Department of Economics

Group for Economic Analysis at Reading (GEAR)

Dynamic Quantile Panel Data Models with Interactive Effects

by Jia Chen, Yongcheol Shin and Chaowen Zheng

Discussion Paper No. 2023-06

Department of Economics
University of Reading
Whiteknights
Reading
RG6 6AA
United Kingdom

www.reading.ac.uk

Dynamic Quantile Panel Data Models with Interactive Effects*

Jia Chen† Yongcheol Shin ‡ Chaowen Zheng §

This Version: June 2023

Abstract

We propose a simple two-step procedure for estimating the dynamic quantile panel data model with unobserved interactive effects. To account for the endogeneity induced by correlation between factors and lagged dependent variable/regressors, we first estimate factors consistently via an iterative principal component analysis. In the second step, we run a quantile regression for the augmented model with estimated factors and estimate the slope parameters. In particular, we adopt a smoothed quantile regression analysis where the quantile loss function is smoothed to have well-defined derivatives. The proposed two-step estimator is consistent and asymptotically normally distributed, but subject to asymptotic bias due to the incidental parameters. We then apply the split-panel jackknife approach to correct the bias. Monte Carlo simulations confirm that our proposed estimator has good finite sample performance. Finally, we demonstrate the usefulness of our proposed approach with an application to the analysis of bilateral trade for 380 country pairs over 59 years.

Keywords: Dynamic Quantile Panel Data Model, Interactive Effects, Principal Component Analysis, Smoothed Quantile Regression, Bilateral Trade Flows.

JEL codes: C31, C33, F14

*We are mostly grateful to Laura Coroneo, Degui Li, Yuning Li, Takashi Yamagata, and the seminar participants at the 15th International Conference on Computational and Financial Econometrics and University of York for helpful comments. Chen and Shin gratefully acknowledge partial financial support from the ESRC (Grant Reference: ES/T01573X/1). Zheng gratefully acknowledges partial financial support from the Royal Economic Society. The usual disclaimer applies.

†Department of Economics and Related Studies, University of York, York, YO10 5DD, UK; E-mail: jia.chen@york.ac.uk

‡Department of Economics and Related Studies, University of York, York, YO10 5DD, UK; E-mail: yongcheol.shin@york.ac.uk

§Department of Economics, University of Reading, Reading, RG6 6EL, UK; E-mail: chaowen.zheng@reading.ac.uk

1 Introduction

Over the past two decades panel data models with interactive effects have been rapidly developed in order to explicitly model pervasive cross-section dependence (CSD) well-documented in the literature (e.g., Pesaran (2015); Mastromarco et al. (2016)). Significant progress has been made in both estimation and inference for the linear model (e.g. Pesaran (2006), Bai (2009)), the non-linear model (e.g. Chen et al. (2021c)), the dynamic model (e.g. Moon and Weidner (2017)), and non-parametric models (e.g. Dong et al. (2021)). See Bai and Wang (2016) and Sarafidis and Wansbeek (2021) for excellent reviews.

Fundamentally, economic relationships are dynamic and thus involve dynamic adjustments due to habit persistence, inertia, sunk cost and so on. For example, a number of empirical studies have been conducted in labour economics (Meghir and Pistaferri (2004)), economic growth (Acemoglu et al. (2008)) and international economics (Keller (2004)). Simultaneously, dynamic panel data models have been developed theoretically (e.g. Chudik and Pesaran (2015), Moon and Weidner (2017)) and widely applied empirically (Eberhardt and Presbitero (2015), Temple and Van de Sijpe (2017)).

Most of existing studies in the dynamic panel data literature focus on estimating the mean effects, although there are many situations where we would like to investigate heterogeneous effects across the (conditional) distribution of variables of interest. For example, policy makers are interested in not only the policy effects at the mean, but also its effects across different quantiles. Moreover, we observe that several economic time series tend to display asymmetric dynamics in which case we need to incorporate such behavior in the modelling strategy to avoid misleading estimation or inference (e.g. Koenker and Xiao (2006)).

Since the seminar paper by Koenker and Bassett Jr (1978), a quantile regression has become a standard approach to modelling distributional effects. Koenker (2004) also introduces the quantile regression approach to the panel data. In the panel quantile regression with individual effects Kato et al. (2012) point out that we need to impose more restrictive conditions to establish $\sqrt{NT}$-consistency and the asymptotic normality of the corresponding estimator than those in nonlinear models with a smooth objective function. Galvao and Montes-Rojas (2010) and Galvao (2011) develop a dynamic quantile panel data model with individual effects and suggest the instrumental variable approach to dealing with the Nickel bias (Nickell (1981)). Harding and Lamarche (2014) extend the model by introducing interactive effects to capture CSD and propose the use of the common correlated common effects (CCE) estimator by Pesaran (2006). See also further extensions by Ando and Bai (2020) and Chen et al. (2021b), which allow the common factors to be quantile-dependent.

In particular, Harding et al. (2020) consider the dynamic quantile panel data models with unobserved common factors. Assuming that the dependent variable and regressors share the same common factors, they propose the CCE approach, where unobserved factors are proxied by cross section averages of the dependent and independent variables (and their time lags) for consistent estimation. Under the stringent condition (e.g. $N^2(\log N)/T \to 0$), they can establish $\sqrt{NT}$-

consistency and asymptotic normality of the quantile estimator. However, it is restrictive to assume that the dependent variable and regressors share the exactly same factors while the use of the cross-section average of the dependent variable as a factor proxy could introduce the small-sample bias, see Chen et al. (2021a). Moreover, there is no formal procedure for selecting the optimal lag order for better factor approximation. For example, Chudik and Pesaran (2015) argue that the optimal lag order could be selected differently for the coefficient on the lagged dependent variable and coefficients on the regressors, which will induces further estimation uncertainty.

In this paper, we develop a simple two-step estimation procedure for dynamic quantile panel data models withe interactive effects (IE). In the first step, we follow Bai (2009) and Moon and Weidner (2017) and estimate unobserved factors consistently by using the iterative principle component (IPC) method. In the second step, we construct an augmented model with estimated factors and run a smoothed quantile regression (see Galvao and Kato (2016)) to consistently estimate the main parameters.

Our approach is more general than the CCE approach advanced by Harding et al. (2020) as follows: First, our approach does not require any data generating process for the regressors and thus is free from imposing the assumption on the relation between the number of regressors and the number of factors (i.e., the rank condition in the CCE literature). In this regard, we allow arbitrary correlation between regressors and factors/loadings.[1] Second, we don't need to select any tuning parameters (e.g., the number of time lags for factor approximation in the CCE approach) and thus avoid any induced estimation uncertainty. Third, by employing a smoothed quantile regression approach, we are able to develop the asymptotic theory without imposing any stringent conditions. Chen (2021) also develops a two-step procedure for the estimation of a static panel data model with IE. But, he still maintains the assumption that the dependent variable and regressors share the same factors. Moreover, unlike Chen (2021), we consider dynamics explicitly, which brings new technical challenges.

We establish that the proposed two-step estimator is $\sqrt{NT}$-consistent and follows the limiting normal distribution, though it is subject to asymptotic biases arising from the estimation of the factors and loadings. This incidental parameter problem has been widely documented in nonlinear panel data models, see Hahn and Kuersteiner (2011) and Chen et al. (2021c). We also find that the classic Nickell bias (Nickell (1981)) exists in dynamic quantile models. To correct these biases, we propose the use of the spilt panel jackknife (SPJ) method and derive that the bias-corrected estimator follows a centered normal distribution asymptotically.

Via Monte Carlo simulations, we show that the finite sample performance of the proposed two-step estimator is quite satisfactory under various experiments with different sample sizes, idiosyncratic error distributions and quantile levels. In particular, the two-step estimator displays smaller (almost negligible) bias and RMSE than those of alternative existing estimators. Further, the size of t-test is close to the nominal level in almost all cases while its power tends to 1 with

[1] Though we need to consistently estimate the number of factors in the principle component analysis, such a theory is rather well-developed, see Bai and Ng (2002), Ahn and Horenstein (2013).

both N and T.

Finally, we demonstrate the usefulness of our proposed approach with an application to the analysis of bilateral trade flows data for 380 country pairs of 14 European Union (EU) and 6 OECD countries over the period 1960-2018 (59 years). In particular, we aim to examine the benefits of European Economic Community (EEC) membership on trade and the potential impact of Brexit on the UK economy. We find that the benefits of an EEC membership can be significantly larger during recession while its long run effect ranges from 27.4% at upper quantiles to 118.7% at lower quantiles, covering most estimated effects documented in the literature. For a comparison we also estimate the impacts of EEC membership using a static IE model by Chen (2021) and the CCE estimation by Harding et al. (2020), and find that those results are less sensible and occasionally difficult to interpret economically. This may highlight the importance of controlling dynamics and pervasive CSD appropriately in the analysis of the trade dataset.

The rest of the paper is organised as follows. Section 2 describes the model and the two-step estimation procedure. Section 3 establishes the asymptotic theory under gives the maintained assumptions. Section 4 presents the finite sample performance of the proposed estimator via Monte Carlo simulations. Section 5 provides an empirical application to uni-directional bilateral trade flows for 380 country pairs of 14 EU and 6 OECD countries. Section 6 concludes. The mathematical proofs are relegated to the Appendix A while additional Lemmas and Monte Carlo simulation results are presented in Appendices S1 and S2.

2 The Model and the Estimator

Consider the following random coefficient panel data model with unobserved common factors:

$$\begin{aligned} y_{it} &= \psi_0(u_{it}) y_{i,t-1} + \boldsymbol{\beta}_0(u_{it})' \boldsymbol{x}_{it} + \boldsymbol{\gamma}_{i0}(u_{it})' \boldsymbol{f}_{t0} \\ &= \boldsymbol{\theta}_0(u_{it})' \boldsymbol{z}_{it} + \boldsymbol{\gamma}_{i0}(u_{it})' \boldsymbol{f}_{t0} \end{aligned} \tag{1}$$

where y_{it} is the dependent variable of the i-th individual at time t, $y_{i,t-1}$ is the lagged value of y_{it} with the corresponding autoregressive coefficient, $\psi_0(u_{it})$, $\boldsymbol{x}_{it} = (x_{it,1}, \ldots, x_{it,p})'$ is a $p \times 1$ vector of independent variables including the constant with $\boldsymbol{\beta}_0(u_{it})$ being the corresponding vector of homogeneous parameters, $\boldsymbol{f}_{t0}$ is an $r \times 1$ vector of common factors with factor loadings, $\boldsymbol{\gamma}_{i0}(u_{it})$, and $u_{it} \sim i.i.d.U[0,1]$ is assumed to be independent of all the other variables in the model. We also denote $\boldsymbol{\theta}_0(u_{it}) = (\psi_0(u_{it}), \boldsymbol{\beta}_0(u_{it})')'$, and $\boldsymbol{z}_{it} = (y_{i,t-1}, \boldsymbol{x}_{it}')'$.

As a common practice in the quantile regression literature, we assume that the co-monotonicity condition for $\boldsymbol{\theta}_0(\cdot)$ holds to avoid quantile crossing (see Koenker (2005)). Then, we have the τ-th quantile representation for (1) as follows:

$$y_{it} = \boldsymbol{\theta}_0(\tau)' \boldsymbol{z}_{it} + \boldsymbol{\gamma}_{i0}(\tau)' \boldsymbol{f}_{t0} + \mu_{it}(\tau) \quad \text{with } P[\mu_{it}(\tau) \leq 0 | \boldsymbol{z}_{it}, \boldsymbol{f}_{t0}] = \tau. \tag{2}$$

We aim to estimate $\boldsymbol{\theta}_0(\tau) = (\psi_0(\tau), \boldsymbol{\beta}_0(\tau)')'$. In what follows, we will drop the dependence on τ

such that $\boldsymbol{\theta}_0(\tau) \equiv \boldsymbol{\theta}_0$ and $\boldsymbol{\gamma}_{i0}(\tau) \equiv \boldsymbol{\gamma}_{i0}$ for simplicity.

If the factors, $\boldsymbol{f}_{t0}$, are observed, then a natural approach is to estimate the nuisance parameters, $\boldsymbol{\gamma}_{i0}$ together with the main parameters, $\boldsymbol{\theta}_0$ by minimising the following objective function:

$$(\hat{\boldsymbol{\theta}}', \hat{\gamma}_1', \ldots, \hat{\gamma}_N')' = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}, \boldsymbol{\gamma}_i\in\mathcal{A}}{\operatorname{argmin}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau(y_{it} - \boldsymbol{\theta}'\boldsymbol{z}_{it} - \boldsymbol{\gamma}_i'\boldsymbol{f}_{t0}), \tag{3}$$

where $\boldsymbol{\Theta}$ and $\mathcal{A}$ are real compact sets and $\rho_\tau(\mu) = \mu[\tau - \mathbb{1}(\mu \leq 0)]$ is the check function. Due to the non-smoothness of the indicator function, it would be more difficult to derive the asymptotic normality of the estimators than establishing their consistency (Galvao (2011)), as most techniques developed in the literature for linear/nonlinear and static/dynamic panel data models crucially rely upon the smoothness of the objective function, (see Hahn and Newey (2004) and Hahn and Kuersteiner (2011)).

To tackle the above issue, Galvao and Kato (2016) propose smoothing the objetive function for quantile regression by replacing the indicator function with a kernel function following the work by Horowitz (1998), and derive the asymptotic distribution of the estimators for quantile panel data models with fixed effects under regularity conditions. In this paper, we adopt this smoothing approach and propose estimating the model (2) by minimising the smoothed objective function:

$$(\hat{\boldsymbol{\theta}}', \hat{\gamma}_1', \ldots, \hat{\gamma}_N')' = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}, \boldsymbol{\gamma}_i\in\mathcal{A}}{\operatorname{argmin}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [\tau - K(\frac{y_{it} - \boldsymbol{\theta}'\boldsymbol{z}_{it} - \boldsymbol{\gamma}_i'\boldsymbol{f}_{t0}}{h})](y_{it} - \boldsymbol{\theta}'\boldsymbol{z}_{it} - \boldsymbol{\gamma}_i'\boldsymbol{f}_{t0}), \tag{4}$$

where $K(z) = 1 - \int_{-1}^{z} k(x)dx$, $k(\cdot)$ is a symmetric continuous kernel function with bounded support [-1, 1], and h is a bandwidth parameter. By rendering $k(\cdot)$ to be smooth up to certain orders, we can show that the estimator obtained by (4) is asymptotically equivalent to that obtained by (3). More importantly, the smoothed objective function enables us to establish the asymptotic distribution of the estimator using the well-developed techniques (e.g. Chen et al. (2021c)).

In practice, however, minimisation in (3) or (4) is not directly feasible because the factors $\boldsymbol{f}_{t0}$ are unobserved. In the dynamic quantile panel data model with unobserved common factors, Harding et al. (2020) propose the CCE approach, where unobserved factors are proxied by cross section averages of the dependent and independent variables (and their time lags) for consistent estimation.[2] This approach requires the crucial assumption that the dependent variable and regressors share the same common factors, though it is restrictive and untenable in practice. Moreover, the correlation between regressors and factor loadings is not allowed in the CCE approach. In general this implies that the CCE estimation could suffer from endogeneity and/or an efficiency loss due to the correlation between any remaining unapproximated factors and the lagged dependent variable.

In this paper, we propose a different approach to estimating $\boldsymbol{f}_{t0}$ by applying the iterative principal component (IPC) directly to the model (1).[3] We then plug-in the estimated factors $\hat{\boldsymbol{f}}_t$

[2] Under the same assumption, Chen (2021) develops an estimator using the principal component (PC) method for a quantile panel data model.

[3] We assume that the number of factors is known, as its consistent selection has been well developed in the

into (4) and estimate the main parameters by minimising the following objective function:[4]

$$(\hat{\boldsymbol{\theta}}', \hat{\gamma}_1', \ldots, \hat{\gamma}_N')' = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}, \gamma_i \in \mathcal{A}}{\arg\min} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [\tau - K(\frac{y_{it} - \boldsymbol{\theta}' \boldsymbol{z}_{it} - \boldsymbol{\gamma}_i' \hat{\boldsymbol{f}}_t}{h})](y_{it} - \boldsymbol{\theta}' \boldsymbol{z}_{it} - \boldsymbol{\gamma}_i' \hat{\boldsymbol{f}}). \quad (5)$$

To better illustrate the idea behind the proposed two-step estimation procedure, we consider the following location and scale shift panel data model with one unobserved factor:

$$y_{it} = \boldsymbol{\theta}_0^{m'} \boldsymbol{z}_{it} + \gamma_{i0}^m f_{t0} + (1 + \boldsymbol{\theta}_0^{q'} \boldsymbol{z}_{it} + \gamma_{i0}^q f_{t0})\epsilon_{it}, \quad (6)$$

where the superscript, m is used for the parameters of covariates that cause location (mean) shifts of the dependent variable, and q for the parameters that cause shifts in scale/shape of the dependent variable. The idiosyncratic error ϵ_{it} is assumed to be $i.i.d.$ with zero-mean, and independent of all other terms. We require $(1 + \boldsymbol{\theta}_0^{q'} \boldsymbol{z}_{it} + \gamma_{i0}^q f_{t0}) > 0$, in order to satisfy the co-monotonicity condition to avoid quantile crossing. Then, we have the following τth quantile representation of (6):

$$y_{it} = \boldsymbol{\theta}_0'(\tau)\boldsymbol{z}_{it} + \gamma_{i0}(\tau) f_{t0} + \mu_{it}(\tau), \quad (7)$$

where $\boldsymbol{\theta}_0(\tau) = \boldsymbol{\theta}_0^m + \boldsymbol{\theta}_0^q Q_\epsilon(\tau)$ and similarly for $\gamma_{i0}(\tau)$, with $Q_\epsilon(\tau)$ being the τth quantile of ϵ_{it}. Notice that (7) is of the same form as (2), where the parameters are allowed to be quantile dependent. We can also rewrite (6) as

$$y_{it} = \boldsymbol{\theta}_0^{m'} \boldsymbol{z}_{it} + \boldsymbol{\gamma}_{i0}^m f_{t0} + e_{it} \quad (8)$$

where $e_{it} = (1 + \boldsymbol{\theta}_0^{q'} \boldsymbol{z}_{it} + \gamma_{i0}^q f_{t0})\epsilon_{it} = \mu_{it}(\tau) + (\boldsymbol{\theta}_0(\tau) - \boldsymbol{\theta}_0^m)' \boldsymbol{z}_{it} + (\boldsymbol{\gamma}_{i0}(\tau) - \boldsymbol{\gamma}_{i0}^m)' \boldsymbol{f}_{t0}$. Since ϵ_{it} is assumed to be $i.i.d.$, e_{it} is both cross sectionally and serially uncorrelated. Thus, (6) or (8) can be seen as the quantile extension of the dynamic panel data model with interactive effects studied by Moon and Weidner (2017). This suggests that all the parameters including both unobserved factors and loadings can be estimated consistently by the IPC.

