Within-group Estimators for Fixed Effects Quantile Models with Large N and Large T*

Heng Chen†
Bank of Canada

Feb.12th 2015
Very preliminary, Not for citation

Abstract

This paper provides within-group estimators for fixed effects quantile models. We apply the quantile coupling transformation to well approximate fixed effects quantile models by fixed effects Gaussian models. Our estimators enjoy three advantages: (1) fixed effects are eliminated by within-group estimators, so that ours is computationally fast without involving the increasing Hessian matrix with respect to the dimension of fixed effects (Koenker, 2004); (2) the rates of convergence and asymptotic normality are derived, where the limit theory allows for both sequential and joint limits; (3) estimates under the potential mis-specification can be interpreted as solutions of minimizing least squared mis-specification errors.

Keywords: Quantile coupling; Fixed effects quantile models; Mis-specification; Within-group estimator.

JEL codes: C13; C23

*I owe my gratitude to my advisor, Yanqin Fan, for her guidance and encouragement. I would also like to thank Ben Fung, David Chavez, Kim P. Huyhn, Sokbae Lee, Jack Porter, John Rust and Gregor Smith for their helpful suggestions. Maren Hansen provided excellent writing assistance.

†Email: chhe@bankofcanada.ca. Comments are welcome.

1 Introduction

> Quantiles of convolutions of random variables are rather intractable objects, and preliminary differencing strategies familiar from Gaussian models have sometimes unanticipated effects. — Koenker and Hallock (2001).

Despite the widespread use of both panel-data methodology and quantile-regression methodology, there has been little work at the intersection of the two. The most likely explanation is the difficulty in extending differencing methods to quantiles. Many usually think there is no general transformation that can suitably eliminate the individual effects, compared to mean regressions. Intuitively the first differencing (FD) and within-group (WG) estimators are impossible under the original fixed effect quantile panel model, due to the fact that the difference of the quantile is not the quantile of the difference.

To show impossibility of either FD or WG under the original data, let us look at the fixed effects $\tau-$th quantile panel model for $1 \leq i \leq N$ (e.g. individuals) and $1 \leq t \leq T$ (e.g. time periods)

$$Y_{it} = X_{it}^{\top}\beta(\tau) + \eta_i(\tau) + \epsilon_{it}$$

where $\eta_i(\tau)$ is the unobserved fixed effect and $F_{\epsilon_{it}}^{-1}(\tau|X_{it}, \eta_i) = 0$. Denote $Q_{\epsilon_{it}}(X_{it}, \eta_i, \tau)$ as the $\tau-$th conditional quantile of ϵ_{it} given $X_{it} \in \mathbb{R}^P$ and $\eta_i \in \mathbb{R}$ (including the intercept term), then

$$Q_{\epsilon_{it}-\epsilon_{is}}\left(X_{it} - X_{is}, \eta_i, \tau\right) \neq 0$$

where the $\tau-$th conditional quantile of difference between ϵ_{it} and ϵ_{is} is generally not zero.

However in this paper we propose to approximate the above fixed effects quantile models (FE-QM) with fixed effects Gaussian models (FE-GM), so that we can apply the traditional within-group estimators for FE-GM in order to eliminate fixed effects. We show that such within-group estimators are based on the transformed data from the quantile coupling transformation, instead of the original data. For a given i, the quantile coupling transformation groups the t dimension data into J bins/cells with equal observations $m = T/J$ in each bin, and then compute the local $\lceil \tau m \rceil -$th order statistics of Y_{it} in each bin/cell. As shown below, such local order statistics can be well approximated by a Gaussian random variable with some negligible approximation errors. In addition, this paper proposes the backfitting algorithm to overcome the curse of dimensionality of X, that is, there is no need to construct the bins/cells for the multivariate case. To the best of our knowledge, this is the first paper that performs this within-group estimation for the FE-QM.

Quantile coupling (QC) is used to approximate a general random variable by a Gaussian random variable on the same measurable space. Quantile coupling has usually been seen as strongly approximating the sample average by a Gaussian random variable; however Zhou (2006) uses this variable to establish the strong approximation of the sample median. Thereafter Brown, Cai and Zhou (2008) use the QC based on the sample median to construct a nonparametric median estimator under an equispaced univariate fixed design with homogeneous errors. Further Chen (2014) generalizes quantile coupling results to the arbitrary $\lceil \tau m \rceil -$th order statistics with non-identically distributed errors under the multivariate random design.

Our within-group estimator enjoys three advantages: (1) since fixed effects are eliminated by WG under the approximated FE-GM, our within-group estimators do not need to involve the increasing Hessian matrix with respect to the dimension of fixed effects under FE-QM. Hence our estimator is computationally fast, compared to the approaches in Koenker (2012) and Kato, Motnes-Rojas and Galvao (2012) where one is required to deal directly with a large number of fixed effects parameters[1]. (2) Under N and T going to infinity, the rates of convergence and asymptotic normality are derived, where the limit theory allows for both a sequential limit, $T \to \infty$ followed by $N \to \infty$, and joint limit, $T, N \to \infty$ simultaneously. (3) Estimates under the potential misspecification can be interpreted as solutions of minimizing least squared mis-specification errors.

Unlike within-group estimators eliminating the fixed effects of FE-GM under the fixed T, the consistency and normality of our within-group estimators require $T \to \infty$ for the quantile coupling transformation being applied. This is because we need $m \to \infty$ in order to well approximate an order statistics with a Gaussian random variable. Hence in this paper we will use big N and big T to handle the incidental parameter problem (see Neyman and Scott (1948) and Lancaster (2000) for a review) under FE-QM. The goal of this paper is to formally establish sufficient conditions for consistency and asymptotic normality of the estimator. As well from a technical point of view, the application of the quantile coupling transformation for FE-QM is of independent interest, which results in a different estimation strategy existing for the FE-QM (e.g. Koenker (2004) and Kato, et al. (2012)) and simplifies the proof without invoking the empirical process technique. Indeed, the core of our proof is deriving stochastic bounds of approximation errors between FE-QM and FE-GM, and then choosing an appropriate bin size m to "kill" their effects on the limit distributions.

