile regressions. *The Econometrics Journal* 24(2), 247–263.
Chen, M. (2016) Estimation of Nonlinear Panel Models with Multiple Unobserved Effects. WERPS Working paper 1120, Department of Economics, University of Warwick.
Chen, M., I. Fernández-Val, & M. Weidner (2021b) Nonlinear factor models for network and panel data. *Journal of Econometrics* 220(2), 296–324.
Chudik, A. & H. Pesaran (2015) Large panel data models with cross-sectional dependence. In B. H., Baltagi (ed.), *The Oxford Handbook of Panel Data*, pp. 3–45, Oxford University Press.
Dhaene, G. & K. Jochmans (2015) Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies* 82(3), 991–1030.
Feng, J. (2019) Regularized Quantile Regression with Interactive Fixed Effects. arXiv:1911.00166
Fernández-Val, I. & M. Weidner (2016) Individual and time effects in nonlinear panel models with large *N*, *T*. *Journal of Econometrics* 192(1), 291–312.
Galvao, A.F. (2011) Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics* 164(1), 142–157.
Galvao, A.F., J. Gu, & S. Volgushev (2020) On the unbiased asymptotic normality of quantile regression with fixed effects. *Journal of Econometrics* 218(1), 178–215.
Galvao, A.F. & K. Kato (2016) Smoothed quantile regression for panel data. *Journal of Econometrics* 193(1), 92–112.
Galvao, A.F., C. Lamarche, & L.R. Lima (2013) Estimation of censored quantile regression for panel data with fixed effects. *Journal of the American Statistical Association* 108(503), 1075–1089.
Galvao, A.F. & G.V. Montes-Rojas (2010) Penalized quantile regression for dynamic panel data. *Journal of Statistical Planning and Inference* 140(11), 3476–3497.
Graham, B.S., J. Hahn, A. Poirier, & J.L. Powell (2018) A quantile correlated random coefficients panel data model. *Journal of Econometrics* 206(2), 305–335.
Hahn, J. & G. Kuersteiner (2011) Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27(6), 1152–1191.
Harding, M. & C. Lamarche (2014) Estimating and testing a quantile regression model with interactive effects. *Journal of Econometrics* 178, 101–113.
Harding, M., C. Lamarche, & M.H. Pesaran (2020) Common correlated effects estimation of heterogeneous dynamic panel quantile regression models. *Journal of Applied Econometrics* 35(3), 294–314.
Horowitz, J.L. (1998) Bootstrap methods for median regression models. *Econometrica*, 1327–1351.
Karabiyik, H., S. Reese, & J. Westerlund (2017) On the role of the rank condition in CCE estimation of factor-augmented panel regressions. *Journal of Econometrics* 197(1), 60–64.
Kato, K., A.F. Galvao, & G.V. Montes-Rojas (2012) Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics* 170(1), 76–91.
Koenker, R. (2004) Quantile regression for longitudinal data. *Journal of Multivariate Analysis* 91(1), 74–89.
Lamarche, C. (2010) Robust penalized quantile regression estimation for panel data. *Journal of Econometrics* 157(2), 396–408.
Lu, X. & L. Su (2016) Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics* 190(1), 148–175.
Ma, S., O. Linton, & J. Gao (2021) Estimation and inference in semiparametric quantile factor models. *Journal of Econometrics* 222(1), 295–323.
Moon, H.R. & M. Weidner (2015) Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* 83(4), 1543–1579.

Muller, H.G. (1984) Smooth optimum kernel estimators of densities, regression curves and modes. *Annals of Statistics* 12(2), 766–774.

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

Rao, B.P. (2009) Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics* 61(2), 441–460.

Rosen, A.M. (2012) Set identification via quantile restrictions in short panels. *Journal of Econometrics* 166(1), 127–137.

Su, L. & Q. Chen (2013) Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory* 29(6), 1079–1135.

Yoon, J. & A.F. Galvao (2020) Cluster robust covariance matrix estimation in panel quantile regression with individual fixed effects. *Quantitative Economics* 11(2), 579–608.