In general, we can rewrite the model (1) as

$$\begin{aligned} y_{it} &= \mathbb{E}[\boldsymbol{\theta}_0(u_{it})]' \boldsymbol{z}_{it} + \mathbb{E}[\boldsymbol{\gamma}_{i0}(u_{it})]' \boldsymbol{f}_{t0} + [\boldsymbol{\theta}_0(u_{it}) - \mathbb{E}[\boldsymbol{\theta}_0(u_{it})]' \boldsymbol{z}_{it} + [\boldsymbol{\gamma}_{i0}(u_{it}) - \mathbb{E}[\boldsymbol{\gamma}_{i0}(u_{it})]' \boldsymbol{f}_{t0} \\ &= \boldsymbol{\theta}_0^{m'} \boldsymbol{z}_{it} + \boldsymbol{\gamma}_{i0}^{m'} \boldsymbol{f}_{t0} + e_{it}, \end{aligned} \quad (9)$$

where $\boldsymbol{\theta}_0^m \equiv \mathbb{E}[\boldsymbol{\theta}_0(u_{it})]$, $\boldsymbol{\gamma}_{i0}^m \equiv \mathbb{E}[\boldsymbol{\gamma}_{i0}(u_{it})]$ and $e_{it} \equiv [\boldsymbol{\theta}_0(u_{it}) - \mathbb{E}[\boldsymbol{\theta}_0(u_{it})]' \boldsymbol{z}_{it} + [\boldsymbol{\gamma}_{i0}(u_{it}) - \mathbb{E}[\boldsymbol{\gamma}_{i0}(u_{it})] \boldsymbol{f}_{t0}$. Then, (9) has exactly the same representation as (8). By assuming u_{it} to be $i.i.d. U[0,1]$ and distributed independently of $\boldsymbol{z}_{it}$ and $\boldsymbol{f}_{t0}$, the e_{it} will be cross-sectionally and serially uncorrelated.

literature, see Bai and Ng (2002), Ahn and Horenstein (2013). Moreover, as is a common practice in the factor model literature, we maintain the standard normalization conditions needed for the identification of factor and factor loadings (e.g. Bai (2009)). As a result, the factors are estimated up to some rotation.

[4]We may also follow the recent studies by Chen et al. (2021b) and Ando and Bai (2020), and jointly estimate the factors and $(\hat{\boldsymbol{\theta}}', \hat{\gamma}_1', \ldots, \hat{\gamma}_N')'$ at each quantile in an iterative manner. However, this would bring more technical complexities in the dynamic case. We leave this topic for a future study.

Hence, the IPC and the corresponding asymptotic theories in Moon and Weidner (2017) can be applied to estimating $\boldsymbol{f}_{t0}$, though we need to impose some additional regularity conditions on e_{it}.

3 Asymptotic Theory

3.1 Consistency

To derive the consistency of the quantile estimator, we impose the following assumptions:

Assumption 1. *Let M be a generic, finite positive constant.*

(i) The random process $\{\boldsymbol{z}_{it}, \boldsymbol{f}_{t0}\}_{t=1}^{\infty}$ is $\alpha-$mixing and the mixing coefficient $\alpha_i(j)$ satisfies $\sup_i \sum_{j=1}^{\infty} \alpha_i(j)^{\delta/(\delta+4)} \leq \infty$, for some $\delta > 0$

(ii) Let $\boldsymbol{X}_{it} = (\boldsymbol{z}'_{it}, \boldsymbol{f}'_{t0})'$, then $\mathbb{E}\|\boldsymbol{X}_{it}\|^{4+\delta} \leq M$ for δ defined (i) and all $i, t \geq 1$. Moreover, as $T \to \infty$, $\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{f}_{t0}\boldsymbol{f}'_{t0} \xrightarrow{p} \boldsymbol{\Sigma}_F$ where $\boldsymbol{\Sigma}_F$ is an $r \times r$ positive definite matrix.

(iii) The Factor loadings for the mean regression (9) satisfy $\|\boldsymbol{\gamma}_{i0}^{m}\| \leq M$ for all i, and as $N \to \infty$, $\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\gamma}_{i0}^{m}\boldsymbol{\gamma}_{i0}^{m\prime}$ converges to an $r \times r$ positive definite matrix $\boldsymbol{\Sigma}_{\Gamma_m}$.

(iv) The error terms μ_{it} defined in (2) are cross sectionally and serially independent conditional on $\boldsymbol{X}_{it}$. The error terms e_{it} defined in (9) are cross sectionally and serially uncorrelated and satisfy $E(e_{it}) = 0$ and $E|e_{it}|^8 \leq M$ uniformly in $i, t \geq 1$. Moreover, e_{it} is uncorrelated with $\boldsymbol{X}_{is}$ for all $s \leq t$, and $N^{-1/2}\sum_{i=1}^{N}[e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})]$ is sub-Gaussian for all $s, t \geq 1$.

(v) For any given quantile, the parameters $\boldsymbol{\theta}_0$ and $\boldsymbol{\gamma}_{i0}$ are interior points of real compact sets $\boldsymbol{\Theta}$ and $\mathcal{A}$, respectively.

(vi) Let $\mathrm{g}_{it}(\cdot)$ and $\mathrm{g}_{it}(\cdot|\boldsymbol{X}_{it})$ denote the unconditional and conditional density of μ_{it} given $\boldsymbol{X}_{it}$, and ρ_{it} the smallest eigenvalue of $\mathbb{E}[\mathrm{g}_{it}(0|\boldsymbol{X}_{it})\boldsymbol{X}_{it}\boldsymbol{X}'_{it}]$, then there exists $\varkappa > 0$ such that $\rho_{it} > \varkappa$ for all i, t.

(vii) $N/T^2 \to 0$ as $N, T \to \infty$ and the bandwidth in (4) satisfies $h \to 0$.

The mixing condition in Assumption 1(i) is standard in the nonlinear panel data literature (e.g., Fernández-Val and Weidner (2016)). It places restrictions on the temporal dependence of $\mathbf{z}_{it}$ and $\mathbf{f}_{t0}$, enabling us to apply moment inequalities developed in the literature, see Hahn and Kuersteiner (2011) and Chen (2021). Assumptions 1(ii)–(iv) are standard for deriving consistent IPC estimator of factors (up to a rotation). As in Moon and Weidner (2017), we exclude serial correlation in both μ_{it} and e_{it} in Assumption 1(iv) since we now study a dynamic model. Cross-sectional correlation is also excluded for simplicity. We impose a sub-Gaussian condition for establishing uniform consistency of factor estimation. In contrast to the classic quantile regression, we need to impose the moment restrictions on e_{it} (and μ_{it}),[5] which exclude some heavy tailed distributions such as

[5] Our procedure is shown to be robust to various idiosyncratic error distributions, including the Normal, t and χ^2 distributions, see Monte Carlo simulation evidence in Section 4.

the Cauchy distribution, since we employ the IPC to estimate unobserved factors in the first step. Similar assumptions are imposed by Harding et al. (2020) and Chen (2021).[6] Assumption 1(v) is a standard condition in the M-estimation literature for establishing consistency (e.g., Newey and McFadden (1994)). Assumption 1(vi) is a standard identification condition in the quantile regression (e.g., Assumption A4 of Galvao and Kato (2016)). The restriction on the bandwidth in Assumption 1(vii) is used to restrict the estimation error obtained from the smoothed quantile regression instead of the original unsmoothed one. The required relative size of N over T is imposed here for technical purposes but easily satisfied.

We now establish the consistency of the quantile estimator of $\boldsymbol{\theta}_0$ in Theorem 1.

Theorem 1. *Under Assumption 1, the quantile estimator $\hat{\boldsymbol{\theta}}$, defined in (5), is consistent for $\boldsymbol{\theta}_0$, i.e., $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = o_p(1)$ as $N, T \to \infty$.*

Remark 1. In Lemma 2 in the Appendix S1, we establish the consistency of the estimated factor loadings, $\hat{\boldsymbol{\gamma}}_i$ up to a rotation matrix as in the mean case (e.g., Bai (2003)). One important implication is that although the estimation of $\boldsymbol{\theta}_0$ relies upon consistent estimation of unobserved factors via the IPC in the first step, the corresponding estimation error is asymptotically negligible and does not affect the consistency of $\hat{\boldsymbol{\theta}}$.

3.2 Asymptotic Distribution

To derive the asymptotic distribution of the estimators, we add the following assumptions.

Assumption 2. *(i) $\boldsymbol{X}_{it}$ is uniformly bounded in probability for all $i, t \geq 1$.*

(ii) Letting q be an integer satisfying $q \geq 8$, $\mathrm{g}_{it}(u)$ and $\mathrm{g}_{it}(u|\boldsymbol{X}_{it})$ have up to $(q+2)$-th order derivatives. Denote $\mathrm{g}_{it}^{(j)}(u) = \partial^j \mathrm{g}_{it}(u)/\partial u^j$, $\mathrm{g}_{it}^{(j)}(u|\boldsymbol{X}_{it}) = \partial^j \mathrm{g}_{it}(u|\boldsymbol{X}_{it})/\partial u^j$ for $j = 1, \ldots, q+2$. Then, for each $1 \leq j \leq q+2$, $|\mathrm{g}_{it}^{(j)}(u)|$ and $|\mathrm{g}_{it}^{(j)}(u|\boldsymbol{X}_{it})|$ are uniformly bounded for all i, t.

(iii) $\int_{-1}^{1} k(u)du = 1, \int_{-1}^{1} k(u)u^j du = 0$ for $j = 1, \ldots, q-1$ and $\int_{-1}^{1} k(u)u^q du \neq 0$, where q was defined in Assumption 2(ii).

(iv) $\sqrt{N/T} \to \pi > 0$ as $N, T \to \infty$, and $h \asymp T^{-c}$, where $1/q < c < 1/6$.

(v) Let

$$\underset{(p+1)\times r}{\boldsymbol{\Xi}_{iT}} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\mathrm{g}_{it}(0|\boldsymbol{X}_{it})\boldsymbol{z}_{it}\boldsymbol{f}_{t0}'], \quad \underset{r\times r}{\boldsymbol{\Pi}_{iT}} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\mathrm{g}_{it}(0|\boldsymbol{X}_{it})\boldsymbol{f}_{t0}\boldsymbol{f}_{t0}'], \quad \underset{(p+1)\times r}{\boldsymbol{\Phi}_{iT}} = \boldsymbol{\Xi}_{iT}\boldsymbol{\Pi}_{iT}^{-1}, \tag{10}$$

$$\underset{(p+1)\times 1}{\boldsymbol{w}_{it}} = \boldsymbol{z}_{it} - \boldsymbol{\Phi}_{iT}\boldsymbol{f}_{t0}, \qquad \underset{(p+1)\times(p+1)}{\boldsymbol{\Delta}_{NT}} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\mathrm{g}_{it}(0|\boldsymbol{X}_{it})\boldsymbol{w}_{it}\boldsymbol{w}_{it}'], \tag{11}$$

$$\underset{(p+1)\times(p+1)}{\boldsymbol{\Phi}_{NT}} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\Lambda}_{tN}\left(\frac{\boldsymbol{\Gamma}_0^{m\prime}\boldsymbol{\Gamma}_0^{m}}{N}\right)^{-1}\boldsymbol{\gamma}_{i0}^{m}\boldsymbol{z}_{it}', \qquad \underset{(p+1)\times(p+1)}{\boldsymbol{D}_{NT}} = \frac{1}{NT}\sum_{i=1}^{N}\tilde{\boldsymbol{z}}_i'\tilde{\boldsymbol{z}}_i, \tag{12}$$

[6] Chen (2021) imposes moment restrictions on idiosyncratic errors for regressors instead of the dependent variable.

where

$$\underset{(p+1)\times r}{\boldsymbol{\Lambda}_{tN}} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\mathrm{g}_{it}(0|\boldsymbol{X}_{it})\boldsymbol{w}_{it}]\boldsymbol{\gamma}_{i0}', \quad \underset{T\times(p+1)}{\tilde{\boldsymbol{z}}_i} = \boldsymbol{M}_{F_0}\boldsymbol{z}_i - \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{M}_{F_0}\boldsymbol{z}_j\boldsymbol{\gamma}_{i0}^{m\prime}(\frac{\boldsymbol{\Gamma}_0^{m\prime}\boldsymbol{\Gamma}_0^m}{N})^{-1}\boldsymbol{\gamma}_{j0}^m,$$

$$\underset{N\times r}{\boldsymbol{\Gamma}_0^m} = \left(\boldsymbol{\gamma}_{10}^m, \boldsymbol{\gamma}_{20}^m, \ldots, \boldsymbol{\gamma}_{N0}^m\right)', \quad \underset{T\times T}{\boldsymbol{M}_{F_0}} = \boldsymbol{I}_T - \boldsymbol{F}_0(\boldsymbol{F}_0'\boldsymbol{F}_0)^{-1}\boldsymbol{F}_0', \quad \underset{T\times r}{\boldsymbol{F}_0} = (\boldsymbol{f}_{10}, \boldsymbol{f}_{20}, \ldots, \boldsymbol{f}_{T0})'.$$

Then, $\boldsymbol{\Pi}_{iT}$ *is positive definite for all* i *and* $T \geq 1$*, and the probability limit* $\boldsymbol{\Phi} = \operatorname{plim}_{N,T\to\infty}\boldsymbol{\Phi}_{NT}$ *exists. Moreover, the limits* $\boldsymbol{\Lambda}_t = \lim_{N\to\infty}\boldsymbol{\Lambda}_{tN}$ *(for each* t*),* $\boldsymbol{\Delta} = \lim_{N,T\to\infty}\boldsymbol{\Delta}_{NT}$*, and* $\boldsymbol{D} = \lim_{N,T\to\infty}\boldsymbol{D}_{NT}$ *exist, and* $\boldsymbol{\Delta}$ *and* $\boldsymbol{D}$ *are positive definite matrices.*

Assumptions 2(ii)–(iii) impose restrictions on the smoothness of the density function of μ_{it} and on the kernel function $k(u)$. They are similar to Assumptions A5–A7 in Galvao and Kato (2016), except that we need an eighth order kernel due to the presence of the common factors, $\boldsymbol{f}_{t0}$ instead of a fourth order kernel. Assumption 2(v) is analogous to the standard rank condition in quantile regression, which guarantees that the variance of the estimator is well-defined.

Let $w_{it,k}$ be the k-th element of $\boldsymbol{w}_{it}$ and $\underset{1\times r}{\Phi_{ik}}$ be the k-th row of $\boldsymbol{\Phi}_{iT}$, $1 \leq k \leq p+1$. Then, we define:

$$\underset{r\times r}{\boldsymbol{\Psi}_{ik}} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\mathrm{g}_{it}^{(1)}(0|\boldsymbol{X}_{it})w_{it,k}\boldsymbol{f}_{t0}\boldsymbol{f}_{t0}'], \quad \underset{r\times r}{\boldsymbol{\Psi}_{tk}} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\mathrm{g}_{it}^{(1)}(0|\boldsymbol{X}_{it})w_{it,k}]\boldsymbol{\gamma}_{i0}\boldsymbol{\gamma}_{i0}',$$

$$\underset{r\times r}{\boldsymbol{\Theta}_{tk}} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{g}_{it}(0)\boldsymbol{\gamma}_{i0}\Phi_{ik}, \qquad 1 \leq k \leq p+1. \tag{13}$$

We now establish the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ in Theorem 2.