We now review the literature related to this paper. Maximum likelihood estimators for a general nonlinear panel data (Hahn and Newey (2000)) lead to inconsistency in the presence of fixed effects, when T is fixed. However Graham, Hahn and Powell (2009) show that when $T = 2$ and the common parameter does not depend on τ ($\beta(\tau) = \beta \ \forall \tau$) for the FE-QM, we can estimate β by running the least absolute deviation of $Y_{i2} - Y_{i1}$ on $X_{i2} - X_{i1}$. However their argument does not extend to the general case. Koenker (2004)[2] and Kato, Motnes-Rojas and Galvao (2012) analyze the FE-QM by treating fixed effects as parameters; in particular, Kato et al. (2012) offer a precise condition for the joint asymptotic (when n and T jointly go to infinity) properties by studying the rate of the reminder term in the Badahur representation of their estimator. Canay (2011) proposes a two-step estimator of the common parameters for FE-QM with the restriction that each individual effect is not allowed to change across quantiles ($\eta_i(\tau) = \eta_i \ \forall \tau$): his first step estimation is able to use the within-group estimator from the original data to estimate fixed effects η_i (being identical under both conditional mean and quantile models), and then run the quantile estimation based on $Y_{it} - \widehat{\eta}_i$ and X_{it}. In addition Galvao, Lamarche and Lima (2013) investigate the censored FE-QM, while Galvao and Wang (2015) propose efficient minimum distance estimators of FE-QM. For other recent

[1] The computation of the variance-covariance matrix for inferences under Koenker (2012) and Kato, et al. (2012) is also difficult to implement due to a large number of fixed effects.

[2] Koenker (2004) treats fixed effects as the same across quantiles, but when estimating a single fixed quantile this restriction does not have any effect on the model and estimator.

developments on FE-QM and the associated estimation approach, readers can refer to Abreyea and Dahl (2008), Galvao and Kato (2010), and Rosen (2012).

The paper is organized as follows. Section 2 introduces fixed effects quantile models under both reduced and structural forms. Section 3 explains how to apply the quantile coupling transformation in order to approximate FE-QM by FE-GM. The WG estimator based on FE-GM is analyzed in Section 4. Section 5 discusses the Monte Carlo simulation results, and Section 6 associates the explanation of minimizing least squared mis-specification errors to our WG estimators. All proofs are included in the appendix.

2 Fixed Effects Quantile Models

The appealing features of fixed effects quantile models (FE-QM) are that they can control for individual heterogeneity via fixed effects, while exploring effects of heterogeneous covariates within the quantile framework.

Our conditional fixed effect quantile model has the expression:

$$Q_Y(X_{it}, \eta_i, \tau) = X_{it}^\top \beta(\tau) + \eta_i(\tau)$$

which allows for the $\beta(\cdot)$ and $\eta_i(\cdot)$ depend to on τ. The model is semiparametric in the sense that the functional form of the conditional distribution of Y_{it} given X_{it} and $\eta_i(\cdot)$ is left unspecified, and no parametric assumption is made on the relation between $\eta_i(\cdot)$ and X_{it}.

This conditional FE-QM can be written as

$$Y_{it} = X_{it}^\top \beta(\tau) + \eta_i(\tau) + \epsilon_{it} \tag{1}$$

where $Q_\epsilon(X_i, \eta_i(\tau), \tau) = 0$ and ϵ_{it} is independent conditional on X_i and η_i. The class of FE-QM we are considering in this paper are required to satisfy Assumption R, below.

Assumption R (reduced form):

(i) $Q_\epsilon(X_{it}, \eta_i(\tau), \tau) = 0$;

(ii) $\epsilon_{it} = \epsilon(X_{it}, \eta_i, U_{it})$, and U_{it} is i.i.d. conditional on X_{it} and η_i for a given i;

(iii) for $\forall i$ and $U \in (0,1)$, $|\epsilon(X_{it}, \eta_i, U) - \epsilon(X_{is}, \eta_i, U)| \leq C\,|X_{it} - X_{is}|^{d_\sigma}$ where $d_\sigma \geq 1$.

Assumption R(i) is constructed from the conditional FE-QM, while Assumption R(ii) allows for a rich structure for ϵ_{it} to be heterogeneous and dependent on X_i and η_i. While Galvao, et al. (2012), Fernandez-val (2005) and Hahn and Newey (2004) assume unconditional temporal and cross sectional independence, Assumption R(ii) only assumes conditional independence. We exclude temporal dependence, and it is hoped that our innovative treatment for FE-QM will serve to illustrate the utility of quantile coupling ideas and to simulate interests in the use of this method. Assumption R(iii) imposes smoothness conditions on ϵ_{it}, required to apply quantile coupling results (see Appendix C).

The above Eq.(1) under Assumption R consists of several structural models assumed to satisfy Assumption S.

Assumption S (structural model):

(i) $Y_{it} = Q_Y(X_{it}, \eta_i, U_{it})$;

(ii) U_{it} follows a standard uniform distribution conditional on X_i and η_i[3];

(iii) $\tau \longmapsto Q(x, \eta, \tau)$ strictly increases on $(0, 1)$, almost surely in x and η[4];

(iv) U_{it} is independent of U_{is} for each $t \neq s$ conditional on X_i and η_i.

Now we provide three examples to illustrate the generality of our FE-QM models.

Example 1:

$$\begin{aligned} Y_{it} &= X_{it}^{\top}\beta(U_{it}) + \eta_i(U_{it}) \\ &= X_{it}^{\top}\beta(\tau) + \eta_i(\tau) + [X_{it}^{\top}\beta(U_{it}) + \eta_i(U_{it}) - X_{it}^{\top}\beta(\tau) - \eta_i(\tau)] \end{aligned}$$

In this nonadditive errors model, $\eta_i(\tau)$ introduces an element of nonparametric functional heterogeneity in the conditional distribution of Y_{it}. Note we make no parametric assumption on the relationship between X_{it} and $\eta_i(\tau)$. In this example,$\epsilon(X_{it}, \eta_i, U_{it}) = [X_{it}^{\top}\beta(U_{it}) + \eta_i(U_{it}) - X_{it}^{\top}\beta(\tau) - \eta_i(\tau)]$ and $Q_\epsilon(X_i, \eta_i, \tau) = 0$ by Assumption S(ii) and (iii). Assumption R (ii) is implied by Assumption S(iv)[5]. In the end $|\epsilon(X_{it}, \eta, U) - \epsilon(X_{is}, \eta, U)| = \left|(X_{it} - X_{is})^{\top} (\beta(U) - \beta(\tau))\right| \leq C \; |X_{it} - X_{is}|^{\top}$ where $d_\sigma = 1$, which is Assumption R(iii).

Example 2:

$$\begin{aligned} Y_{it} &= X_{it}^{\top}\alpha + \eta_i + (X_{it}^{\top}\gamma + \mu\eta_i)U_{it} \\ &= X_{it}^{\top}\left[\alpha + \gamma F_U^{-1}(\tau)\right] + \eta_i\left[1 + \mu F_U^{-1}(\tau)\right] \\ &\quad + \left[(X_{it}^{\top}\gamma + \mu\eta_i)U_{it} - X_{it}^{\top}\gamma F_U^{-1}(\tau) - \mu\eta_i F_U^{-1}(\tau)\right] \end{aligned}$$

where $\beta(\tau) = \alpha + \gamma F_U^{-1}(\tau)$ and $\eta_i(\tau) = \eta_i\left[1 + \mu F_U^{-1}(\tau)\right]$. This model is the panel generalization of the location-scale model of He (1997). In this example, we can see that $\epsilon(X_{it}, \eta_i, U_{it}) = (X_{it}^{\top}\gamma + \mu\eta_i)U_{it} - X_{it}^{\top}\gamma F_U^{-1}(\tau) - \mu\eta_i F_U^{-1}(\tau)$ and $Q_\epsilon(X_i, \eta_i, \tau) = 0$ by Assumption S(ii) and (iii). Assumption R (ii) is implied by Assumption S(iv). In the end $|\epsilon(X_{it}, \eta, U) - \epsilon(X_{is}, \eta, U)| = \left|(X_{it} - X_{is})^{\top} (\tau U - \gamma F_U^{-1}(\tau))\right| \leq C$ $|X_{it} - X_{is}|^{\top}$ where $d_\sigma = 1$, which is Assumption R(iii).

Example 3:

$$\begin{aligned} Y_{it} &= X_{it}^{\top}\beta(U_{it}) + \eta_i \cdot \gamma(U_{it}) \\ &= X_{it}^{\top}\beta(\tau) + \eta_i \cdot \gamma(\tau) + \left[X_{it}^{\top}\beta(U_{it}) + \eta_i \cdot \gamma(U_{it}) - X_{it}^{\top}\beta(\tau) - \eta_i \cdot \gamma(\tau)\right] \end{aligned}$$

[3] This representation incorporates traditional additive error models $Y_{it} = Q_Y(X_{it}) + \eta_i + F^{-1}(U_{it})$. Here $Q_Y(X_{it})$ is not $Q_Y(X_{it}, U_{it})$ and η_i is not $\eta_i(U_{it})$. Hence $Y_{it} = Q_Y(X_{it}, \eta_i, U_{it})$ can affect the entire distribution of the demand curve, while in traditional modesl it only affects the location of the distribution of the stochastic demand curve.

[4] Hence under Assumption S(ii) and (iii), $Q_Y(X_{it}, \eta_i, \tau)$ is the $\tau-$th conditional quantile of Y_{it} given X_i and η_i. To see this

$$\begin{aligned} \Pr(Y_{it} &\leq Q_Y(X_{it}, \eta_i, \tau)|X_i, \eta_i) = \Pr(Q_Y(X_{it}, \eta_i, U_{it}) \leq Q_Y(X_{it}, \eta_i, \tau)|X_i, \eta_i) \\ &= \Pr(U_{it} \leq \tau|X_i, \eta_i) = \tau. \end{aligned}$$

[5] This is because U_{it} is independent of U_{is} for each $t \neq s$ conditional on X_i and η_i.

In this linear quantile model, $\eta_i \cdot \gamma(U_{it})$ introduces an element of parametric functional heterogeneity in the conditional distribution of Y_{it}. Here $\epsilon(X_{it}, \eta_i, U_{it}) = X_{it}^\top \beta(U_{it}) + \eta_i \cdot \gamma(U_{it}) - X_{it}^\top \beta(\tau) - \eta_i \cdot \gamma(\tau)$ and $Q_\epsilon(X_i, \eta_i, \tau) = 0$ by Assumption S(ii) and (iii). Assumption R (ii) is implied by Assumption S(iv). Further $|\epsilon(X_{it}, \eta, U) - \epsilon(X_{is}, \eta, U)| = \left|(X_{it} - X_{is})^\top (\beta(U) - \beta(\tau))\right| \leq C \; |X_{it} - X_{is}|^\top$ where $d_\sigma = 1$.

3 Quantile Coupling of FE-QM

In this section we apply the quantile coupling to FE-QM, so that the within-group estimators based on the transformed data are feasible. In this respect, our approach is in contrast to Koenker (2004) and Kato, et al. (2012) who concentrate out the individual fixed effects parameters in the first step by considering varying time periods for a given i, then apply the whole varying time periods and varying individuals due to the impossibility of within-group methods under the original data. For example, Koenker (2004) considers the individual dummy variables estimator, which is a natural analog of the dummy variables estimator for the standard fixed effects Gaussian regression models. Recall

$$Y_{it} = X_{it}^\top \beta(\tau) + \eta_i(\tau) + \epsilon_{it};$$

then the Koenker (2004) estimator is defined as follows

$$\left(\widehat{\beta}(\tau), \widehat{\eta}(\tau)\right) \equiv \arg \min_{(\beta(\tau), \eta(\tau)) \in R \times R^N} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau \left(Y_{it} - X_{it}^\top \beta(\tau) - \eta_i(\tau)\right),$$

where $\eta(\tau) \equiv (\eta_1(\tau), \cdots, \eta_N(\tau))^\top$, $\rho_\tau(u) \equiv \{\tau - 1\,(u \leq 0)\}\, u$ is the check function as in Koenker and Bassett (1978).

Ours is to apply the quantile coupling transformation in the first step so the fixed effects can be eliminated by the within-group estimation from the transformed data.

We apply the quantile coupling techniques for computing local $\lceil \tau m \rceil$ −th order statistics of each bin/cell for a given i. In order to compute local order statistics, we have to divide T observations into J bins/cells with equal $m = T/J$ observations in each bin/cell, and then choose the order statistics value of the dependent variable Y in each bin/cell. We denote $\lceil \cdot \rceil$ to be the rounded-up integer. With these J bins for a given i, the transformed data set is generated as

$$\mathcal{Y}_{ij:\tau} \equiv \text{the } \lceil \tau m \rceil \text{-th smallest value of } Y \text{ in the } j\text{th bin}, \qquad 1 \leq j \leq J.$$