Theorem 2. *Suppose that Assumptions 1 and 2 hold. Then, as* $N, T \to \infty$*, we have:*

$$\sqrt{NT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{\Delta}^{-1}\left\{\boldsymbol{D}^{-1}\left[\pi(\boldsymbol{a}_1 + \boldsymbol{a}_2) + \pi^{-1}\boldsymbol{a}_3\right] + \pi\boldsymbol{b} + \pi^{-1}\boldsymbol{c}\right\}, \ \boldsymbol{\Delta}^{-1}\boldsymbol{V}\boldsymbol{\Delta}^{-1}), \tag{14}$$

where $\boldsymbol{b} = \boldsymbol{b}_1 + \boldsymbol{b}_2 + \boldsymbol{b}_3$, $\boldsymbol{c} = \boldsymbol{c}_1 + \boldsymbol{c}_2$, $\boldsymbol{a}_j = [a_{1,j}, \ldots, a_{p+1,j}]', j = 1, 2, 3,$ *and*

$$a_{k,1} = \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s<t}^{T}E[z_{it,k}\boldsymbol{f}_{t0}'\boldsymbol{\Sigma}_F^{-1}\boldsymbol{f}_{s0}e_{is})],$$

$$a_{k,2} = \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{(\boldsymbol{z}_{i,k} - \breve{\boldsymbol{z}}_{i,k})'\boldsymbol{F}_0}{T}\boldsymbol{\Sigma}_F\boldsymbol{\Sigma}_{\Gamma_m}\boldsymbol{\gamma}_{i0}^m(\mathbb{E}(e_{it})^2),$$

$$a_{k,3} = \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{z}_{i,k}'\boldsymbol{M}_{\boldsymbol{F}_0}\boldsymbol{\Omega}\boldsymbol{F}_0\boldsymbol{\Sigma}_F\boldsymbol{\Sigma}_{\Gamma_m}\boldsymbol{\gamma}_{i0}^m,$$

for $k = 1, \ldots, p+1$*, where* $\breve{\boldsymbol{z}}_{i,k} = \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\gamma}_{i0}^{m\prime}(\frac{\boldsymbol{\Gamma}_0^{m\prime}\boldsymbol{\Gamma}_0^m}{N})^{-1}\boldsymbol{\gamma}_{j0}^m\boldsymbol{z}_{i,k}$*,* $\boldsymbol{z}_{i,k} = (z_{i1,k}, \ldots, z_{iT,k})'$*,* $\boldsymbol{\Omega} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Omega}_i$*,*

$\boldsymbol{\Omega}_i = diag(\mathbb{E}(e_{i1}^2), \ldots, \mathbb{E}(e_{iT}^2))$, *and*

$$\boldsymbol{b}_1 = -(\tau - 0.5) \lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}[\mathrm{g}_{it}(0|\boldsymbol{X}_{it}) \boldsymbol{w}_{it} \boldsymbol{f}_{t0}' \boldsymbol{\Pi}_{iT}^{-1} \boldsymbol{f}_{t0}]$$

$$\boldsymbol{b}_2 = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s<t}^{T} \mathbb{E}[\mathrm{g}_{it}(0|\boldsymbol{X}_{it}) \boldsymbol{w}_{it} \boldsymbol{f}_{t0}' \boldsymbol{\Pi}_{iT}^{-1} \boldsymbol{f}_{s0} (\tau - \mathbb{1}(\mu_{is} \leq 0))],$$

$$\boldsymbol{b}_3 = [b_{1,3}, \ldots, b_{p+1,3}]' \quad \text{with } b_{k,3} = \frac{\tau(1-\tau)}{2} \lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}[\boldsymbol{f}_{t0}' \boldsymbol{\Pi}_{iT}^{-1} \boldsymbol{\Psi}_{ik} \boldsymbol{\Pi}_{iT}^{-1} \boldsymbol{f}_{t0}],\ k = 1, \ldots, p+1,$$

$$\boldsymbol{c}_1 = -\lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\boldsymbol{\gamma}_{i0}' \boldsymbol{\Sigma}_{\Gamma_m}^{-1} \boldsymbol{\gamma}_{i0}^m \mathbb{E}[\mathrm{g}_{it}(0|\boldsymbol{X}_{it}) c_{it}(\tau) \boldsymbol{z}_{it}] + \boldsymbol{\Phi}_{iT} \boldsymbol{\Sigma}_{\Gamma_m}^{-1} \boldsymbol{\gamma}_{i0}^m \mathbb{E}[\mu_{it}(\tau - \mathbb{1}(\mu_{it} \leq 0))] \right),$$

$$\boldsymbol{c}_2 = [c_{1,2}, \ldots, c_{p+1,2}]' \quad \text{with } c_{k,2} = \lim_{N,T\to\infty} \frac{1}{2} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}[e_{it}^2] \boldsymbol{\gamma}_{i0}^{m'} \boldsymbol{\Sigma}_{\Gamma_m}^{-1} (\boldsymbol{\Psi}_{tk} + 2\boldsymbol{\Theta}_{tk}) \boldsymbol{\Sigma}_{\Gamma_m}^{-1} \boldsymbol{\gamma}_{i0}^m,\ 1 \leq k \leq p+1,$$

$$\boldsymbol{V} = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}[\boldsymbol{v}_{it} \boldsymbol{v}_{it}'] \ \text{ with } \boldsymbol{v}_{it} = \boldsymbol{\Phi} \boldsymbol{D}^{-1} \tilde{\boldsymbol{z}}_{it}' e_{it} + (\tau - \mathbb{1}(\mu_{it} \leq 0)) \boldsymbol{w}_{it} - \boldsymbol{\Lambda}_{tN} (\frac{\boldsymbol{\Gamma}_0^{m'} \boldsymbol{\Gamma}_0^m}{N})^{-1} \boldsymbol{\gamma}_{i0}^m e_{it},$$

where $\tilde{\boldsymbol{z}}_{it}$ *is the* t*-th row of* $\tilde{\boldsymbol{z}}_i$*.*

Remark 2. From (14), we observe that the quantile estimator, $\hat{\boldsymbol{\theta}}$ suffers from asymptotic biases. The bias term $\boldsymbol{b}$ ($\boldsymbol{c}$) arises mainly from the estimation of factor loadings (factors). The convergence rates for the estimated factors (time effects) and factor loadings (individual effects) are $\sqrt{N}$ and $\sqrt{T}$, respectively, with the leading bias terms of orders $1/N$ and/or $1/T$ (e.g., Theorems 1 and 2 in Bai (2003)). These rates are slower than that of the main parameters, which is $\sqrt{NT}$. As a result, the biases of estimated factors and loadings do not affect consistency of $\hat{\boldsymbol{\theta}}$ as $N, T \to \infty$, but they do not disappear after the multiplication of $\sqrt{NT}$. This leads to asymptotic biases of orders $\sqrt{T/N}$ or $\sqrt{N/T}$. Moreover, the bias term $\boldsymbol{b}_2$ can be seen as an extension of the Nickell Bias (Nickell (1981)), which arises from the lagged dependent regressor, $y_{i,t-1}$. Without $y_{i,t-1}$ in the model, $\boldsymbol{b}_2$ is zero since $\boldsymbol{w}_{it}$ is no longer correlated with μ_{is} ($s < t$) under our assumptions.

Remark 3. Since factors and parameters $\boldsymbol{\theta}_0^m$ are estimated iteratively in the first step, our quantile estimator is subject to an additional bias. As shown in the proof of Theorem 2 in Appendix A, the expansion of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ consists of the term, $\boldsymbol{a} = \boldsymbol{\Delta}_{NT}^{-1} \boldsymbol{\Phi}_{NT} \sqrt{NT} (\boldsymbol{\theta}_0^m - \hat{\boldsymbol{\theta}}^m)$, which arises from the factor estimation error (see Lemma 8). While this term can be absorbed in the conditional mean regression (see Proposition A.2 in Bai (2009)), $\boldsymbol{\theta}_0^m - \hat{\boldsymbol{\theta}}^m$ is generally different from its quantile counterpart $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ (albeit the same convergence rate) and thus it cannot be absorbed in the quantile regression. The $\boldsymbol{a}_1, \boldsymbol{a}_2$ and $\boldsymbol{a}_3$ are the corresponding asymptotic bias terms relating to $(\boldsymbol{\theta}_0^m - \hat{\boldsymbol{\theta}}^m)$, see Theorem 4.3 in Moon and Weidner (2017).

Remark 4. In the special case where $\boldsymbol{f}_{t0}$ is observed (see Xu et al. (2021)), the asymptotic results can be greatly simplified, since all the bias terms related to factor estimation will disappear. Then, we have:

$$\sqrt{NT}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0] \xrightarrow{d} N(\pi \boldsymbol{\Delta}^{-1}(\boldsymbol{b}_1 + \boldsymbol{b}_2 + \boldsymbol{b}_3),\ \boldsymbol{\Delta}^{-1} \tilde{\boldsymbol{V}} \boldsymbol{\Delta}^{-1}),$$

where $\tilde{\boldsymbol{V}} = \tau(1-\tau)\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\boldsymbol{w}_{it}\boldsymbol{w}_{it}']$. In this case, the quantile estimator only suffers from asymptotic bias arising from estimating individual effects (see Galvao and Kato (2016)).

To estimate the asymptotic variance of the quantile estimator, we first estimate $\boldsymbol{v}_{it}$ as

$$\hat{\boldsymbol{v}}_{it} = \hat{\boldsymbol{\Phi}}\hat{\boldsymbol{D}}^{-1}\hat{\tilde{\boldsymbol{z}}}_{it}'\hat{e}_{it} + (\tau - \mathbb{1}(\hat{\mu}_{it}\le 0))\hat{\boldsymbol{w}}_{it} - \hat{\boldsymbol{\Lambda}}_{tN}(\frac{\hat{\boldsymbol{\Gamma}}^{m\prime}\hat{\boldsymbol{\Gamma}}^{m}}{N})^{-1}\hat{\boldsymbol{\gamma}}_i^{m}\hat{e}_{it},$$

where $\hat{\boldsymbol{\gamma}}_i^m$ and $\hat{e}_{it}$ are the estimated factor loadings and residuals obtained from the first step IPC estimation, $\hat{\tilde{\boldsymbol{z}}}_{it}$ and $\hat{\boldsymbol{D}}$ are constructed using factors and loadings estimated in the first step, and $\hat{\mu}_{it}$ is the regression residual at the given τ-th quantile. Since the estimation of the other quantities are more involved, we follow Galvao and Kato (2016) and propose the following estimators:

$$\hat{\boldsymbol{w}}_{it} = \boldsymbol{z}_{it} - \hat{\boldsymbol{\Xi}}_i\hat{\boldsymbol{\Omega}}_i^{-1}\hat{\boldsymbol{f}}_t \quad \text{with} \quad \hat{\boldsymbol{\Xi}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathcal{K}_{\mathfrak{h}}(\hat{\mu}_{it})\boldsymbol{z}_{it}\hat{\boldsymbol{f}}_t' \quad \text{and} \quad \hat{\boldsymbol{\Omega}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathcal{K}_{\mathfrak{h}}(\hat{\mu}_{it})\hat{\boldsymbol{f}}_t\hat{\boldsymbol{f}}_t',$$

$$\hat{\boldsymbol{\Lambda}}_{tN} = \frac{1}{N}\sum_{i=1}^{N}\mathcal{K}_{\mathfrak{h}}(\hat{\mu}_{it})\hat{\boldsymbol{w}}_{it}\hat{\boldsymbol{\gamma}}_i', \quad \hat{\boldsymbol{\Phi}} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\boldsymbol{\Lambda}}_{tN}(\frac{\hat{\boldsymbol{\Gamma}}^{m\prime}\hat{\boldsymbol{\Gamma}}^{m}}{N})^{-1}\hat{\boldsymbol{\gamma}}_i^{m}\boldsymbol{z}_{it}',$$

where $\mathcal{K}_{\mathfrak{h}}(\hat{\mu}_{it})$ is the kernel estimator for the probability density of the regression residual around 0 with $\mathcal{K}_{\mathfrak{h}}(\cdot) = \mathcal{K}(\cdot/\mathfrak{h})/\mathfrak{h}$, $\mathcal{K}(\cdot)$ is a kernel function and $\mathfrak{h}$ is a bandwidth.[7] Based on these estimates, we construct an estimate of $\boldsymbol{\Delta}$ as

$$\hat{\boldsymbol{\Delta}} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathcal{K}_{\mathfrak{h}}(\hat{\mu}_{it})\hat{\boldsymbol{w}}_{it}\hat{\boldsymbol{w}}_{it}',$$

Then, we estimate the variance of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ as

$$\hat{\boldsymbol{V}} = \hat{\boldsymbol{\Delta}}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\boldsymbol{v}}_{it}\hat{\boldsymbol{v}}_{it}'\hat{\boldsymbol{\Delta}}^{-1}. \tag{15}$$

3.3 Bias Correction

Theorem 2 implies that the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ is not centered. Such biases in finite samples can be substantial if T or N is not sufficiently large. To get rid of the bias, we have two options: the analytical bias correction by Hahn and Kuersteiner (2011) and the Split Panel Jackknife (SPJ hereafter) by Dhaene and Jochmans (2015). While both approaches can be shown to remove the asymptotic bias effectively, SPJ enjoys a main advantage such that it is easy to implement and more effective when the sample size is small (as the estimation of asymptotic bias in the analytical bias correction can be rather imprecise in small samples).

We therefore propose using SPJ to correct the bias of the quantile estimator. Define the subpanel

[7] The kernel function and bandwidth are not necessarily the same as those used for parameter estimation in (4).

index sets by

$$\mathcal{T}_1 = \{1, \ldots, \lfloor (T+1)/2 \rfloor\}, \quad \mathcal{T}_2 = \{\lfloor T/2 \rfloor + 1, \ldots, T\},$$
$$\mathcal{N}_1 = \{1, \ldots, \lfloor (N+1)/2 \rfloor\}, \quad \mathcal{N}_2 = \{\lfloor N/2 \rfloor + 1, \ldots, N\},$$

where $\lfloor \cdot \rfloor$ takes an integer part of any number. Also denote $\mathcal{T}_0 = \mathcal{T}_1 \cup \mathcal{T}_2$ and $\mathcal{N}_0 = \mathcal{N}_1 \cup \mathcal{N}_2$. For any given quantile level, τ, let $\hat{\boldsymbol{\theta}}_{(js)}$ be the estimator of $\boldsymbol{\theta}_0$, defined in (4), obtained using the sample data, $\{(y_{it}, \mathbf{x}_{it}) : i \in \mathcal{N}_j, t \in \mathcal{T}_s\}$, where $j, s = 0, 1, 2$. Then, the SPJ bias-corrected estimator is given by

$$\hat{\boldsymbol{\theta}}_{bc} = 3\hat{\boldsymbol{\theta}}_{(00)} - \frac{1}{2}[\hat{\boldsymbol{\theta}}_{(01)} + \hat{\boldsymbol{\theta}}_{(02)}] - \frac{1}{2}[\hat{\boldsymbol{\theta}}_{(10)} + \hat{\boldsymbol{\theta}}_{(20)}]. \tag{16}$$

To derive its asymptotic property, we follow Fernández-Val and Weidner (2016) and impose an unconditional homogeneity condition.

Assumption 3. *The sequence, $\{\boldsymbol{y}_{it}, \boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_{t0}, 1 \le t \le T, 1 \le i \le N\}$ is identically distributed across i and strictly stationary across t, for each N, T .*

Theorem 3. *Under Assumptions 1-3 and as $N, T \to \infty$, then*

$$\sqrt{NT}(\hat{\boldsymbol{\theta}}_{bc} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0},\ \boldsymbol{\Delta}^{-1}\boldsymbol{V}\boldsymbol{\Delta}^{-1})). \tag{17}$$

4 Monte Carlo Simulation

4.1 The Simulation Design

We investigate the finite sample performance of the proposed two-step quantile estimator. We generate the data by

$$\begin{aligned} y_{it} &= (0.5 + 0.2u_{it})y_{i,t-1} + (1 + 0.2u_{it})x_{it,1} + (2 + 0.2u_{it})x_{it,2} \\ &\quad + (\gamma_{1i} + 0.2u_{it})f_{1,t} + (\gamma_{2i} + 0.2u_{it})f_{2,t} + F^{-1}(u_{it}) \end{aligned} \tag{18}$$

$$x_{it,k} = \Gamma_{1i,k}f_{1,t} + \Gamma_{2i,k}f_{3,t} + v_{it,k}, \quad k = 1, 2, \tag{19}$$

for $i = 1, \ldots, N$ and $t = -49, -48, \ldots, T$,[8] where u_{it} is generated from $i.i.d.U[0, 1]$ and $F^{-1}(\cdot)$ is the inverse of the cumulative distribution function. To check the robustness of the proposed estimator, we consider the different distributions of idiosyncratic errors such as $N(0, 1)$, $t(4)$ and $\chi^2(3)$ distributions. We set the number of unobserved factors in both equations for y_{it} and $\boldsymbol{x}_{it}$ at 2, but allow $\boldsymbol{x}_{it}$ to be influenced by a different factor $f_{3,t}$. All the factors are generated by an AR(1) process:

$$f_{r,t} = \phi_{f_r} f_{r,t-1} + \xi_{f_{rt}},\ \ r = 1, 2, 3;\ \ t = -49, -48, \ldots, T, \text{[9]}$$

[8]We discard the first 50 observations as a burn-in sample.

[9]The first 50 observations are discarded as a burn-in sample.

with $\phi_{f_r} = 0.5$ and $\xi_{f_{rt}} \sim i.i.d.N(0, 1-\phi_{f_r}^2)$. The loadings are generated as $i.i.d.N(0.5, 0.5)$. We allow $\boldsymbol{v}_{it}$ to be serially correlated via the following AR (1) process:

$$v_{itk} = \phi_v v_{i,t-1,k} + \xi_{v_{ik}},\ k = 1, 2;\ i = 1, 2, \ldots, N;\ t = -49, -48, \ldots, T,$$

where $\phi_v = 0.5$, and $\xi_{v_{ik}} \sim i.i.d.N(0, 1-\phi_{v_{ik}}^2)$.

In addition to our proposed two-step estimator in (5) and its bias-corrected estimator in (16), we also consider alternative estimators for comparison. These estimators are listed below:[10]

1. TS: Our proposed two-step estimator, where we apply the IPC to (18) and obtain consistent estimates of factors $\hat{\boldsymbol{f}}_t, t = 1, 2, \ldots, T$ in the first step.[11] In the second step, we plug in the estimated factors and run the smoothed quantile regression in (5) to estimate the main parameters. In the smoothed quantile regression, we use the kernel function (see also Muller et al. (1984)):

$$k(z) = \mathbb{1}\{|z| \leq 1\} \times \frac{3465}{8192}(7 - 105z^2 + 462z^4 - 858z^6 + 715z^8 - 221z^{10}), \tag{20}$$

   and the bandwidth $1.4(NT)^{-1/13}$.[12]

2. TS_{bc}: The SPJ Bias-corrected estimator given by (16).

3. CCE: The estimator by Harding et al. (2020), uses cross-section averages of dependent and independent variables and their lags to approximate the factors.

4. PCA: The estimator by Chen (2021) estimates the factors by applying PCA to $\boldsymbol{x}_{it}$. This estimator also suffers from asymptotic bias. Hence, we report results for the SPJ bias-corrected version.

To evaluate the finite sample performance of the above estimators, we report their biases and RMSEs[13] over 1000 replications for each pair of (N, T) with $N, T = 30, 50, 100$. For our proposed estimators we also report their size and power of the t-test. Notice that the variance estimation defined in (15) involves the selection of an additional kernel function and bandwidth. While not

---

[10] We also consider the standard quantile regression estimator and the IVQR estimator proposed by Galvao (2011). Those estimation results are much worse than those of our proposed estimator and are available upon request.

[11] To determine the number of factors, we apply the following criterion proposed in Bai and Ng (2002):

$$\hat{r} = \underset{0 \leq m \leq r_{max}}{\operatorname{argmin}}\ IC(m) = \underset{0 \leq m \leq r_{\max}}{\operatorname{argmin}}\ \ln\left(\frac{1}{NT}\sum_{i=1}^{N} \hat{e}_{it}^2(m)\right) + m\frac{N+T}{NT}\ln(\min[N, T]),$$

where $\hat{e}_{it}(m)$ are the mean regression residuals obtained by assuming the number of factors is m.

[12] Bandwidth selection is always an important problem in nonparametric estimation and remains an open problem in quantile panel regression (see e.g. Galvao and Kato (2016)). Here we simply use the bandwidth $1.4(NT)^{-1/13}$ which satisfies Assumption 2. The simulation results confirm its satisfactory performance. We also conduct simulation for various other choices of bandwidth, the results are generally satisfactory and are available upon request.

[13] When reporting the results, we multiply them by 100 for convenience.

necessary, for simplicity, we employ the the same kernel function and bandwidth as those employed in the estimation of parameters.

4.2 Simulation Results

To save space, we only report the simulation results for $t(4)$ error distribution.[14] The simulation results for biases are presented in Table 1. Biases of the proposed TS and TS_{bc} estimators for (ψ, β_1, β_2) are much smaller than those of the CCE and PCA estimators for all the quantile levels and sample sizes. These results confirm that our proposed two-step estimation procedure is more reliable and robust than the CCE estimation by Harding et al. (2020) and the PCA estimation by Chen (2021), both of which require restrictive assumptions on the data generating process for the independent variables.[15] In particular, biases of TS_{bc} are the smallest and almost negligible across all quantiles and sample sizes, which confirms that the SPJ procedure can remove biases effectively even in small samples with $N, T = 30$.

The RMSE results for all the estimators are reported in Table 2. Again, RMSEs of the proposed TS and TS_{bc} are much smaller than those of CCE and PCA, suggesting that our proposed estimators are likely to be more efficient. As N or T increases, RMSEs become smaller, which is in line with Theorem 2 that the convergence rate of the estimator is $\sqrt{NT}$. The RMSEs are slightly smaller at the median ($\tau = 0.5$) than at the lower and upper quantiles, $\tau = 0.2$ and 0.8, which is also consistent with the usual findings in quantile regression as there are more data points at the median (e.g. Xu et al. (2021)). Notice that RMSEs of TS_{bc} are larger than those of TS. The inflated RMSEs of SPJ estimator has been also observed in many studies (e.g., Moon and Weidner (2017), Galvao and Kato (2016)), which reflects the efficiency loss of the SPJ estimator due to the use of only the half data sample for each half panel estimator.

The size and power values of our proposed bias-corrected estimator TS_{bc} are summarised in Table 3. We observe that the size values of the t-tests are generally close to the nominal 5% significance level. Although the power values at the 0.2 and 0.8 quantiles are lower than those at the median, they approach 1 quickly as sample size (either N or T) increases.

We also report the simulation results for error distributions of $N(0, 1)$ and $\chi^2(3)$ in Tables C1-C6 of Appendix S2. The results are very similar to those reported here, which provides support for the robust performance of the proposed estimators. In sum, we establish that the finite sample performance of the proposed (bias-corrected) estimator is quite satisfactory in almost all cases considered.

[14] We provide the complete simulation results for $N(0, 1)$ and $\chi^2(3)$ in Appendix S2. Overall, the results are qualitatively similar to those reported here.

[15] When such assumptions fail, both CCE and PCA estimators suffer from remaining endogeneity issue which may lead to large biases. In static models, we can still obtain consistent estimates even if such assumptions fail (see Chen et al. (2021a), Cui et al. (2021)). However, this does not apply to dynamic models, due to the correlation between lagged dependent variable and unapproximated factors.

Table 1: Bias ($\times 100$) of Estimators with $t(4)$ Error

			$\tau = 0.25$				$\tau = 0.5$				$\tau = 0.75$			
			TS	TS_{bc}	CCE	PCA	TS	TS_{bc}	CCE	PCA	TS	TS_{bc}	CCE	PCA
$N = 30$	$T = 30$	ψ	1.56	0.04	1.61	0.49	-0.19	0.30	-0.08	-1.00	-2.01	-0.56	-1.92	-2.74
		β_1	4.10	0.25	10.45	11.84	2.85	0.69	7.26	10.84	1.31	0.33	6.07	10.43
		β_2	4.29	0.49	10.63	13.09	2.67	0.71	7.59	9.32	1.07	-0.68	6.42	9.54
	$T = 50$	ψ	1.36	0.66	1.99	-0.47	-0.09	0.06	0.22	-2.02	-1.61	-0.51	-1.60	-3.69
		β_1	2.92	-0.03	9.80	11.11	2.10	0.62	7.37	9.30	0.76	-0.68	6.17	8.54
		β_2	3.14	0.22	9.68	8.68	1.98	0.29	7.00	7.93	0.84	-0.65	5.84	7.11
	$T = 100$	ψ	1.32	0.66	2.14	-0.92	0.05	0.13	0.44	-2.34	-1.33	-0.64	-1.14	-3.95
		β_1	2.47	0.42	9.50	9.97	1.36	0.15	7.36	8.73	-0.23	-0.24	5.80	8.18
		β_2	2.55	0.49	9.75	8.35	1.05	-0.17	6.29	6.68	-0.28	-0.21	5.52	5.9
$N = 50$	$T = 30$	ψ	1.13	0.78	1.66	-0.05	-0.32	0.22	-0.05	-1.39	-1.67	-0.48	-1.91	-2.93
		β_1	3.27	0.66	9.94	11.87	2.17	0.45	7.75	9.04	0.58	-0.91	5.85	8.83
		β_2	3.57	0.71	10.35	9.97	2.27	0.39	7.82	8.92	0.88	-0.72	5.96	8.34
	$T = 50$	ψ	1.17	0.72	1.99	-0.68	-0.10	0.05	0.20	-2.36	-1.51	-0.59	-1.40	-3.63
		β_1	2.42	0.09	9.81	10.71	1.21	-0.52	7.48	8.37	-0.61	0.20	6.04	8.77
		β_2	2.56	0.28	9.76	8.68	1.35	-0.41	6.89	7.59	-0.60	0.60	5.62	7.79
	$T = 100$	ψ	1.02	0.61	2.04	-1.06	-0.01	0.01	0.43	-2.41	-1.15	-0.68	-1.20	-4.21
		β_1	1.99	0.68	9.83	10.21	0.81	-0.13	6.99	7.50	-0.49	-0.03	5.88	8.27
		β_2	1.94	0.50	9.30	7.97	0.97	0.13	6.77	7.30	-0.56	-0.05	5.93	6.37
$N = 100$	$T = 30$	ψ	0.81	0.65	1.75	-0.28	-0.35	0.06	-0.21	-1.74	-1.52	-0.39	-2.13	-3.43
		β_1	2.32	-0.24	10.14	10.98	1.39	-0.59	7.40	9.13	0.16	-0.08	6.74	8.21
		β_2	2.55	-0.03	9.92	10.27	2.00	0.29	7.70	9.22	0.60	0.19	6.78	8.74
	$T = 50$	ψ	0.78	0.43	1.87	-1.06	-0.13	0.090	0.17	-2.31	-1.17	-0.55	-1.40	-3.98
		β_1	1.78	0.42	9.58	10.58	0.60	-0.38	6.99	7.95	-0.09	-0.50	5.47	8.04
		β_2	2.00	0.61	9.41	8.48	0.79	-0.29	7.10	6.66	-0.08	0.00	6.13	7.86
	$T = 100$	ψ	0.82	0.57	2.05	-1.17	-0.07	0.00	0.37	-2.53	-0.95	-0.54	-1.13	-4.08
		β_1	1.36	0.39	9.60	10.57	0.61	-0.21	7.10	9.08	-0.50	-0.44	5.74	8.15
		β_2	1.61	0.64	9.63	7.87	0.58	-0.06	6.80	6.77	-0.48	-0.42	5.50	6.01

Note: The results are based on the DGP specified in Section 4.1. TS stands for our proposed estimator and TS_{bc} is its SPJ bias-corrected version. CCE and PCA are the estimators proposed by Harding et al. (2020) and Chen (2021), respectively.

Table 2: RMSE ($\times 100$) of Estimators with $t(4)$ Error

			$\tau = 0.25$				$\tau = 0.5$				$\tau = 0.75$			
			TS	TS_{bc}	CCE	PCA	TS	TS_{bc}	CCE	PCA	TS	TS_{bc}	CCE	PCA
$N = 30$	$T = 30$	ψ	2.29	3.36	2.59	3.86	1.51	2.85	1.77	3.73	2.63	3.18	2.80	5.09
		β_1	7.54	11.85	13.68	18.53	6.46	10.68	10.48	17.52	6.58	12.00	10.31	18.37
		β_2	7.71	12.29	13.34	19.85	6.41	10.63	10.46	16.21	6.47	11.92	10.77	17.95
	$T = 50$	ψ	1.75	2.01	2.52	2.76	1.07	1.77	1.29	3.22	1.98	2.04	2.18	4.74
		β_1	5.33	7.68	11.78	15.26	4.48	6.75	9.38	13.27	4.84	8.67	9.09	14.38
		β_2	5.63	7.81	11.84	13.64	4.58	7.02	9.15	12.55	4.80	8.23	8.87	13.49
	$T = 100$	ψ	1.53	1.42	2.40	1.95	0.68	1.000	0.97	2.89	1.52	1.35	1.52	4.39
		β_1	3.81	4.85	10.79	13.13	3.11	4.49	8.44	11.32	3.11	5.20	7.63	11.69
		β_2	4.04	4.95	11.00	11.44	2.88	4.44	7.63	9.87	3.17	5.22	7.28	10.14
$N = 50$	$T = 30$	ψ	1.72	2.53	2.34	3.22	1.25	2.11	1.46	3.17	2.07	2.29	2.48	4.45
		β_1	5.79	9.06	12.05	16.67	4.64	7.18	9.65	14.09	4.44	8.40	8.61	14.61
		β_2	5.97	9.06	12.12	15.28	4.86	7.78	9.65	13.86	4.64	8.31	8.95	14.77
	$T = 50$	ψ	1.51	1.67	2.37	2.41	0.80	1.24	1.07	3.13	1.73	1.55	1.84	4.35
		β_1	4.14	6.03	11.06	14.09	3.21	5.05	8.75	11.63	3.23	5.26	8.02	13.68
		β_2	4.31	5.84	11.22	12.54	3.31	5.07	8.19	11.08	3.63	6.12	7.48	12.31
	$T = 100$	ψ	1.19	1.09	2.22	1.81	0.53	0.78	0.88	2.80	1.28	1.07	1.42	4.54
		β_1	3.20	4.02	10.62	12.22	2.05	2.96	7.80	9.93	2.29	3.40	6.94	11.16
		β_2	3.17	3.93	10.10	10.59	2.27	3.00	7.55	9.97	2.27	3.68	7.26	10.31
$N = 100$	$T = 30$	ψ	1.26	1.60	2.30	2.57	0.93	1.39	1.14	2.84	1.74	1.53	2.45	4.34
		β_1	3.86	5.46	11.62	14.89	3.37	4.98	8.34	13.17	3.40	5.84	8.32	12.65
		β_2	4.24	5.98	11.32	13.84	3.47	4.69	9.01	12.77	3.36	5.90	8.56	13.6
	$T = 50$	ψ	1.01	1.11	2.11	1.99	0.54	0.76	0.94	2.75	1.35	1.13	1.68	4.41
		β_1	2.71	3.50	10.31	13.38	2.22	3.27	7.65	10.70	2.60	4.17	6.73	11.64
		β_2	3.28	4.02	10.29	11.52	2.07	2.88	7.94	10.27	2.72	4.03	7.40	11.1
	$T = 100$	ψ	0.94	0.86	2.17	1.66	0.40	0.51	0.70	2.78	1.03	0.78	1.31	4.32
		β_1	2.15	2.67	10.04	12.24	1.52	2.06	7.53	10.72	1.64	2.60	6.44	10.29
		β_2	2.23	2.21	10.10	9.77	1.50	1.99	7.25	8.93	1.76	2.51	6.19	8.3

Note: The results are based on the DGP specified in Section 4.1. TS stands for our proposed estimator and TS_{bc} is its SPJ bias-corrected version. CCE and PCA are the estimators proposed by Harding et al. (2020) and Chen (2021), respectively.

Table 3: Size and Power of the SPJ Bias-Corrected Estimator with $t(4)$ Error

			Size			Power		
			$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$
$N = 30$	$T = 30$	ψ	0.062	0.056	0.046	0.872	0.930	0.864
		β_1	0.044	0.054	0.048	0.292	0.406	0.240
		β_2	0.056	0.056	0.050	0.274	0.378	0.228
	$T = 50$	ψ	0.054	0.058	0.050	0.940	0.966	0.930
		β_1	0.050	0.058	0.067	0.524	0.656	0.514
		β_2	0.046	0.054	0.053	0.466	0.644	0.472
	$T = 100$	ψ	0.060	0.044	0.050	0.988	0.996	0.984
		β_1	0.054	0.062	0.048	0.906	0.958	0.900
		β_2	0.040	0.048	0.044	0.854	0.952	0.848
$N = 50$	$T = 30$	ψ	0.060	0.054	0.056	0.898	0.940	0.936
		β_1	0.048	0.044	0.048	0.398	0.522	0.458
		β_2	0.046	0.052	0.048	0.350	0.520	0.432
	$T = 50$	ψ	0.066	0.055	0.054	0.954	0.964	0.968
		β_1	0.054	0.065	0.044	0.668	0.798	0.724
		β_2	0.038	0.046	0.050	0.632	0.782	0.710
	$T = 100$	ψ	0.062	0.044	0.058	0.992	0.996	0.996
		β_1	0.046	0.056	0.056	0.922	0.974	0.964
		β_2	0.046	0.054	0.044	0.922	0.966	0.962
$N = 100$	$T = 30$	ψ	0.055	0.050	0.040	0.940	0.955	0.935
		β_1	0.047	0.043	0.045	0.660	0.755	0.690
		β_2	0.045	0.060	0.045	0.645	0.765	0.630
	$T = 50$	ψ	0.065	0.044	0.054	0.975	0.990	0.975
		β_1	0.060	0.055	0.045	0.840	0.945	0.820
		β_2	0.045	0.060	0.045	0.885	0.970	0.845
	$T = 100$	ψ	0.050	0.053	0.055	1.000	1.000	0.985
		β_1	0.055	0.060	0.045	0.970	0.990	0.970
		β_2	0.045	0.045	0.053	0.980	0.995	0.935

Note: The results are based on the DGP specified in Section 4.1. The alternative for the power test is $\boldsymbol{\theta}^a = \boldsymbol{\theta} + 0.2$.

5 Empirical Application

During the past few years, the Brexit has provoked a renewed interest among researchers in studying the effects of EU membership on trade. Such studies are important in helping the UK setting up a new trade agreement with the EU. Although Brexit has been done, relevant research is still ongoing.

Historically, empirical findings regarding the effects of EU membership are quite heterogeneous. Carrere (2006) applies a gravity panel data model with fixed effects to the bilateral trade data for 130 countries over the period from 1962 to 1996, and documents that the EU membership can increase intra-EU trade by more than 100%. Baier et al. (2008) report a smaller effect (around 60%), using data for 96 countries from 1960 to 2000, see also Ebell (2016) and Mayer et al. (2019) for similar findings. Other studies find smaller impacts of the EU trade union. Hufbauer and Schott (2009) find that the effect is estimated at 31% for data of all EU countries from 1976 to 2005, while Eicher and Henn (2011) report an estimated effect of 37% using data for 177 countries from 1950 to 2000. Employing a similar dataset and applying the Bayesian Model Averaging to account for model uncertainty, Eicher et al. (2012) find that the EU membership can boost the bilateral trade flows by 51%.