We set

$$\mathcal{E}_{ij:\tau} \equiv \text{the } \lceil \tau m \rceil \text{-th smallest value of } \epsilon \text{ in the } j\text{th bin}, \qquad 1 \leq j \leq J,$$

$$\theta_{ij} \left[\beta(\tau), \eta_i(\tau)\right] \equiv \mathcal{Y}_{ij:\tau} - \mathcal{E}_{ij:\tau},$$

then $\mathcal{Y}_{ij:\tau}$ can be written as

$$\mathcal{Y}_{ij:\tau} = \theta_{ij} \left[\beta(\tau), \eta_i(\tau)\right] + \mathcal{E}_{ij:\tau}.$$

By working on the local order statistics, the error ϵ_{it} is transformed to $\mathcal{E}_{ij:\tau}$. Lemma 3.1 below shows that $\mathcal{E}_{ij:\tau}$ can be closely approximated using Gaussian random variables. Then, according to Lemma 3.2, $\theta_{ij}\left[\beta(\tau),\eta_i(\tau)\right]$ is approximated by $\mathcal{X}_{i[j]}^{\top}\beta(\tau)+\eta_i(\tau)$, in which the vector $\mathcal{X}_{i[j]}^{\top}$ is the coordinates of endpoints in the jth bin/cell (which are precisely defined later).

Assumption QC (quantile coupling)

(i) ϵ_{it} has an unknown density function f_ϵ, such that $\int_{-\infty}^{0} f_\epsilon(u)du = \tau$ and $|f_\epsilon(u) - f_\epsilon(0)| \leq Cu^2$ in an open neighborhood of 0 for some constant C.

(ii) $\int |u|^\kappa f_\epsilon(u)du < \infty$ for some $\kappa > 0$.

(iii) There is a constant M such that $\max_{1\leq i\leq N, 1\leq t\leq T} \|X_{it}\| < M$.

(iv) $\max_{1\leq j\leq J} \left\|\mathcal{X}_{i[j]} - \mathcal{X}_{i[j]}\right\|_1 = O_p(J^{-1})$, where $\mathcal{X}_{i[j]}$ is the coordinates of endpoints in the jth bin.

The smoothness condition $|f_\epsilon(u) - f_\epsilon(0)| \leq Cu^2$ in Assumption QC (i) is satisfied, for example, by the Cauchy distribution, the Laplace distribution and the t distribution. Assumption QC(ii) guarantees the existence of moments of the order statistics, as in Cramer et al. (2002). Assumption QC(iii) assumes that the covariates are uniformly bounded. This is common to the quantile literature, and is imposed in A3 of Koenker (2004). Assumption QC(iv) imposes the conditions on the spacings of $\mathcal{X}_{i[j]}$, which can be substituted by low-level conditions on the density function of X, as in Gasser and Müller (1979). As the dimension of X increases, such spacing conditions for the adjacent cells/bins will still hold using the backfitting algorithm described in Section 4.1.

Lemma 3.1 *(a) When Assumptions R and QC hold, we have*

$$\mathcal{E}_{ij:\tau} = \frac{1}{\sqrt{m}}\left[\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)} Z_{ij} + \zeta_{ij}\right]$$

where $\{Z_{ij}\}_{j=1}^{J}$ *is i.i.d. standard Gaussian errors and* $f_{\epsilon,ij}(0)$ *is the density of* ϵ *evaluated as* $\mathcal{X}_{i[j]}$ *and* η_i*;*

(b) For all $l > 0$*,*

$$E\left|\zeta_{ij}\right|^l = O\left[\left(\frac{\sqrt{m}}{J}\right)^l + \left(\frac{1}{m}\right)^l\right].$$

Lemma 3.2 *When Assumption QC(iii) holds, we have*

$$\theta_{ij}\left[\beta(\tau),\eta_i(\tau)\right] - \mathcal{X}_{i[j]}^{\top}\beta(\tau) - \eta_i(\tau) = O_p(\frac{1}{J}).$$

Remark: Although bins with an unequal number of observations are allowed (for example, the data are binned with equal lengths in each dimension of X), the current bins with equal numbers of observations provide a homoscedastic Gaussian approximation for $f_{\epsilon,ij}(0)\cdot\mathcal{E}_{j,\lambda}$, where the variance of $f_{\epsilon,ij}(0)\cdot\mathcal{E}_{j,\lambda}$ depends on m, the number of observations in the bin[6]. However binning the data into equal lengths will result in heteroscedastic Gaussian errors where variance will depend on how the X values are distributed within the bin. Given that the density of the design points X is usually unknown, it is not clear how results could be extended to that context.

[6]Smoothing quantile criterion functions is essentially constructing overlapping bins by adjusting the bandwidth, as in Galvao and Kato (2010), but this requires extension of the quantile coupling results to the non-independent dataset. We leave this for future research.

4 Asymptotic of Within-group Estimators

Based on the quantile coupling results, we have the following approximated fixed effects Gaussian model:

Proposition 4.1 *When Assumptions R and QC hold, then for a given* i*,*

$$\sqrt{m}\mathcal{Y}_{ij:\tau} = \sqrt{m}\left[\mathcal{X}_{i[j]}^{\top}\beta(\tau) + \eta_i(\tau)\right] + \frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{ij} + \xi_{ij}$$

where $\{Z_{ij}\}_{j=1}^{J}$ *is i.i.d. standard Gaussian errors, and for all* $l > 0$

$$E\left|\xi_{ij}\right|^{l} = O\left[\left(\frac{\sqrt{m}}{J}\right)^{l} + \left(\frac{1}{m}\right)^{l}\right].$$

In this paper, we discuss both sequential and joint limits (Phillips and Moon, 1999, 2000) to study within-group estimators for FE-QM, since FE-QM has a double-indexed process related to N and T. For the sequential limits asymptotics, we let first T, then N, tend to infinity. Denote as $(T,N)_{\text{seq}} \to \infty$. For the joint asymptotics where T and N tend to infinity at the same time without specifying the exact relationship between N and T, we denote $(T,N) \to \infty$. Although sequential limits could give deceptive asymptotic results , its asymptotics simplify the proofs substantially, and provide valuable insights into the results. In contrast the view of joint asymptotics is more general, although it is significantly more difficult to obtain, even with more stringent assumptions. In our case, the cost is related to the requirements on the diverging rate of T and N. We expect that the scope of using sequential asymptotics is large for FE-QM; for instance, censored, duration, and survival fixed effects quantile models are examples that remain to be formally developed.

Based on Proposition 4.1, we end up with the approximated fixed effects Gaussian model (FE-GM)

$$\sqrt{m}\mathcal{Y}_{ij:\tau} \approx \sqrt{m}\left[\mathcal{X}_{i[j]}^{\top}\beta(\tau) + \eta_i(\tau)\right] + \frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{ij}$$

where we can apply within-group estimators to eliminate the fixed effects $\eta_i(\tau)$, allowing estimation of $\beta(\tau)$.