When estimating a gravity model of trade flows, it is a common practice to control for multilateral resistance, because bilateral trade flows depend on bilateral barriers as well as trade barriers across all trading partners (Anderson and Van Wincoop (2003)). However, all of the above studies include only the (country-time) fixed effects in the gravity model, which is likely to produce biased and misleading results (the "gold-medal error" in the terminology of Baldwin and Taglioni (2006)), given that multilateral resistance is unobserved, time-varying, heterogeneous and cross-sectionally correlated (Mastromarco et al. (2016)). Thus, we follow the factor-based approach proposed by Serlenga and Shin (2007, 2013), that can control for multilateral resistance through unobserved time-varying common factors with heterogeneous loadings.

Moreover, it is also important to control dynamics in trade analysis. As argued in De Nardis and Vicarelli (2003) (see also Olivero and Yotov (2012), Anderson and Yotov (2020)), on the one hand, countries that trade a great deal with each other have a tendency to keep doing so as their consumers have grown accustomed to the partner countries' products (habit formation); on the other hand, for countries trading extensively in the past, businesses have set up distribution and service networks in their partner countries, which has led to entrance and exit barriers due to sunk costs. Hence, bilateral trade data are generally highly persistent and past trading volume could affect current trading. Ignoring this in empirical analysis may lead to incorrect inferences.

Finally, Egger and Nigai (2015) find that, in linear gravity models, observable trade costs perform well mostly for country pairs with large bilateral trade flows but poorly for country pairs with relatively small bilateral trade flows. This phenomenon suggests that the observable trade cost measures could also affect the shape of bilateral trade besides the mean, and quantile regression is more appropriate as it allows for a flexible mapping analysis of various factors affecting bilateral trade (see also Baltagi and Egger (2016)).

Hence, to study the effects of EU membership on bilateral trade, we use the modelling strategy proposed in this paper and estimate the following dynamic gravity panel data model of trade flows with common factors:

$$\begin{aligned} export_{it} &= \beta_0 + \rho export_{i,t-1} + \beta_1 gdp_{it} + \beta_2 pop_{it} + \beta_3 sim_{it} \\ &\quad + \beta_4 dis_i + \beta_5 bor_i + \beta_6 lan_i + \beta_7 euro_{it} + \beta_8 eec_{it} + e_{it}, \\ e_{it} &= \boldsymbol{\gamma}_i' \boldsymbol{f}_t + \varepsilon_{it}, \end{aligned} \tag{21}$$

where $export_{it}$ is the logarithm of bilateral export (measured in millions of US dollars) for the i-th country pair at time t, and $export_{i,t-1}$ is its value lagged by one time period. To avoid the "silver medal mistake" (Baldwin and Taglioni (2006)), our model is estimated on uni-directional trade instead of the log of the sum of exports and imports. In model (21), gdp_{it} is the sum of the (logarithm) gross domestic products (GDP) of the two countries in the i-th country pair, and pop_{it} is defined similarly for population. These two variables measure the (economic) size of each country pair and are expected to have positive influence on bilateral trade due to their relationships with production and demand (Martinez-Zarzoso et al. (2009)). sim_{it} is a similarity (or relative development) measure in terms of the size of the country pair and is constructed as $sim_{it} = ln\left[1 - (\frac{gdp_{i,ot}}{gdp_{it}})^2 - (\frac{gdp_{i,dt}}{gdp_{it}})^2\right]$, where $gdp_{i,ot}$ ($gdp_{i,dt}$) is the GDP of the origin (destination) country within the i-th country pair. Its impact on bilateral trade is ambiguous since economic theory suggests that, on the one hand, the more countries differ the more they will trade with each other (Yamarik and Ghosh (2005)), and on the other hand, countries with similar levels of development will have similar preferences and thus be more alike that trade will occur (Linder (1961)). dis_i is the logarithm of the distance between the capital cities of the i-th country pair, and bor_i, lan_i, $euro_{it}$, eec_{it} are binary variables taking the value of one if both countries in the i-th country pair share a border, a common language, the same currency (i.e. euro), and belong to the European Economic Community, respectively. These variables measures geographical/cultural/policy factors that could affect bilateral trade through lowering transportation/transaction costs and financial/cultural risks. Except for dis_i, all the above factors are expected to boost bilateral trade.[16]

We employ the extended dataset analysed by Serlenga and Shin (2007) to cover the longer time period 1960-2018 (59 years) for 380 country-pairs of uni-directional trade out of 14 EU countries (Austria, Belgium-Luxemburg, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain, Sweden and UK) and 6 OECD countries (Australia, Canada, Japan, Norway, Switzerland and the US).

We first estimate model (21) by the mean regression method in Moon and Weidner (2017). The results are reported in the first column of Table 4. As could be seen, the estimated coefficient for the dynamic lagged term is as high as 0.886[17] and statistically significant, confirming the finding in other research that the trade data is highly persistent (e.g. Olivero and Yotov (2012), Comunale et al. (2021)). All the coefficients have the expected signs except for bor_i, which is however not

[16] see Yamarik and Ghosh (2005) for the impacts of various variables on trade.

[17] We also perform various panel unit root tests to check the stationarity of the data, and all the results reject the null of nonstationarity.

significant. Compared to the effect of gdp_{it}, the effect of pop_{it} is much smaller and insignificant (see also Serlenga and Shin (2007)). The estimated results further suggest that a 1% increase in distance between two countries leads to an 8.7% decrease in bilateral trade, and there will be a 9.09%[18] increase in trade if two countries speak the same language. Finally, the positive impact of eec_{it} is higher than that of $euro_{it}$, which is consistent with findings in many existing researches, see e.g. De Nardis and Vicarelli (2003), Mayer et al. (2019).

We then turn to the estimation results from our quantile regression (columns 2-4 of Table 4) that enables the investigation of distributional effects. It is evident that the results are generally different at the different quantile levels considered, which justifies the use of quantile analysis. Moreover, the estimated coefficients have the expected signs at all three quantile levels. Compared with the value from mean regression, the estimated lag coefficient is larger (over 0.9), which further confirms the high persistence of trade data and highlights the importance of dynamic modelling. In line with our expectation, the effects of both gdp and pop are significantly positive now, with the former smaller and the latter higher than their mean regression counterparts. The effect of sim_{it} is positive but only significant at medium and high quantile levels. This implies that in trade prosperous regimes (medium and high quantile levels of trade), the more similar the two countries, the more they are likely to trade (Thursby and Thursby (1987)). Similar to mean regression results, the effect of distance is significantly negative, while the effect of bor_i is significantly positive, in line with our expectation that sharing border could possibly reduce the transportation costs, and thereby boost bilateral trade. The effect of lan) is however negative and becomes insignificant now. This insignificant finding might not be surprising considering that 75% of the countries in our sample have over a half of their population who could speak English.[19] Language related transaction costs (e.g. translation costs) during trade may be relatively small. The estimated effects of $euro_{it}$ and eec_{it} are both significantly positive and of similar values to their mean regression counterparts.

To more clearly illustrate the estimated effects, we plot their values across the different quantiles in Figure 1. The dashed line is the estimated coefficient while the grey area gives the 95% confidence band using the method by Powell (1991). At the 5% significance level, except for the effect of sim_{it} at lower quantiles and lan_i at all quantiles, all the other effects are significantly different from zero. The effects of the lag term and dis_i generally decrease in magnitude as the quantile increases, suggesting that during the trade recession period (lower quantiles), the trade data is more persistent and the adverse effect of dis is larger. These results are sensible considering that, first, bilateral trade during recession period generally implies its necessity, thereby is unlikely to change and more persistent; second, average transportation costs generally decrease as trading volume increases due to the scale effect. It is interesting to see that the coefficients of gdp_{it} and bor_i display a "V" shape,

[18] $= (e^{0.087} - 1) \times 100\%$

[19] English is the first language for Australia, Canada, Ireland, UK, US. According to Wikipedia, over 86% of the population in Denmark (86%), Netherlands (90%), Noway (90%) and Sweden (89%) could speak English; over 51% of the population in Austria (73%), Belgium-Luxemburg (60%), Germany (56%), Greece (51%), Finland (70%) and Switzerland (61%) could speak English. For the rest of the five countries, the English-speaking population is France (39%), Italy (34%), Portugal (27%), Spain (22%), and Japan (15%).

suggesting that during both trade recession and prosperous periods, trade is more likely to happen between countries that have high GDP and are geographically close. These findings may reflect that at trade recession and prosperous periods, only countries with high GDP have the ability to meet partner country's trade demand, and the border effect in reducing transportation costs is more significant than at normal time. Most importantly, while the *euro* effect changes mildly across quantiles,[20] there is a significant drop in the effect of eec_{it} as quantile increases. These results could have the following policy implications: first, the benefit of adopting the same currency is generally stable regardless of the trading/economic situation; second, the benefit of joining the European economic community/free trade agreement may be larger/more significant during trade recession periods.

Owing to dynamic quantile modelling, we are now able to distinguish the long-run effect of eec_{it} on bilateral trade from its short-run effect at different economic regimes. This is an important contribution to the vast literature on studying the benefits of being a EU member country on trade, and the potential impact of Brexit on the UK economy. Our estimated short-run effect of eec_{it} on boosting bilateral trade is between 3.15% and 5.55% across different economic situations. Due to the high persistence of trade data, the effect of eec_{it} is much larger in the long run, ranging from 27.40% at high quantiles to 118.72% at lower quantiles, which covers most estimated effects in the literature, see e.g. Carrere (2006), Mayer et al. (2019). Therefore, our results provide another explanation for the heterogeneity of the estimated effect in the literature, and the effect of EU membership on trade might be heterogeneous by nature under different economic situations. We also report the effect of $euro_{it}$ on trade. Similar to e.g. Bun and Klaassen (2007), Larch et al. (2019), the estimated short-run effect ranges from 1.92% to 3.17%, which becomes much larger in the long run, ranging between 26.41% and 58.83% (see e.g. Glick and Rose (2016)).

In columns 5-7 of Table 4, the estimation results using our proposed method but without dynamics are also reported. As could be seen, most of the estimated results tend to be much higher in magnitude than those from the dynamic model. Notably, the effect of pop_{it} becomes significantly negative at all the three quantile levels, which is contrary to most existing empirical findings in the literature (e.g. Kalirajan (2007)). Moreover, while theory suggests that sharing a common language could possibly lower transaction costs, and thereby boost trading volume (Frankel et al. (1997)), the estimated effect of lan_i however tends to be negative, and significant at the median. Most importantly, the estimated short-run effects of $euro_{it}$ and eec_{it} become much larger. These results highlight the importance of considering dynamics in modelling bilateral trade data. Ignoring the dynamics will, on the one hand, produce less sensible estimation results, and on the other hand, overestimate the effects from currency union and free trade agreement on bilateral trade.

For comparison purposes, we also report the results obtained from using the CCE-type approach by Harding et al. (2020) in columns 8-10 of Table 4. Since bor_i, lan_i, $euro_{it}$, and eec_{it} are binary variables with little time variation, we exclude them when approximating factors to avoid multicollinearity issue (see also Serlenga and Shin (2007)). The estimated dynamic coefficient is

[20]Even though there is a drop at the median, the effect generally fluctuates between 0.025-0.03.

smaller than previous estimations but still large and significant. Similar to the static case, the estimated effect of gdp_{it} is higher than that from our proposed method, but the effect of pop_{it} is significantly negative. Estimates for the coefficients of bor_i and lan_i are both insignificant although high and switching sign across quantiles. Results for both $euro_{it}$ and eec_{it} are significantly positive and higher than those from our proposed method. In general, these results are less sensible and we conjecture that this is due to some assumptions maintained for CCE estimation being possibly violated.

Table 4: Estimation Results for Bilateral Trade Data

	Mean	QR-PC			QR-PC (static)			QR-CCE		
		$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$	$\tau = 0.2$	$\tau = 0.5$	$\tau = 0.8$
ψ	0.886***	0.933***	0.924***	0.889***				0.742***	0.763***	0.745***
	(0.003)	(0.005)	(0.004)	(0.004)				(0.011)	(0.009)	(0.010)
gdp	0.149***	0.070***	0.063***	0.088***	3.637***	2.980***	2.595***	0.377***	0.357***	0.446***
	(0.011)	(0.015)	(0.012)	(0.014)	(0.152)	(0.070)	(0.074)	(0.052)	(0.050)	(0.049)
pop	0.007	0.031***	0.034***	0.036***	−2.290***	−1.401***	−1.143***	−0.298	−0.318**	−0.370***
	(0.011)	(0.014)	(0.012)	(0.014)	(0.052)	(0.110)	(0.106)	(0.141)	(0.146)	(0.131)
sim	0.026***	0.011	0.018**	0.023***	−0.522***	−0.422***	−0.249***	−0.118***	−0.055*	0.000
	(0.006)	(0.008)	(0.009)	(0.008)	(0.053)	(0.055)	(0.059)	(0.032)	(0.033)	(0.032)
dis	−0.087***	−0.087***	−0.076***	−0.067***	−1.137***	−1.168***	−1.116***	−0.427	−2.810**	−6.039***
	(0.005)	(0.007)	(0.008)	(0.007)	(0.077)	(0.069)	(0.072)	(1.330)	(1.446)	(1.242)
bor	−0.008	0.073***	0.051***	0.071***	0.380*	0.293	0.108	−7.673	−1.549	2.170
	(0.014)	(0.017)	(0.022)	(0.019)	(0.237)	(0.211)	(0.197)	(4.326)	(5.368)	(4.056)
lan	0.087***	−0.022	−0.015	−0.010	−0.202	−0.316*	−0.093	−0.380	−0.443	0.511
	(0.017)	(0.015)	(0.019)	(0.019)	(0.205)	(0.189)	(0.190)	(4.109)	(4.759)	(3.810)
$euro$	0.032***	0.031***	0.019***	0.027***	0.191***	0.206***	0.210***	0.066***	0.037**	0.036**
	(0.007)	(0.006)	(0.007)	(0.007)	(0.028)	(0.026)	(0.025)	(0.014)	(0.016)	(0.014)
eec	0.055***	0.050***	0.040***	0.036***	0.311***	0.305***	0.205***	0.107***	0.083***	0.082***
	(0.005)	(0.005)	(0.006)	(0.008)	(0.017)	(0.018)	(0.018)	(0.009)	(0.009)	(0.009)

Note: The dataset consists of yearly bilateral trade data for $N = 380$ country pairs over 59 years. Mean stands for mean regression estimation results by using method in Moon and Weidner (2017). QR-PC is the method proposed in this paper, and QR-PC (static) is our proposed method but without dynamics. QR-CCE is the method developed in Harding et al. (2020). The value in parentheses is the standard error, and significance at the 1%, 5% and 10% levels are denoted by ***, **, *, respectively.

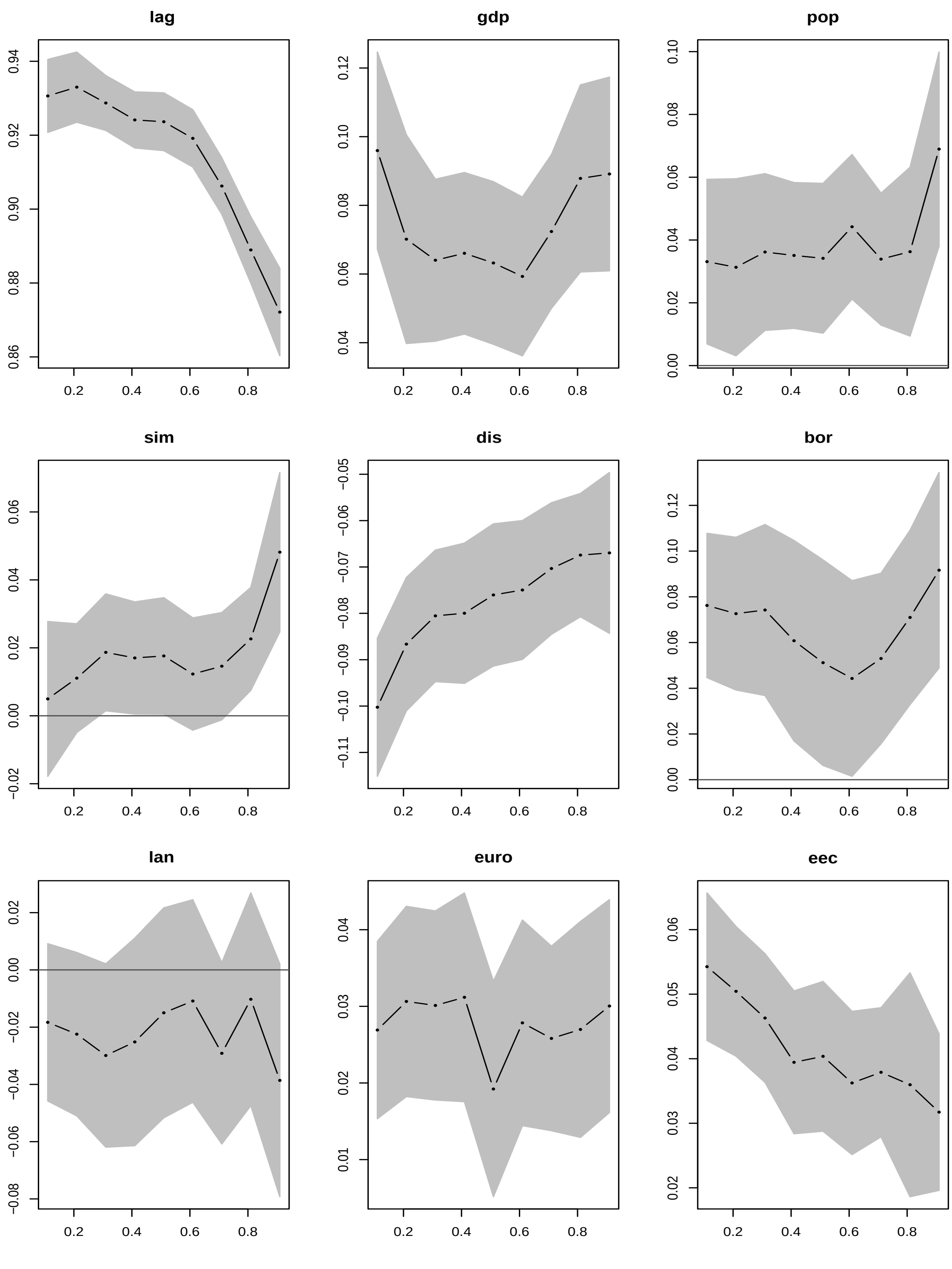

Figure 1: Estimated Coefficients across Quantiles for Bilateral Trade Data

Note: The dashed line represents the estimated coefficient, and the grey area represents a kernel density based 95% confidence band advanced by Powell (1991). The coefficients are plotted across quantiles, $\tau = 0.1, 0.2, \ldots, 0.9$.

6 Conclusion

In this paper, we develop a simple two-step procedure for estimating dynamic quantile panel data models with unobserved common factors, without imposing any restrictions on the data generating process for regressors. Specifically, we propose to use the iterative principal component (IPC) analysis in the first step to obtain consistent factor estimation, and then in the second step, run the augmented quantile regression by substituting these factors to obtain an estimator for the other parameters of interest.

Under regularity conditions, we show that the estimator is $\sqrt{NT}$ consistent. Same as other nonlinear panel data models with unobserved factors (see e.g. Chen et al. (2021c)), our estimator suffers from asymptotic bias arising from the estimation of both factors and factor loadings. Due to the existence of dynamic autocorrelation, our estimator also suffers from the Nickel bias, extending the findings in linear mean regression models (Moon and Weidner (2017)). To correct the asymptotic bias, we propose to use the split panel jackknife procedure, and show that the bias-corrected estimator follows a centered normal distribution asymptotically.

Monte Carlo simulations confirm that the finite sample performance of the proposed estimator is quite satisfactory, especially when compared with the performance of other existing estimators. We also demonstrate the usefulness of our approach with an application to a gravity model of bilateral trade flows for 380 pairs of 14 EU countries and 6 OECD countries over 1960-2019. We find that, on the one hand, the effect of being in the EU economic community could be larger/more significant in recession periods, and on the other hand, the corresponding long-run effect ranges from 27.40% at high quantiles to 118.72% at lower quantiles, covering most estimated effects in the literature.

We conclude by noting a few avenues for future research. A natural extension is to also allow the factors to be quantile dependent. Quantile-dependent factor models are becoming more and more popular and attracting attention from both theoretical and empirical researchers. There are already several influential papers in this area (see e.g. Chen et al. (2021b), Ando and Bai (2020)) and extending them to dynamic models can help improve our understanding of the dynamic nature of many important variables. Another interesting direction of research is to develop a formal procedure for testing whether the factors are quantile-varying, or more generally for cross-section dependence test in quantile regression. Such tests are well developed in mean regression (see e.g. Pesaran (2015), Bailey et al. (2016)), and recent developments in quantile factor models make such an extension both interesting and possible.

References

Acemoglu, D., Johnson, S., Robinson, J. A., and Yared, P. (2008). Income and democracy. *American Economic Review*, 98:808–42.

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

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

Anderson, J. E. and Yotov, Y. V. (2020). Short run gravity. *Journal of International Economics*, 126:103341.

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

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

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

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

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

Baier, S. L., Bergstrand, J. H., Egger, P., and McLaughlin, P. A. (2008). Do economic integration agreements actually work? Issues in understanding the causes and consequences of the growth of regionalism. *World Economy*, 31:461–497.

Bailey, N., Kapetanios, G., and Pesaran, M. H. (2016). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics*, 31:929–960.

Baldwin, R. and Taglioni, D. (2006). Gravity for dummies and dummies for gravity equations. *mimeo.,* National Bureau of Economic Research.

Baltagi, B. H. and Egger, P. (2016). Estimation of structural gravity quantile regression models. *Empirical Economics*, 50:5–15.

Bun, M. J. and Klaassen, F. J. (2007). The Euro effect on trade is not as large as commonly thought. *Oxford Bulletin of Economics and statistics*, 69:473–496.

Carrere, C. (2006). Revisiting the effects of regional trade agreements on trade flows with proper specification of the gravity model. *European Economic Review*, 50:223–247.

Chen, J., Shin, Y., and Zheng, C. (2021a). Estimation and inference in heterogeneous spatial panels with a multifactor error structure. *Journal of Econometrics,* Forthcoming.

Chen, L. (2021). Two-step estimation of quantile panel data models with interactive fixed effects. *mimeo.,* Peking University.

Chen, L., Dolado, J. J., and Gonzalo, J. (2021b). Quantile factor models. *Econometrica*, 89:875–910.

Chen, M., Fernández-Val, I., and Weidner, M. (2021c). Nonlinear factor models for network and panel data. *Journal of Econometrics*, 220:296–324.

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

Comunale, M., Dainauskas, J., and Lastauskas, P. (2021). What explains excess trade persistence? A theory of habits in the supply chains. *mimeo.,* Bank of Lithuania.

Cui, G., Norkute, M., Sarafidis, V., and Yamagata, T. (2021). Two-stage instrumental variable estimation of linear panel data models with interactive effects. *Econometrics Journal,* Forthcoming.

De Nardis, S. and Vicarelli, C. (2003). Currency unions and trade: The special case of emu. *Review of World Economics*, 139:625–649.

Dhaene, G. and Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies*, 82:991–1030.

Dong, C., Gao, J., and Peng, B. (2021). Varying-coefficient panel data models with nonstationarity and partially observed factor structure. *Journal of Business & Economic Statistics*, 39:700–711.

Ebell, M. (2016). Assessing the impact of trade agreements on trade. *National Institute Economic Review*, 238:R31–R42.

Eberhardt, M. and Presbitero, A. F. (2015). Public debt and growth: Heterogeneity and nonlinearity. *Journal of International Economics*, 97:45–58.

Egger, P. H. and Nigai, S. (2015). Structural gravity with dummies only: Constrained anova-type estimation of gravity models. *Journal of International Economics*, 97:86–99.

Eicher, T. S. and Henn, C. (2011). In search of WTO trade effects: Preferential trade agreements promote trade strongly, but unevenly. *Journal of International Economics*, 83:137–153.

Eicher, T. S., Henn, C., and Papageorgiou, C. (2012). Trade creation and diversion revisited: Accounting for model uncertainty and natural trading partner effects. *Journal of Applied Econometrics*, 27:296–321.

Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192:291–312.

Frankel, J. A., Stein, E., and Wei, S.-J. (1997). *Regional trading blocs in the world economic system*. Peterson Institute.

Galvao, A. F. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, 164:142–157.

Galvao, A. F. and Kato, K. (2016). Smoothed quantile regression for panel data. *Journal of econometrics*, 193:92–112.

Galvao, A. F. and Montes-Rojas, G. V. (2010). Penalized quantile regression for dynamic panel data. *Journal of Statistical Planning and Inference*, 140:3476–3497.

Glick, R. and Rose, A. K. (2016). Currency unions and trade: A post-emu reassessment. *European Economic Review*, 87:78–91.

Hahn, J. and Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, 27:1152–1191.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72:1295–1319.

Harding, M. and Lamarche, C. (2014). Estimating and testing a quantile regression model with interactive effects. *Journal of Econometrics*, 178:101–113.

Harding, M., Lamarche, C., and Pesaran, M. H. (2020). Common correlated effects estimation of heterogeneous: Dynamic panel quantile regression models. *Journal of Applied Econometrics*, 35:294–314.

Horowitz, J. L. (1998). Bootstrap methods for median regression models. *Econometrica*, pages 1327–1351.

Hufbauer, G. C. and Schott, J. J. (2009). Fitting Asia-Pacific agreements into the WTO system. In *Multilateralizing Regionalism: Challenges for the Global Trading System*, pages 554–635. Cambridge.

Kalirajan, K. (2007). Regional cooperation and bilateral trade flows: An empirical measurement of resistance. *International Trade Journal*, 21:85–107.

Kato, K., Galvao Jr, A. F., and Montes-Rojas, G. V. (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics*, 170:76–91.

Keller, W. (2004). International technology diffusion. *Journal of economic literature*, 42:752–782.

Koenker (2005). *Quantile Regression*. Cambridge University Press.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91:74–89.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica*, pages 33–50.

Koenker, R. and Xiao, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association*, 101:980–990.

Larch, M., Wanner, J., Yotov, Y. V., and Zylkin, T. (2019). Currency unions and trade: A PPML re-assessment with high-dimensional fixed effects. *Oxford Bulletin of Economics and Statistics*, 81:487–510.

Linder, S. B. (1961). *An essay on trade and transformation*. Almqvist & Wiksell Stockholm.

Martinez-Zarzoso, I., Felicitas, N.-L. D., and Horsewood, N. (2009). Are regional trading agreements beneficial? static and dynamic panel gravity models. *North American Journal of Economics and Finance*, 20:46–65.

Mastromarco, C., Serlenga, L., and Shin, Y. (2016). Modelling technical efficiency in cross sectionally dependent stochastic frontier panels. *Journal of Applied Econometrics*, 31:281–297.

Mayer, T., Vicard, V., and Zignago, S. (2019). The cost of non-Europe, revisited. *Economic Policy*, 34:145–199.

Meghir, C. and Pistaferri, L. (2004). Income variance dynamics and heterogeneity. *Econometrica*, 72:1–32.

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

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

Newey, K. and McFadden, D. (1994). Large sample estimation and hypothesis. *Handbook of Econometrics, IV, Edited by RF Engle and DL McFadden*, pages 2112–2245.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, pages 1417–1426.

Olivero, M. P. and Yotov, Y. V. (2012). Dynamic gravity: Endogenous country size and asset accumulation. *Canadian Journal of Economics/Revue canadienne d'économique*, 45:64–92.

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

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews*, 34:1089–1117.

Powell, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. In Barnett, W. A., Powell, J. L., and Tauchen, G. E., editors, *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press.

Sarafidis, V. and Wansbeek, T. (2021). Celebrating 40 years of panel data analysis: Past, present and future. *Journal of Econometrics*, 220:215–226.

Serlenga, L. and Shin, Y. (2007). Gravity models of intra-EU trade: Application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics*, 22:361–381.

Serlenga, L. and Shin, Y. (2013). The Euro effect on intra-EU trade: Evidence from the cross sectionally dependent panel gravity models. *mimeo.,* University of York.

Temple, J. and Van de Sijpe, N. (2017). Foreign aid and domestic absorption. *Journal of International Economics*, 108:431–443.

Thursby, J. G. and Thursby, M. C. (1987). Bilateral trade flows, the linder hypothesis, and exchange risk. *Review of Economics and Statistics*, 69:488–495.

van de Geer, S. (2000). *Empirical Processes in M-estimation*, volume 6. Cambridge university press.

Xu, X., Wang, W., Shin, Y., and Zheng, C. (2021). Dynamic network quantile regression model. *mimeo.,* University of York.

Yamarik, S. and Ghosh, S. (2005). A sensitivity analysis of the gravity model. *International Trade Journal*, 19:83–126.

Yoshihara, K.-I. (1978). Moment inequalities for mixing sequences. *Kodai Mathematical Journal*, 1:316–328.

Appendix A Proof of the Main Results

Before proving the main theorems, we introduce some definitions and notations.

For any fixed $\boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\gamma}_i \in \mathcal{A}$, where $\boldsymbol{\Theta}$ and $\boldsymbol{\gamma}_i \in \mathcal{A}$ are some compact sets, define:

$$\mathrm{S}_{NT}(\boldsymbol{\theta},\ \boldsymbol{\Gamma}, \boldsymbol{F}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}(\boldsymbol{\theta},\ \boldsymbol{\gamma}_i,\ \boldsymbol{f}_t),\ \mathrm{S}^*_{NT}(\boldsymbol{\theta},\ \boldsymbol{\Gamma}, \boldsymbol{F}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \rho_{it}(\boldsymbol{\theta},\ \boldsymbol{\gamma}_i,\ \boldsymbol{f}_t)$$

$$\mathrm{S}_{iT}(\boldsymbol{\theta}, \boldsymbol{\gamma}_i, \boldsymbol{F}) = \frac{1}{T}\sum_{t=1}^{T} l_{it}(\boldsymbol{\theta},\ \boldsymbol{\gamma}_i,\ \boldsymbol{f}_t),\ \mathrm{S}^*_{iT}(\boldsymbol{\theta}, \boldsymbol{\gamma}_i, \boldsymbol{F}) = \frac{1}{T}\sum_{t=1}^{T} \rho_{it}(\boldsymbol{\theta},\ \boldsymbol{\gamma}_i,\ \boldsymbol{f}_t)$$

where $\boldsymbol{\Gamma}' = [\boldsymbol{\gamma}_1,\ \ldots,\ \boldsymbol{\gamma}_N]$, $\boldsymbol{F}' = [\boldsymbol{f}_1, \ldots, \boldsymbol{f}_T]$, and

$$l_{it}(\boldsymbol{\theta},\ \boldsymbol{\gamma}_i,\ \boldsymbol{f}_t) = [\tau - K(\frac{y_{it} - \boldsymbol{\theta}'\boldsymbol{z}_{it} - \boldsymbol{\gamma}_i'\boldsymbol{f}_t}{h})](y_{it} - \boldsymbol{\theta}'\boldsymbol{z}_{it} - \boldsymbol{\gamma}_i'\boldsymbol{f}_t)\ ,$$
$$\rho_{it}(\boldsymbol{\theta},\ \boldsymbol{\gamma}_i,\ \boldsymbol{f}_t) = [\tau - \mathbb{1}(y_{it} \le \boldsymbol{\theta}'\boldsymbol{z}_{it} - \boldsymbol{\gamma}_i'\boldsymbol{f}_t)](y_{it} - \boldsymbol{\theta}'\boldsymbol{z}_{it} - \boldsymbol{\gamma}_i'\boldsymbol{f}_t)\ .$$

Further, for any random function $f(\cdot)$, define $\bar{f}(\cdot) = \mathbb{E}f(\cdot)$. $\tilde{f}(\cdot) = f(\cdot) - \bar{f}(\cdot)$. Moreover, we let $f^{(j)}(\cdot)$ denote the j-th order derivative of $f(\cdot)$.

A.1 Proof of Theorem 1

The proof of consistency follows standard technique for M-estimator, see e.g. van de Geer (2000), and we assume that Assumption 1 holds throughout this section.

Proof. First, define a ball $\boldsymbol{B}_{\delta,i} = \{\boldsymbol{\theta} \in \boldsymbol{\Theta},\ \boldsymbol{\gamma}_i \in \mathcal{A} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_1 + \|\boldsymbol{\gamma}_{i0} - \boldsymbol{\gamma}_{i0}\|_1 \le \delta\}$, where δ is a small positive number. Then under Assumption (1) (vi), we have:

$$\begin{aligned}
&\bar{\rho}_{it}(\breve{\boldsymbol{\theta}}, \breve{\boldsymbol{\gamma}}_i, \boldsymbol{f}_{t0}) - \bar{\rho}_{it}(\boldsymbol{\theta}_0,\ \boldsymbol{\gamma}_{i0},\ \boldsymbol{f}_{t0}) \\
&= [(\breve{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)',\ (\breve{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i0})] \cdot \mathbb{E}[\mathrm{g}_{it}(0|\boldsymbol{X}_{it})\boldsymbol{X}_{it}\boldsymbol{X}_{it}'] \cdot [(\breve{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)',\ (\breve{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i0})]' \ + o(\delta^2) \\
&\ge \varkappa \cdot [\|\breve{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 + \|\breve{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i0}\|_1]^2 + o(\delta^2) = (\varkappa + o(1)) \cdot \delta^2,
\end{aligned} \tag{A.1}$$

for any $(\breve{\boldsymbol{\theta}}', \breve{\boldsymbol{\gamma}}_i')'$ on the boundary of $\boldsymbol{B}_{\delta,i}$, and the first equality follows from that $\partial\bar{\rho}_{it}(\boldsymbol{\theta}_0,\ \boldsymbol{\gamma}_{i0},\ \boldsymbol{f}_{t0})/\partial\boldsymbol{\theta} = \boldsymbol{0}$ and $\partial\bar{\rho}_{it}(\boldsymbol{\theta}_0,\ \boldsymbol{\gamma}_{i0},\ \boldsymbol{f}_{t0})/\partial\boldsymbol{\gamma}_{i0} = \boldsymbol{0}$ in the first equality. For any point outside of $\boldsymbol{B}_{\delta,i}$, that is $(\boldsymbol{\theta}', \boldsymbol{\gamma}_i')' \in \boldsymbol{B}_{\delta,i}^C$, we still have $(\breve{\boldsymbol{\theta}}', \breve{\boldsymbol{\gamma}}_i')' = \delta/m(\boldsymbol{\theta}',\ \boldsymbol{\gamma}_i')' + (1 - \delta/m)(\boldsymbol{\theta}_0',\ \boldsymbol{\gamma}_{i0}')'$ on the boundary by defining $m = \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_1 + \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i0}\|_1 > \delta$, and therefore (A.1) holds.