In summary there are two steps in our approach:

Step 1: for a given i, bin the data across t and compute the $\lceil \tau m \rceil$-th smallest value of Y in the jth bin where $1 \le j \le J$;

Step 2: use within-group estimators to the quantile coupling transformed data $\left\{\mathcal{Y}_{ij:\tau}, \mathcal{X}_{i[j]}\right\}_{1\le i\le N, 1\le j\le J}$.

Our within-group estimator can be expressed as

$$\begin{aligned}
\widehat{\beta}(\tau) \quad &= \quad \beta(\tau) + \\
&\frac{1}{\sqrt{m}}\left\{\sum_{i=1}^{N}\sum_{j=1}^{J}(\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top}\right\}^{-1} \cdot \\
&\sum_{i=1}^{N}\sum_{j=1}^{J}\left\{\left[\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{ij} + \xi_{ij} - \frac{1}{J}\sum_{s=1}^{J}\left(\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{is} + \xi_{is}\right)\right](\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top}\right\}.
\end{aligned}$$

Notice that our proposed estimator could also be thought of as estimating $Q_Y(X_{it}, \eta_i, \tau)$ in a nonparametric first step (in the small bin/cell), and then minimizing the estimated $Q_Y(\mathcal{X}_{i[j]}, \eta_i, \tau)$ and $\mathcal{X}_{i[j]}^{\top}\beta(\tau) + \eta_i(\tau)$ across for all $1 \le i \le N$ and $1 \le j \le J$.

Denote

$$
\begin{aligned}
V_i &\equiv \lim_{J\to\infty}\frac{1}{J}\sum_{j=1}^{J}\frac{\tau(1-\tau)}{[f_{\epsilon,ij}(0)]^2}(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^\top,\\
V &\equiv \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}V_i,\\
H_i &\equiv \lim_{J\to\infty}\frac{\sum_{j=1}^{J}(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^\top}{J},\\
H &\equiv \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}H_i.
\end{aligned}
$$

THEOREM 4.2 ***(sequential asymptotics)*** *Under Assumptions R and QC, and $J = T^a$ where $a \in \left(\frac{1}{2}, \frac{2}{3}\right)$, then when $(T,N)_{\text{seq}} \to \infty$ and we condition on the design $\{X_{ij}\}_{i=1,t=1}^{N,T}$,*

$$
\begin{aligned}
&\sqrt{NT}\left[\widehat{\beta}(\tau)-\beta(\tau)\right] \xrightarrow{d}\\
&\xrightarrow{d} N\left[0, H^{-1}VH^{-1}\right].
\end{aligned}
$$

Denote

$$
\begin{aligned}
V^* &\equiv \lim_{N\to\infty}\frac{1}{N}\left\{\sum_{i=1}^{N}\frac{1}{J}\sum_{j=1}^{J}\frac{\tau(1-\tau)}{[f_{\epsilon,ij}(0)]^2}(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^\top\right\},\\
H^* &\equiv \lim_{N\to\infty}\frac{1}{N}\left\{\sum_{i=1}^{N}\frac{1}{J}\sum_{j=1}^{J}(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^\top\right\},\\
\Theta_1 &\equiv \left\{N,T,J : \frac{\frac{\sqrt{J}}{\sqrt{N}}+\sqrt{\frac{T}{J}}+\frac{J^2}{T}}{\frac{T}{J^2}+\frac{J^3}{T^2}} \to 0, \frac{\left(\frac{T}{J^2}+\frac{J^3}{T^2}\right)\log N}{\left(\frac{\sqrt{J}}{\sqrt{N}}+\sqrt{\frac{T}{J}}+\frac{J^2}{T}\right)^2} \to 0, J = T^a \text{ with } a\in(1/3,1)\right\},\\
\Theta_2 &\equiv \left\{N,T,J : \frac{\frac{T}{J^2}+\frac{J^3}{T^2}}{\frac{\sqrt{J}}{\sqrt{N}}+\sqrt{\frac{T}{J}}+\frac{J^2}{T}} \to 0, \frac{\log N}{\frac{\sqrt{J}}{\sqrt{N}}+\sqrt{\frac{T}{J}}+\frac{J^2}{T}} \to 0, J = T^a \text{ with } a\in(1/3,1)\right\}.
\end{aligned}
$$

THEOREM 4.3 ***(joint asymptotics)*** *Under Assumptions R and QC, either set of Θ_1 or Θ_2 is not empty when $(T,N) \to \infty$, and we condition on the design $\{X_{ij}\}_{i=1,t=1}^{N,T}$,*

$$
\begin{aligned}
&\sqrt{NT}\left[\widehat{\beta}(\tau)-\beta(\tau)\right] \xrightarrow{d}\\
&\xrightarrow{d} N\left[0, (H^*)^{-1}V^*(H^*)^{-1}\right].
\end{aligned}
$$

There are several important remarks to be observed from these results. First, from Theorems 4.2 and 4.3, one can see that $\widehat{\beta}(\tau)$ are $\sqrt{NT}$ consistent. Second, notice that the asymptotics of within-group estimators have the same limiting distributions regardless of whether they are sequential or joint. In other words, under the standard assumption available in the literature on the sample rate for joint limits, the limiting distribution of our WG estimator under sequential limits is to equal to that under joint limits. This find is similar to Galvao and Wang (2014)[7]. With respect to the joint asymptotics argument, our within-group

[7] Galvao and Wang (2014) address the efficiency under the class of minimum distance estimators. Notice that the asymptotic covariance matrix between ours and Galvao and Wang (2014) is generally different, unless the errors are i.i.d. (Knight, 2001). We leave the efficiency issue for the future research.

estimators have the normal limiting distribution under the sophisticated conditions of N, T and J[8]. These conditions are used only to "kill" the approximation term ξ_{ij} in the derivation of the asymptotic results. It serves as a device on the type of situations where the asymptotics are likely to provide a good approximation for $\sqrt{NT}\left[\widehat{\beta}(\tau)-\beta(\tau)\right]$. Kato, et al. (2012) show that their fixed effect quantile regression estimator is asymptotically (mean-zero) normal if the following condition holds: $N^2/T \to 0$ as $N \to \infty$. Further, Kato and Galvao (2011) suggest that a fixed effects smoothed quantile regression estimator is also asymptotically (mean-zero) normal under the condition if $N/T \to 0$ as $N \to \infty$.