Observe that ρ_{it} is convex in $(\boldsymbol{\theta}', \boldsymbol{\gamma}_i')'$ for any given $\boldsymbol{X}_{it} = (\boldsymbol{z}_{it}', \boldsymbol{f}_t')'$, and the fact the expectation operator preserves convexity, we have

$$\delta/m\bar{\rho}_{it}(\boldsymbol{\theta},\ \boldsymbol{\gamma}_i,\ \boldsymbol{f}_{t0}) + (1 - \delta/m)\bar{\rho}_{it}(\boldsymbol{\theta}_0,\ \boldsymbol{\gamma}_{i0},\ \boldsymbol{f}_{t0}) \ge \bar{\rho}_{it}(\breve{\boldsymbol{\theta}}, \breve{\boldsymbol{\gamma}}_i,\ \boldsymbol{f}_{t0}) \ \text{ for all } \ i,$$

which together with ($A.1$) further implies that

$$\bar{\rho}_{it}(\boldsymbol{\theta},\ \boldsymbol{\gamma}_i,\ \boldsymbol{f}_{t0}) - \bar{\rho}_{it}(\boldsymbol{\theta}_0,\ \boldsymbol{\gamma}_{i0},\ \boldsymbol{f}_{t0}) \geq m/\delta[\bar{\rho}_{it}(\breve{\boldsymbol{\theta}},\breve{\gamma}_i,\ \boldsymbol{f}_{t0}) - \bar{\rho}_{it}(\boldsymbol{\theta}_0,\ \boldsymbol{\gamma}_{i0},\ \boldsymbol{f}_{t0})] \geq (\varkappa + o(1)) \cdot \delta^2.$$

The fact $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 > \delta$ implies $(\hat{\boldsymbol{\theta}}', \hat{\gamma}_i')' \in \boldsymbol{B}_{\delta,i}^C$ for all i, then gives us that

$$\bar{\mathrm{S}}^*_{NT}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Gamma}}\boldsymbol{H}', \boldsymbol{F}_0) - \bar{\mathrm{S}}^*_{NT}(\boldsymbol{\theta}_0, \boldsymbol{\Gamma}_0, \boldsymbol{F}_0) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[\bar{\rho}_{it}(\hat{\boldsymbol{\theta}}, \boldsymbol{H}\hat{\gamma}_i, \boldsymbol{f}_{t0}) - \bar{\rho}_{it}(\boldsymbol{\theta}_0,\ \boldsymbol{\gamma}_{i0},\ \boldsymbol{f}_{t0})] \geq (\varkappa + o(1)) \cdot \delta^2. \tag{A.2}$$

Second, by the multiplicative form of factor and factor loadings and the basic inequality $\mathrm{S}_{NT}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{F}}) = \mathrm{S}_{NT}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Gamma}}\boldsymbol{H}', \hat{\boldsymbol{F}}\boldsymbol{H}^{-1'}) \leq \mathrm{S}_{NT}(\boldsymbol{\theta}_0,\ \boldsymbol{\Gamma}_0, \hat{\boldsymbol{F}}\boldsymbol{H}^{-1'}) = \mathrm{S}_{NT}(\boldsymbol{\theta}_0,\ \boldsymbol{\Gamma}_0\boldsymbol{H}^{-1'}, \hat{\boldsymbol{F}})$, we also have

$$\begin{aligned}
&\bar{\mathrm{S}}^*_{NT}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Gamma}}\boldsymbol{H}', \boldsymbol{F}_0) - \bar{\mathrm{S}}^*_{NT}(\boldsymbol{\theta}_0, \boldsymbol{\Gamma}_0,\ \boldsymbol{F}_0) \\
&\leq [\mathrm{S}_{NT}(\boldsymbol{\theta}_0,\ \boldsymbol{\Gamma}_0, \hat{\boldsymbol{F}}\boldsymbol{H}^{-1}) - \mathrm{S}^*_{NT}(\boldsymbol{\theta}_0,\ \boldsymbol{\Gamma}_0, \hat{\boldsymbol{F}}\boldsymbol{H}^{-1})] + [\mathrm{S}^*_{NT}(\boldsymbol{\theta}_0,\ \boldsymbol{\Gamma}_0, \hat{\boldsymbol{F}}\boldsymbol{H}^{-1}) - \mathrm{S}^*_{NT}(\boldsymbol{\theta}_0,\ \boldsymbol{\Gamma}_0, \boldsymbol{F}_0)] \\
&- [\mathrm{S}_{NT}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{F}}) - \mathrm{S}^*_{NT}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{F}})] - [\mathrm{S}^*_{NT}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Gamma}}\boldsymbol{H}', \hat{\boldsymbol{F}}\boldsymbol{H}^{-1}) - \mathrm{S}^*_{NT}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Gamma}}\boldsymbol{H}', \boldsymbol{F}_0)] \\
&- \tilde{\mathrm{S}}^*_{NT}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Gamma}}\boldsymbol{H}',\ \boldsymbol{F}_0) + \tilde{S}^*_{NT}(\boldsymbol{\theta}_0,\ \boldsymbol{\Gamma}_0,\ \boldsymbol{F}_0).
\end{aligned} \tag{A.3}$$

It then follows from ($A.2$) and ($A.3$) that for any arbitrarily small $\delta > 0$,

$$\begin{aligned}
P[\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 > \delta] \leq\ & P[\sup_{\boldsymbol{\theta},\boldsymbol{\Gamma}} |\mathrm{S}_{NT}(\boldsymbol{\theta},\ \boldsymbol{\Gamma},\ \hat{\boldsymbol{F}}) - \mathrm{S}^*_{NT}(\boldsymbol{\theta},\ \boldsymbol{\Gamma},\ \hat{\boldsymbol{F}})| > 1/6(\varkappa + o(1)) \cdot \delta^2] \\
&+ P[\sup_{\boldsymbol{\theta},\boldsymbol{\Gamma}} |\mathrm{S}^*_{NT}(\boldsymbol{\theta}, \boldsymbol{\Gamma}, \hat{\boldsymbol{F}}\boldsymbol{H}^{-1}) - \mathrm{S}^*_{NT}(\boldsymbol{\theta}, \boldsymbol{\Gamma}, \boldsymbol{F}_0)| > 1/6(\varkappa + o(1)) \cdot \delta^2] \\
&+ P[|\tilde{\mathrm{S}}^*_{NT}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Gamma}}\boldsymbol{H}',\ \boldsymbol{F}_0) - \tilde{S}^*_{NT}(\boldsymbol{\theta}_0,\ \boldsymbol{\Gamma}_0,\ \boldsymbol{F}_0)| > 1/3(\varkappa + o(1)) \cdot \delta^2].
\end{aligned} \tag{A.4}$$

The consistency then would follow if we could show that each of the three terms on the right hand side of ($A.4$) is $o(1)$.

According to Lemma 1 in Horowitz (1998), the first term (the smoothing error) on the right-hand side of ($A.4$) is $o(1)$ under Assumption 1 (vii) since

$$\sup_{\boldsymbol{\theta},\boldsymbol{\Gamma}} |\mathrm{S}_{NT}(\boldsymbol{\theta},\ \boldsymbol{\Gamma},\ \hat{\boldsymbol{F}}) - \mathrm{S}^*_{NT}(\boldsymbol{\theta},\ \boldsymbol{\Gamma},\ \hat{\boldsymbol{F}})| \lesssim h.$$

holds for any value of parameters and variables.

The second term on the right-hand side of (A.4) is also $o(1)$ since for some $M > 0$,

$$\begin{aligned}
&\sup_{\boldsymbol{\theta},\boldsymbol{\Gamma}} |\mathrm{S}^*_{NT}(\boldsymbol{\theta},\boldsymbol{\Gamma},\hat{\boldsymbol{F}}\boldsymbol{H}^{-1}) - \mathrm{S}^*_{NT}(\boldsymbol{\theta},\boldsymbol{\Gamma},\boldsymbol{F}_0)| \\
&= \sup_{\boldsymbol{\theta},\boldsymbol{\Gamma}} |\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \rho_{it}(\boldsymbol{\theta},\ \gamma_i, \boldsymbol{H}'^{-1}\hat{\boldsymbol{f}}_t) - \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \rho_{it}(\boldsymbol{\theta},\ \gamma_i, \boldsymbol{f}_{t0})| \\
&\le \max_i \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta},\gamma_i\in\mathcal{A}} |\frac{1}{T}\sum_{t=1}^T \rho_{it}(\boldsymbol{\theta},\ \gamma_i, \boldsymbol{H}'^{-1}\hat{\boldsymbol{f}}_t) - \frac{1}{T}\sum_{t=1}^T \rho_{it}(\boldsymbol{\theta},\ \gamma_i,\ \boldsymbol{f}_{t0})| \\
&\le \bar{\gamma}\cdot M\cdot \sqrt{\frac{1}{T}\sum_{t=1}^T \|\boldsymbol{H}'^{-1}\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t0}\|^2} = o_p(1)\ ,
\end{aligned}$$

where for the last inequality, we have used the fact that

$$\frac{1}{T}\sum_{t=1}^T \|\hat{\boldsymbol{f}}_t - \boldsymbol{H}'\boldsymbol{f}_{t0}\|^2 = O_p(\frac{1}{\sqrt{min\{N,T\}}}).$$

under Assumptions 1(ii)-(iv) as a result of Proposition A.1 in Bai (2009), and $\bar{\gamma} = \sup_{\boldsymbol{\gamma}_i\in\mathcal{A}}\|\boldsymbol{\gamma}_i\|$ is bounded under Assumption 1(v).

Finally, note that for any $\epsilon > 0$,

$$\begin{aligned}
&P[|\tilde{\mathrm{S}}^*_{NT}(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\Gamma}}\boldsymbol{H}',\ \boldsymbol{F}_0) - \tilde{S}^*_{NT}(\boldsymbol{\theta}_0,\ \boldsymbol{\Gamma}_0,\ \boldsymbol{F}_0)| > \epsilon] \\
&\le \max_i P[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta},\gamma_i\in\mathcal{A}} |\frac{1}{T}\sum_{t=1}^T \tilde{\rho}_{it}(\boldsymbol{\theta},\boldsymbol{\gamma}_i,\ \boldsymbol{f}_{t0}) - \frac{1}{T}\sum_{t=1}^T \tilde{\rho}_{it}(\boldsymbol{\theta}_0,\ \boldsymbol{\gamma}_{i0},\ \boldsymbol{f}_{t0})| > \epsilon] \\
&\le \sum_{i=1}^N P[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta},\gamma_i\in\mathcal{A}} |\frac{1}{T}\sum_{t=1}^T \tilde{\rho}_{it}(\boldsymbol{\theta},\boldsymbol{\gamma}_i,\ \boldsymbol{f}_{t0}) - \frac{1}{T}\sum_{t=1}^T \tilde{\rho}_{it}(\boldsymbol{\theta}_0,\ \boldsymbol{\gamma}_{i0},\ \boldsymbol{f}_{t0})| > \epsilon] \qquad \text{(A.5)}
\end{aligned}$$

To show that the third term on the right-hand side of (A.4) is $o(1)$, we therefore only need to show

$$P[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta},\gamma_i\in\mathcal{A}} |\frac{1}{T}\sum_{t=1}^T \tilde{\rho}_{it}(\boldsymbol{\theta},\boldsymbol{\gamma}_i,\ \boldsymbol{f}_{t0}) - \frac{1}{T}\sum_{t=1}^T \tilde{\rho}_{it}(\boldsymbol{\theta}_0,\ \boldsymbol{\gamma}_{i0},\ \boldsymbol{f}_{t0})| > \epsilon] = o(\frac{1}{N}) \qquad \text{(A.6)}$$

Write $\boldsymbol{\vartheta}_i = (\boldsymbol{\theta}', \boldsymbol{\gamma}_i')'$ and $\mathcal{B}_i = (\boldsymbol{\Theta}', \mathcal{A}')'$. Since $\mathcal{B}_i$ is compact, there exist a finite, say, R small open balls, denoted $\mathcal{B}^j_{\eta,i}, j = 1,\ldots,R$ with center $\boldsymbol{\vartheta}^j_i$ and $\|\boldsymbol{\vartheta}^j_i - \boldsymbol{\vartheta}^k_i\| \ge \eta$ for any $j \ne k$, such that $\mathcal{B}_i \in \bigcup_{j=1}^L \mathcal{B}^j_{\eta,i}$. Therefore, any $\boldsymbol{\vartheta}_i$ must belong to some $\mathcal{B}^j_{\eta,i}$ and we further denote its corresponding center as $\boldsymbol{\vartheta}_i^*$. Then,

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta},\boldsymbol{\gamma}_i\in\mathcal{A}}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{\rho}_{it}(\boldsymbol{\theta},\boldsymbol{\gamma}_i,\ \boldsymbol{f}_{t0})-\frac{1}{T}\sum_{t=1}^{T}\tilde{\rho}_{it}(\boldsymbol{\theta}_0,\ \boldsymbol{\gamma}_{i0},\ \boldsymbol{f}_{t0})\right| \\
&=\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta},\boldsymbol{\gamma}_i\in\mathcal{A}}\left|\frac{1}{T}\sum_{t=1}^{T}[D_{it}(\boldsymbol{\vartheta}_i^*)-\bar{D}_{it}(\boldsymbol{\vartheta}_i^*)]+\frac{1}{T}\sum_{t=1}^{T}[D_{it}(\boldsymbol{\vartheta}_i)-D_{it}(\boldsymbol{\vartheta}_i^*)-(\bar{D}_{it}(\boldsymbol{\vartheta}_i)-\bar{D}_{it}(\boldsymbol{\vartheta}_i^*))]\right|, \\
&\leq \max_{1\leq j\leq R}\frac{1}{T}\sum_{t=1}^{T}[D_{it}(\boldsymbol{\vartheta}_i^j)-\bar{D}_{it}(\boldsymbol{\vartheta}_i^j)]+\sup_{\|\boldsymbol{\vartheta}_i^a-\boldsymbol{\vartheta}_i^b\|\leq\eta}\frac{1}{T}\sum_{t=1}^{T}[D_{it}(\boldsymbol{\vartheta}_i^a)-D_{it}(\boldsymbol{\vartheta}_i^b)-(\bar{D}_{it}(\boldsymbol{\vartheta}_i^a)-\bar{D}_{it}(\boldsymbol{\vartheta}_i^b))],
\end{aligned}$$

where $D_{it}(\boldsymbol{\vartheta}_i)=\rho_{it}(\boldsymbol{\theta},\boldsymbol{\gamma}_i,\ \boldsymbol{f}_{t0})-\rho_{it}(\boldsymbol{\theta}_0,\ \boldsymbol{\gamma}_{i0},\ \boldsymbol{f}_{t0})$ and $\bar{D}_{it}(\boldsymbol{\vartheta}_i)=\mathbb{E}[\rho_{it}(\boldsymbol{\theta},\boldsymbol{\gamma}_i,\ \boldsymbol{f}_{t0})-\rho_{it}(\boldsymbol{\theta}_0,\ \boldsymbol{\gamma}_{i0},\ \boldsymbol{f}_{t0})]$.

Further, by the property of check function that for any two different $\boldsymbol{\vartheta}_i^a$ and $\boldsymbol{\vartheta}_i^b$,

$$|D_{it}(\boldsymbol{\vartheta}_i^a)-D_{it}(\boldsymbol{\vartheta}_i^b)|\leq M\|\boldsymbol{\vartheta}_i^a-\boldsymbol{\vartheta}_i^b\|\|\boldsymbol{X}_{it}\|,$$

and under Assumption 1 (ii), $\mathbb{E}\|\boldsymbol{X}_{it}\|\leq M$, we therefore have

$$\sup_{\|\boldsymbol{\vartheta}_i^a-\boldsymbol{\vartheta}_i^b\|\leq\eta}\frac{1}{T}\sum_{t=1}^{T}[D_{it}(\boldsymbol{\vartheta}_i^a)-D_{it}(\boldsymbol{\vartheta}_i^b)-(\bar{D}_{it}(\boldsymbol{\vartheta}_i^a)-\bar{D}_{it}(\boldsymbol{\vartheta}_i^b))]\leq M\eta\frac{1}{T}\sum_{t=1}^{T}(\|\boldsymbol{X}_{it}\|-\mathbb{E}\|\boldsymbol{X}_{it}\|)+2M\eta. \tag{A.7}$$

As a result,

$$\begin{aligned}
&P[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta},\boldsymbol{\gamma}_i\in\mathcal{A}}|\frac{1}{T}\sum_{t=1}^{T}\tilde{\rho}_{it}(\boldsymbol{\theta},\boldsymbol{\gamma}_i,\ \boldsymbol{f}_{t0})-\frac{1}{T}\sum_{t=1}^{T}\tilde{\rho}_{it}(\boldsymbol{\theta}_0,\ \boldsymbol{\gamma}_{i0},\ \boldsymbol{f}_{t0})|>\epsilon] \\
&\leq\sum_{j=1}^{R}P\left[\left|\frac{1}{T}\sum_{t=1}^{T}[D_{it}(\boldsymbol{\vartheta}_i^j)-\bar{D}_{it}(\boldsymbol{\vartheta}_i^j)]\right|\geq\epsilon/3\right]+P\left[M\eta\left|\frac{1}{T}\sum_{t=1}^{T}(\|\boldsymbol{X}_{it}\|-\mathbb{E}\|\|\boldsymbol{X}_{it}\|)\right|\geq\epsilon/3\right]+P\left[2M\eta\geq\epsilon/3\right]
\end{aligned} \tag{A.8}$$

It is easily seen that the last term in (A.8) would be zero if $\eta<\epsilon/6M$.