Remark 1: Our within-group estimator does not provide the analytical bias correction for the FE-QM; this possible correction is left for future research. Similarly, Galvao, et al. (2012) also discuss the impossibility of using analytical bias correction for the large panel data. And the biggest difficulty they face is how to establish the exact probability limit of the reminder term for the Bahadur representation, and they only manage to obtain moment inequalities for the reminder term. In our approach since we can only derive the quantile coupling inequality for the approximation error between the local order statistics and Gaussian random variable, we can not derive the exact probability limit for these "approximation error" terms either; that is, we can not provide the analytical bias reduction.

Remark 2 To apply the above procedure in practice one needs to choose the number of observations per bin, m. This parameter can be chosen data dependently by extending the median cross-validation criterion proposed by Zheng and Yang (1998).

Remark 3: Since our within-group estimators do not need to jointly estimate the fixed effects as in Koenker (2004)[9], Lamarche (2010), and Galvao, et al. (2012), ours is computationally fast. Although our method is less efficient because of binning[10], it is computationally less demanding since only few parameters are estimated and there is no need to handle the increasing Hessian matrix with respect to the increasing dimension of fixed effects under FE-QM.

Remark 4: Our proof for the limiting distribution does not work on the diverging number of fixed effects $\eta_i(\tau)$ where $1 \leq i \leq N$, since these fixed effects are cancelled out by WG estimation under the FE-GM. In addition, the objective function of WG estimation is smooth. Thus our proof is relatively straightforward, and does not use empirical process techniques such as Talagrand inequality, as in Kato, et al. (2012).

[8] Notice that the range of m is different from Cai and Zhou (2009), who try to derive the consistency and optimal rate of the convergence; here we are proving the asymptotical normality such that the binning has no impact on the asymptotic limiting distribution.

[9] Koenker (2004) explains how sparse matrix approaches can be used to aid computation in some cases.

[10] More efficient within-group estimator can be applied by weighting $\left\{\mathcal{Y}_{ij:\tau}, \mathcal{X}_{i[j]}\right\}_{1\leq i\leq N, 1\leq j\leq J}$ by the density function $f_{\epsilon,ij}(0)$. However we do not pursue the efficiency issue in this paper.

Next we turn to estimating the asymptotic covariance matrix. This matrix is not directly estimated by their sample analogues because the density $f_{\epsilon,ij}(0)$ is unknown. Let

$$\begin{aligned}\widehat{\eta}_i(\tau) &\equiv \frac{1}{J}\sum_{j=1}^{T}\mathcal{Y}_{ij:\tau} - \frac{1}{J}\sum_{j=1}^{T}\mathcal{X}_{i[j]}^{\top}\widehat{\beta}(\tau),\\ \widehat{\Xi}_{ij} &\equiv \mathcal{Y}_{ij:\tau} - \mathcal{X}_{i[j]}^{\top}\widehat{\beta}(\tau) - \widehat{\eta}_i(\tau),\\ \widehat{\epsilon}_{it} &\equiv Y_{it} - X_{it}^{\top}\widehat{\beta}(\tau) - \widehat{\eta}_i(\tau)\end{aligned}$$

where $\widehat{\eta}_i(\tau)$ can be viewed as an estimator for the fixed effect $\eta_i(\tau)$, $\widehat{\Xi}_{ij}$ is an estimator for $\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)\sqrt{m}}Z_{ij}$ and $\widehat{\epsilon}_{it}$ is an estimator for ϵ_{it}. Here we offer two alternatives to estimate $f_{\epsilon,ij}(0)$. One is like Kato, et al. (2012) using the kernel method for $f_{\epsilon,ij}(0)$ to estimate $\widehat{\epsilon}_{it}$ from the original FE-QM. However their method becomes impractical when a large number of fixed effects are present.

The second alternative is to estimate the sample variance of $\widehat{\Xi}_{ij}$ in a small neighborhood of $\mathcal{X}_{i[j]}^{\top}$, so that $1/\left[\widehat{f}_{\epsilon,ij}(0)\right]^2$ is computed as this variance is divided by $\tau(1-\tau)/m$. This approach uses the fact that under the approximated FE-GM, $\widehat{\Xi}_{ij}$ is an estimate of the Gaussian error and its variance is $\tau(1-\tau)/\left[f_{\epsilon,ij}(0)\sqrt{m}\right]^2$. To formalize this approach, let $\{h_n\}$ denote a sequence of positive numbers (bandwidths) such that $h_N \to 0$ as $N \to \infty$. We use the notation $K_{h_N}(u) \equiv h_N^{-1}K\left(u/h_N\right)$. Then

$$1/\left[\widehat{f}_{\epsilon,ij}(0)\right]^2 = \frac{m}{\tau(1-\tau)}\frac{\sum_{s=1}^{J}\left(\widehat{\Xi}_{is}\right)^2 K_{h_N}(\mathcal{X}_{i[j]}^{\top} - \mathcal{X}_{i[s]}^{\top})}{\sum_{s=1}^{J}K_{h_N}(\mathcal{X}_{i[j]}^{\top} - \mathcal{X}_{i[s]}^{\top})}.$$

To guarantee the consistency of $1/\left[\widehat{f}_{\epsilon,ij}(0)\right]^2$, we assume the following:

Assumption K (kernel function):

(i)The kernel K is continuous, bounded and of bounded variation on $\mathbb{R}$.

(ii) $h_N \to 0$ and $Jh_N \to \infty$ when either $(T,N)_{\text{seq}} \to \infty$ or $(T,N) \to \infty$.

Most standard kernels such as Gaussian and Epanechnikov kernels satisfy Assumption K(i). Assumption K(ii) is a restriction on the bandwidth h_n, which needs to be slower than J^{-1}.