For the first term, by the property of the check function and Assumption 1 (ii), we have $\mathbb{E}|D_{it}(\boldsymbol{\vartheta}_i)|^{4+\delta}\leq M$ for any $\boldsymbol{\vartheta}_i\in\mathcal{B}_i$, then we have

$$\mathbb{E}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[D_{it}(\boldsymbol{\vartheta}_i^j)-\bar{D}_{it}(\boldsymbol{\vartheta}_i^j)]\right|^4\leq M, \tag{A.9}$$

under the α-mixing condition in Assumption 1(i) and Theorem 3 in Yoshihara (1978). Further, by the Markov's inequality,

$$P\left[\left|\frac{1}{T}\sum_{t=1}^{T}[D_{it}(\boldsymbol{\vartheta}_i^j)-\bar{D}_{it}(\boldsymbol{\vartheta}_i^j)]\right|\geq\epsilon/3\right]=O_p(\frac{1}{T^2}). \tag{A.10}$$

We can show same result for the second term on the right hand side of (A.8) in a similar way, and we thus have shown (A.6) according to Assumption 1(vii), which completes the proof.

■

A.2 Proof of Theorem 2

Proof. Collecting results (S.10), (S.13), (S.15) in Lemma 4, (S.44) in Lemma 7, (S.57) in Lemma 8, (S.68) in Lemma 9, (S.75) in Lemma 10 and (S.79)-(S.82) in Lemma 11, we have

$$\begin{aligned}
&\boldsymbol{\Delta}_{NT}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}^{(1)}\boldsymbol{w}_{it}-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\Lambda}_{tN}(\frac{\boldsymbol{\Gamma}_0^{m\prime}\boldsymbol{\Gamma}_0^{m}}{N})^{-1}\boldsymbol{\gamma}_{i0}^{m}\boldsymbol{z}_{it}^{\prime}(\boldsymbol{\theta}_0^{m}-\hat{\boldsymbol{\theta}}^{m})-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\Lambda}_{tN}(\frac{\boldsymbol{\Gamma}_0^{m\prime}\boldsymbol{\Gamma}_0^{m}}{N})^{-1}\boldsymbol{\gamma}_{i0}^{m}e_{it}\\
&+\frac{1}{T}(\boldsymbol{b}_{1,NT}+\boldsymbol{b}_{2,NT}+\boldsymbol{b}_{3,NT})+\frac{1}{N}(\boldsymbol{c}_{1,NT}+\boldsymbol{c}_{2,NT})+o_p(\frac{1}{T})+o_p(\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|)
\end{aligned}\tag{A.11}$$

where $\boldsymbol{\Delta}_{NT}$ is defined in (11) and

$$\begin{aligned}
\boldsymbol{b}_{1,NT}&=\frac{(0.5-\tau)}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\mathrm{g}_{it}(0|\boldsymbol{X}_{it})\boldsymbol{w}_{it}\boldsymbol{f}_{t0}^{\prime}\boldsymbol{\Pi}_{iT}^{-1}\boldsymbol{f}_{t0}],\\
\boldsymbol{b}_{2,NT}&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s<t}^{T}\mathbb{E}[\mathrm{g}_{it}(0|\boldsymbol{X}_{it})\boldsymbol{w}_{it}\boldsymbol{f}_{t0}^{\prime}\boldsymbol{\Pi}_{iT}^{-1}\boldsymbol{f}_{s0}(\tau-\mathbb{1}(\mu_{is}\le 0))],\\
\tilde{b}_{k,3,NT}&=\frac{\tau(1-\tau)}{2}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\boldsymbol{f}_{t0}^{\prime}\boldsymbol{\Pi}_{iT}^{-1}\boldsymbol{\Psi}_{ik}\boldsymbol{\Pi}_{iT}^{-1}\boldsymbol{f}_{t0}],\\
\boldsymbol{c}_{1,NT}&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\boldsymbol{\gamma}_{i0}^{\prime}\boldsymbol{\Sigma}_{\Gamma_m}^{-1}\boldsymbol{\gamma}_{i0}^{m}\mathbb{E}[\mathrm{g}_{it}(0|\boldsymbol{X}_{it})c_{it}\boldsymbol{z}_{it}]+\boldsymbol{\Phi}_{iT}\boldsymbol{\Sigma}_{\Gamma_m}^{-1}\boldsymbol{\gamma}_{i0}^{m}\mathbb{E}[\mu_{it}(\tau-\mathbb{1}(\mu_{it}\le 0))]\right\},\\
\tilde{c}_{k,2,NT}&=\frac{1}{2}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[e_{it}^{2}]\boldsymbol{\gamma}_{i0}^{m\prime}(\frac{\boldsymbol{\Gamma}_0^{m\prime}\boldsymbol{\Gamma}_0^{m}}{N})^{-1}(\boldsymbol{\Psi}_{tk}+2\boldsymbol{\Theta}_{tk})(\frac{\boldsymbol{\Gamma}_0^{m\prime}\boldsymbol{\Gamma}_0^{m}}{N})^{-1}\boldsymbol{\gamma}_{i0}^{m}.
\end{aligned}$$

Thus we can write:

$$\begin{aligned}
&\sqrt{NT}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\\
=&\boldsymbol{\Delta}_{NT}^{-1}\boldsymbol{\Phi}_{NT}\sqrt{NT}(\hat{\boldsymbol{\theta}}^{m}-\boldsymbol{\theta}_0^{m})+\boldsymbol{\Delta}_{NT}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(l_{it}^{(1)}\boldsymbol{w}_{it}-\boldsymbol{\Lambda}_{tN}(\frac{\boldsymbol{\Gamma}_0^{m\prime}\boldsymbol{\Gamma}_0^{m}}{N})^{-1}\boldsymbol{\gamma}_{i0}^{m}e_{it})\\
&+\sqrt{\frac{N}{T}}\boldsymbol{\Delta}_{NT}^{-1}\boldsymbol{b}_{NT}+\sqrt{\frac{T}{N}}\boldsymbol{\Delta}_{NT}^{-1}\boldsymbol{c}_{NT}+o_p(1),
\end{aligned}\tag{A.12}$$

where $\boldsymbol{\Phi}_{NT}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\Lambda}_{tN}(\frac{\boldsymbol{\Gamma}_0^{m\prime}\boldsymbol{\Gamma}_0^{m}}{N})^{-1}\boldsymbol{\gamma}_{i0}^{m}\boldsymbol{z}_{it}^{\prime}$, $\boldsymbol{b}_{NT}=\boldsymbol{b}_{1,NT}+\boldsymbol{b}_{2,NT}+\boldsymbol{b}_{3,NT}$, and $\boldsymbol{c}_{NT}=\boldsymbol{c}_{1,NT}+\boldsymbol{c}_{2,NT}$.

Further according to Theorem 4.3 of Moon and Weidner (2017),

$$\sqrt{NT}(\hat{\boldsymbol{\theta}}^m - \boldsymbol{\theta}_0^m) = \boldsymbol{D}_{NT}^{-1}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\boldsymbol{z}}_{it}'e_{it} + \sqrt{\frac{N}{T}}(\boldsymbol{a}_{1,NT} + \boldsymbol{a}_{2,NT}) + \sqrt{\frac{T}{N}}\boldsymbol{a}_{3,NT}\right] + o_p(1), \tag{A.13}$$

where $\boldsymbol{a}_{q,NT} = [a_{q,1}, \ldots, a_{q,p+1}]', q = 1, 2, 3$, and

$$a_{k,1,NT} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s<t}E[z_{it,k}\boldsymbol{f}_{t0}'(\frac{\boldsymbol{F}_0'\boldsymbol{F}_0}{T})^{-1}\boldsymbol{f}_{s0}e_{is})] \tag{A.14}$$

$$a_{k,2,NT} = \frac{1}{N}\sum_{i=1}^{N}\frac{(\boldsymbol{z}_{i,k} - \breve{\boldsymbol{z}}_{i,k})'\boldsymbol{F}_0}{T}(\frac{\boldsymbol{F}_0'\boldsymbol{F}_0}{T})^{-1} \times (\frac{\boldsymbol{\Gamma}_0^{m\prime}\boldsymbol{\Gamma}_0^m}{N})^{-1}\boldsymbol{\gamma}_{i0}^m(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(e_{it})^2) \tag{A.15}$$

$$a_{k,3,NT} = \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{z}_{i,k}'\boldsymbol{M}_{\boldsymbol{F}_0}\boldsymbol{\Omega}\boldsymbol{F}_0(\frac{\boldsymbol{F}_0'\boldsymbol{F}_0}{T})^{-1}(\frac{\boldsymbol{\Gamma}_0^{m\prime}\boldsymbol{\Gamma}_0^m}{N})^{-1}\boldsymbol{\gamma}_{i0}^m, \tag{A.16}$$

with $\breve{\boldsymbol{z}}_{i,k}$, $\boldsymbol{z}_{i,k}$ *and* $\boldsymbol{\Omega}$ defined in Theorem 2.[21]

Moreover, as shown in Lemma 3 (k) Horowitz (1998) and by Assumption 2 (iv),

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}l_{it}^{(1)}\boldsymbol{w}_{it} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tau - \mathbb{1}(\mu_{it} \leq 0))\boldsymbol{w}_{it} + o_p(1) \tag{A.17}$$

As a results, (A.12) could be further written as

$$\begin{aligned}
&\sqrt{NT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\Delta}_{NT}^{-1}(\boldsymbol{\Phi}_{NT}\boldsymbol{D}_{NT}^{-1}\tilde{\boldsymbol{z}}_{it}'e_{it} + (\tau - \mathbb{1}(\mu_{it} \leq 0))\boldsymbol{w}_{it} - \boldsymbol{\Lambda}_{tN}(\frac{\boldsymbol{\Gamma}_0^{m\prime}\boldsymbol{\Gamma}_0^m}{N})^{-1}\boldsymbol{\gamma}_{i0}^m e_{it})\\
&+ \boldsymbol{\Delta}_{NT}^{-1}\left\{\boldsymbol{\Phi}_{NT}\boldsymbol{D}_{NT}^{-1}\left[\sqrt{\frac{N}{T}}(\boldsymbol{a}_{1,NT} + \boldsymbol{a}_{2,NT}) + \sqrt{\frac{T}{N}}\boldsymbol{a}_{3,NT}\right] + \sqrt{\frac{N}{T}}\boldsymbol{b}_{NT} + \sqrt{\frac{T}{N}}\boldsymbol{c}_{NT}\right\} + o_p(1)
\end{aligned} \tag{A.18}$$

Similar to the proof of Theorem 4.1 in Fernández-Val and Weidner (2016), the desired results follows by invoking a central limit theorem for martingale difference sequences.

■

Proof of Theorem 3:

Proof. From the expansion (A.18), for any estimator $\hat{\boldsymbol{\theta}}_{(js)}$, $j, s = 0, 1, 2$ and at least one of j or s

[21] Therefore, if we assume time homogeneity, this term would disappear, see also Theorem 3 in Bai (2009).

must be 0, we have

$$\begin{aligned}
&\sqrt{NT}(\hat{\boldsymbol{\theta}}_{(js)} - \boldsymbol{\theta}_0) \\
&= \frac{2^{\mathbb{1}(j>0)}2^{\mathbb{1}(s>0)}}{\sqrt{NT}} \sum_{i\in\mathcal{N}_j}\sum_{s\in\mathcal{T}_j} \boldsymbol{\Delta}^{-1}_{\mathcal{N}_j\mathcal{T}_s}(\boldsymbol{\Phi}_{\mathcal{N}_j\mathcal{T}_s}\boldsymbol{D}^{-1}_{\mathcal{N}_j\mathcal{T}_s}\tilde{\boldsymbol{z}}'_{it}e_{it} + (\tau - \mathbb{1}(\mu_{it}\leq 0))\boldsymbol{w}_{it} - \boldsymbol{\Lambda}_{t\mathcal{N}_j}(\frac{\boldsymbol{\Gamma}_0^{m\prime}\boldsymbol{\Gamma}_0^m}{\mathcal{N}_j})^{-1}\boldsymbol{\gamma}^m_{i0}e_{it}) \\
&+ \boldsymbol{\Delta}^{-1}_{\mathcal{N}_j\mathcal{T}_s}\left\{\boldsymbol{\Phi}_{\mathcal{N}_j\mathcal{T}_s}\boldsymbol{D}^{-1}_{\mathcal{N}_j\mathcal{T}_s}\left[\sqrt{\frac{2^{\mathbb{1}(s>0)}N}{T}}(\boldsymbol{a}_{1,\mathcal{N}_j\mathcal{T}_s} + \boldsymbol{a}_{2,\mathcal{N}_j\mathcal{T}_s}) + \sqrt{\frac{2^{\mathbb{1}(j>0)}T}{N}}\boldsymbol{a}_{3,\mathcal{N}_j\mathcal{T}_s}\right] + \sqrt{\frac{2^{\mathbb{1}(s>0)}N}{T}}\boldsymbol{b}_{\mathcal{N}_j\mathcal{T}_s} + \sqrt{\frac{2^{\mathbb{1}(j>0)}T}{N}}\boldsymbol{c}_{\mathcal{N}_j\mathcal{T}_s}\right\} \\
&+ o_p(1),
\end{aligned}$$

As $\mathcal{N}_j, \mathcal{T}_s \to \infty$ and under Assumptions 1(iii), 2(v), 3 and Theorem 2, we have

$$\lim_{\mathcal{N}_j,\mathcal{T}_s\to\infty} \boldsymbol{\Delta}_{\mathcal{N}_j\mathcal{T}_s} = \boldsymbol{\Delta}, \lim_{\mathcal{N}_j,\mathcal{T}_s\to\infty} \boldsymbol{\Phi}_{\mathcal{N}_j\mathcal{T}_s} = \boldsymbol{\Phi}, \lim_{\mathcal{N}_j,\mathcal{T}_s\to\infty} \boldsymbol{D}_{\mathcal{N}_j\mathcal{T}_s} = \boldsymbol{D}, \lim_{\mathcal{N}_j,\mathcal{T}_s\to\infty} \boldsymbol{\Lambda}_{t\mathcal{N}_j} = \boldsymbol{\Lambda}_t,$$
$$\lim_{\mathcal{N}_j\to\infty} \frac{\boldsymbol{\Gamma}_0^{m\prime}\boldsymbol{\Gamma}_0^m}{\mathcal{N}_j} = \boldsymbol{\Sigma}_{\boldsymbol{\Gamma_m}}, \lim_{\mathcal{N}_j,\mathcal{T}_s\to\infty} \boldsymbol{a}_{q,\mathcal{N}_j\mathcal{T}_s}, q = 1,2,3, \lim_{\mathcal{N}_j,\mathcal{T}_s\to\infty} \boldsymbol{b}_{\mathcal{N}_j\mathcal{T}_s} = \boldsymbol{b}, \lim_{\mathcal{N}_j,\mathcal{T}_s\to\infty} \boldsymbol{c}_{\mathcal{N}_j\mathcal{T}_s} = \boldsymbol{c},$$

and same as Fernández-Val and Weidner (2016),

$$\hat{\boldsymbol{\Theta}} = \begin{bmatrix} \hat{\boldsymbol{\theta}}_{(00)} \\ \hat{\boldsymbol{\theta}}_{(01)} \\ \hat{\boldsymbol{\theta}}_{(02)} \\ \hat{\boldsymbol{\theta}}_{(10)} \\ \hat{\boldsymbol{\theta}}_{(20)} \end{bmatrix} \xrightarrow{d} MN\left(\begin{bmatrix} \boldsymbol{\Delta}^{-1}\left\{\boldsymbol{\Phi}\boldsymbol{D}^{-1}\left[\pi(\boldsymbol{a}_1+\boldsymbol{a}_2) + \pi^{-1}\boldsymbol{a}_3\right] + \pi\boldsymbol{b} + \pi^{-1}\boldsymbol{c}\right\} \\ \boldsymbol{\Delta}^{-1}\left\{\boldsymbol{\Phi}\boldsymbol{D}^{-1}\left[2\pi(\boldsymbol{a}_1+\boldsymbol{a}_2) + \pi^{-1}\boldsymbol{a}_3\right] + 2\pi\boldsymbol{b} + \pi^{-1}\boldsymbol{c}\right\} \\ \boldsymbol{\Delta}^{-1}\left\{\boldsymbol{\Phi}\boldsymbol{D}^{-1}\left[2\pi(\boldsymbol{a}_1+\boldsymbol{a}_2) + \pi^{-1}\boldsymbol{a}_3\right] + 2\pi\boldsymbol{b} + \pi^{-1}\boldsymbol{c}\right\} \\ \boldsymbol{\Delta}^{-1}\left\{\boldsymbol{\Phi}\boldsymbol{D}^{-1}\left[\pi(\boldsymbol{a}_1+\boldsymbol{a}_2) + 2\pi^{-1}\boldsymbol{a}_3\right] + \pi\boldsymbol{b} + 2\pi^{-1}\boldsymbol{c}\right\} \\ \boldsymbol{\Delta}^{-1}\left\{\boldsymbol{\Phi}\boldsymbol{D}^{-1}\left[\pi(\boldsymbol{a}_1+\boldsymbol{a}_2) + 2\pi^{-1}\boldsymbol{a}_3\right] + \pi\boldsymbol{b} + 2\pi^{-1}\boldsymbol{c}\right\} \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 & 1 \\ 1 & 0 & 2 & 1 & 1 \\ 1 & 1 & 1 & 2 & 0 \\ 1 & 1 & 1 & 0 & 2 \end{bmatrix} \otimes (\boldsymbol{\Delta}^{-1}\boldsymbol{V}\boldsymbol{\Delta}^{-1})\right).$$

The results in (17) then follows by noticing that $\hat{\boldsymbol{\theta}}_{bc} - \boldsymbol{\theta}_0 = [3, -0.5, -0.5, -0.5, -0.5]\hat{\boldsymbol{\Theta}}$ and using the property of multivariate normal distribution. ■