Proposition 4.4 *When Assumptions R, QC and K hold, and we condition on the design* $\{X_{ij}\}_{i=1,t=1}^{N,T}$,

(a) ***sequential asymptotics:*** *if*$J = T^a$ *where* $a \in \left(\frac{1}{2}, \frac{2}{3}\right)$ *when* $(T,N)_{\text{seq}} \to \infty$*:*

$$\frac{1}{\left[\widehat{f}_{\epsilon,ij}(0)\right]^2} \xrightarrow{p} \frac{1}{\left[f_{\epsilon,ij}(0)\right]^2}.$$

(b) ***joint asymptotics:*** *if either set of* Θ_1 *or* Θ_2 *is not empty, when* $(T,N) \to \infty$*:*

$$\frac{1}{\left[\widehat{f}_{\epsilon,ij}(0)\right]^2} \xrightarrow{p} \frac{1}{\left[f_{\epsilon,ij}(0)\right]^2}.$$

Remark: When $\epsilon_{it} = \epsilon(\eta_i, U_{it})$ where ϵ_{it} does not depend on the observed covariates X_{it}, then $f_{\epsilon,ij}(0)$ is simply $f_{\epsilon,i}(0)$. We can then estimate the entity $1/\left[f_{\epsilon,ij}(0)\right]^2$ by $m\sum_{s=1}^{J}\left(\widehat{\Xi}_{is}\right)^2/\tau(1-\tau)J$. In the end, if

$f_{\epsilon,i}(0) = f_\epsilon(0)$ for all $1 \leq i \leq N$, then the asymptotic covariance matrix is

$$\frac{\tau(1-\tau)}{[f_\epsilon(0)]^2} \lim_{N\to\infty} \left\{ \frac{1}{NJ} \sum_{i=1}^{N} \sum_{j=1}^{J} (\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J} \mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J} \mathcal{X}_{i[s]})^\top \right\}^{-1}.$$

4.1 Numerical algorithm

In order to implement the above WG estimator, it is essential to have J bins. For the univariate fixed-design case, Brown et al. (2008) simply divide the data along the real line at equal interval lengths. However, when X is a multivariate random design, this is difficult. In particular, it is more challenging to bin the data in order to have an equal number of observations in each bin. Chen (2014) proposes using conditional (sequential) ordering for constructing bins/cells for the multivariate case. His approach essentially constructs the quantile function for the multivariate case (Sen and Chaudhuri, 2011), so that both multivariate fractile and geometric quantile mappings can be applied. However as the dimension of X increases, these mappings become more difficult and the computational complexity increases at an exponential rate. Hence this paper suggests using the *backfitting algorithm* (Breiman and Friedman, 1985) to compute $\beta(\tau)$ where in each iteration we only need to construct bins for the univariate case.

The backfitting algorithm for our WG estimator is implemented as follows:

Step 1: Initialize the $\widehat{\beta}_{-q}(\tau)$, $1 \leq q \leq p$ which is the $(p-1)\times 1$ estimated coefficients, except for $\beta_q(\tau)$.

Step 2: For each $1 \leq i \leq N$, compute $Y_{it} - (X_{it})_{-q}^\top \widehat{\beta}_{-q}(\tau)$ where $(X_{it})_{-q}$ denotes the $(p-1)\times 1$ covariates of X_{it}, except for X_{qit}.

Step 3: Divide the X_{qit} along the real line into J intervals of having equal observations m in each interval. Denote $\mathcal{X}_{qi[j]}$ as the coordinate of the $J-$th interval's endpoint. And let $\mathcal{Y}_{qij:\tau}$ be the $\lceil \tau m \rceil$-th smallest value of $\left[Y_{it} - (X_{it})_{-q}^\top \widehat{\beta}_{-q}(\tau)\right]$ in the jth bin.

Step 4: Run the WG estimator based on $\left\{\mathcal{Y}_{qij:\tau}, \mathcal{X}_{qi[j]}\right\}_{1\leq i\leq N, 1\leq j\leq J}$ to have the estimated $\widehat{\beta}_q(\tau)$.

Step 5: Repeat the above Steps 1, 2, 3 and 4 for $\widehat{\beta}_k(\tau)$, $k \neq q$ with the updated $\widehat{\beta}_q(\tau)$ included in $\widehat{\beta}_{-k}(\tau)$ for Step 1, and $Y_{it} - (X_{it})_{-k}^\top \widehat{\beta}_{-k}(\tau)$ and $\mathcal{X}_{ki[j]}$ for Steps 2, 3 and 4.

Step 6: Stop the iteration utill $\widehat{\beta}_k(\tau)$ converges.

The backfitting algorithm helps avoid the need for binning the multivariate design, so that binning along the real line is straightforward. Since the backfitting algorithm is a contraction mapping under the general mean regression, this guarantees the convergence of our WG estimation. However the final estimates may depend on initial values of $\widehat{\beta}_{-q}(\tau)$ and orders of covariates X in Step 1, and the convergence criterion in Step 2. Moreover, reducing non-uniqueness of the solutions (Hastie and Tibshirani, 1990) and developing asymptotic theory (Mammen, Linton and Nielsen, 1999) are left for future investigation.

5 Asymptotics under Mis-specification

Of course, one can estimate the conditional FE-QM without assuming the correct specification using various non-parametric methods such as kernel estimation. However, our result provide a convenient parametric alternative to nonparametric methods when researchers are not sure about a correct specification or when they want to keep a parametric model for reasons of parsimony or interpretability, even though it may not pass a specification test such as the nonparametric kernel based on test proposed by Zheng (1998).

Our within-group estimator has an appealing interpretation, minimizing the squared mis-specification error. For the existing cross-sectional quantile model Angrist, Chernozhukov and Fernandez-Val (2006) offer the conditional quantile regression estimator as an explanation of minimizing the weighted squared mis-specification errors, while Chamerlain (1994) provides the minimum distance quantile estimator for the histogram weighted squared mis-specification errors. However Chamerlain (1994) is only built upon the discrete X and the exact nature of the linear approximation has remained elusive. In this paper we use the close connection between the FE-QM and FE-GM from the quantile coupling transformation, and generalize Chamerlain (1994)'s ideas to the continuous random variables X. In the end, our result focuses on the panel framework instead of the classical cross-sectional one.

Notice we first generalize the estimator in Kato et al. (2012) as the (infeasible) criterion of minimizing the weighted squared mis-specification errors, following Angrist et al. (2006) to the fixed effect quantile models as

$$\widehat{\beta}_{CAF}(\tau) \approx \arg \min_{\beta, \{\eta_i\}_{i=1}^N} \sum_{i=1}^N \sum_{t=1}^T \overline{w}_\tau(X_{it}) \cdot \left[Q_{Y_{it}}(X_i, \eta_i, \tau) - X_{it}^\top \beta(\tau) - \eta_i(\tau)\right]^2$$

where $\overline{w}_\tau(X_{it}, \beta(\tau), \eta_i)$ has the similar representation form as in Angrist et al. (2006)[11] and depends on $\beta(\tau)$.

On the other hand, our within-group estimator for the FE-QM is

$$\widehat{\beta}(\tau) \approx \arg\min_{\beta} \sum_{i=1}^N \sum_{j=1}^J \left[Q_{Y_{it}}(\mathcal{X}_{i[j]}, \eta_i, \tau) - \mathcal{X}_{i[j]}^\top \beta(\tau) - \eta_i(\tau)\right]^2.$$

To see this, from

$$\sqrt{m}\mathcal{Y}_{ij:\tau} \approx \sqrt{m}\left[\mathcal{X}_{i[j]}^\top \beta(\tau) + \eta_i(\tau)\right] + \frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)} Z_{ij}$$

we have as similar in Galvao and Kato (2014)

$$\widehat{\beta}(\tau) \xrightarrow{P} E\left[\widetilde{\mathcal{X}}_{i[j]}\widetilde{\mathcal{X}}_{i[j]}^\top\right]^{-1} E\left[\widetilde{\mathcal{X}}_{i[j]}\widetilde{\mathcal{Y}}_{ij:\tau}\right] \equiv \beta_0(\tau)$$

where

$$\begin{aligned}
\widetilde{\mathcal{X}}_{i[j]} &\equiv \mathcal{X}_{i[j]} - E\left[\mathcal{X}_{i[j]}|\eta_i(\tau)\right] \\
\widetilde{\mathcal{Y}}_{ij:\tau} &\equiv \mathcal{Y}_{ij:\tau} - E\left[\mathcal{Y}_{ij:\tau}|\eta_i(\tau)\right] \\
\beta_0(\tau) &\equiv \arg \min_{\beta, \eta_i(\tau) \in L_2} E\left[\mathcal{Y}_{ij:\tau} - \mathcal{X}_{i[j]}^\top \beta(\tau) - \eta_i(\tau)\right]^2 \\
&= \arg \min_{\beta, \eta_i(\tau) \in L_2} E\left[Q_{Y_{it}}(\mathcal{X}_{i[j]}, \eta_i, \tau) - \mathcal{X}_{i[j]}^\top \beta(\tau) - \eta_i(\tau)\right]^2 \text{ since } \mathcal{Y}_{ij:\tau} = Q_{Y_{it}}(\mathcal{X}_{i[j]}, \eta_i, \tau)
\end{aligned}$$

[11] See Appendix B in Belloni, Chernozhukov and Fernandez-Val (2011) for bounds of approximated conditional linear quantile models.

where $\mathcal{X}_{i[j]}^{\top}\beta_0(\tau)+\eta_{0i}(\tau)$ is the best linear approximation to $Q_{Y_{it}}(\mathcal{X}_{i[j]},\eta_i,\tau)$. Just as the coefficient vector of the best linear approximation to the conditional mean is a parameter of interest in the cross section case, $\beta_0(\tau)$ can be regarded as a plausible of interest of the panel data.

The difference between $\widehat{\beta}_{CAF}(\tau)$ and $\widehat{\beta}(\tau)$ is mainly due to their weight functions of mis-specification errors, where $\widehat{\beta}_{CAF}(\tau)$ is the importance weight while $\widehat{\beta}(\tau)$ is the histogram weight. Notice that Chamerlain (1994)'s result is a special case only weighted by the probability mass function covariates, which is

$$\arg\min_{\beta}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{S}\left[Q_{Y_{it}}(X_{its},\eta_i,\tau)-X_{its}\beta(\tau)-\eta_i(\tau)\right]^2$$

where X_{its} denotes the discrete mass point X evaluated at the $s-$th grid for a given i and t.

6 Monte Carlo Simulations

This section conducts simulations to investigate the finite sample performance of the within-group estimator for FE-QM. We consider the bivariate X_{it} example and apply the backfitting algorithm described in Section 4.1:

$$Y_{it}=\eta_i+X_{it}+(1+0.1X_{it})\epsilon_{it}$$

where $X_{kit}\sim U[-2,2]+2\cdot Z_i$ for $k=1,2$, $\eta_i=10\cdot Z_i$, $Z_i\sim 10\cdot N(0,1)$ and $\epsilon_{it}\sim 0.1\cdot$Cauchy. The correlation between η_i and X_{it} makes the random effects estimators inconsistent. The Cauchy distribution of the ϵ_{it} makes the WG for the FE-GM[12] inconsistent as well. We consider cases where $n\in\{10,100\}, T\in\{25,50\}$ and $J\in\{5,10\}$.

[12] This FE-GM is without the quantile coupling transformation, that is, m is equal to 1.

Table 1: Monte Carlo Simulation Results

		Coefficient of (X_1, X_2)					
		WG for FE-QM		WG for FE-GM		Pooled OLS for RE-GM	
	$n = 10$ and $J = 5$						
$t = 25$	Bias	-0.1100	-0.1633	1.0139	-2.5823	1.9077	-1.3909
	Stand error	0.3051	0.3933	5506.3	6203.39	4619.2	5261.9
	$n = 100$ and $J = 5$						
$t = 25$	Bias	-0.0683	-0.1251	-0.7609	-0.0259	0.0708	0.6426
	MSE	0.0312	0.0344	1303.5	1455.5	195.38	180.89
	$n = 10$ and $J = 5$						
$t = 50$	Bias	-0.0706	-0.1108	7.0584	-2.9577	5.8327	-5.0389
	MSE	0.0766	0.0908	9668.2	14676	11305	1.0087
$t = 50$	$n = 10$ and $J = 10$						
	Bias	-0.0149	-0.0249	7.0584	-2.9577	5.8327	-5.0389
	MSE	0.1398	1.3869	9668.2	14676	11305	10087
$t = 50$	$n = 100$ and $J = 5$						
	Bias	-0.0746	-0.1094	0.1738	1.8411	-0.2387	0.8257
	MSE	-0.0075	0.0098	8607.1	14083	7861.8	9347.2
	$n = 100$ and $J = 10$						
$t = 50$	Bias	-0.0199	-0.0327	0.1738	1.8411	-0.2387	0.8257
	MSE	0.0143	0.0163	8607.1	14083	7861.8	9347.2

Table 2:

Appendix A: Quantile Coupling Representation

Proof for Lemma 3.1: Decompose ζ_{ij} as $\zeta_{ij} \equiv \zeta_{ij}^A + \zeta_{ij}^B$, where

$$\zeta_{ij}^A \equiv \sqrt{m}\mathcal{E}_{ij:\tau} - \sqrt{m}\mathcal{E}_{ij:\tau}^* \tag{2}$$

and

$$\zeta_{ij}^B \equiv \sqrt{m}\mathcal{E}_{ij:\tau}^* - \frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)} Z_{ij} \tag{3}$$

and $\mathcal{E}_{ij:\tau}^*$ is the $\lceil \tau m \rceil$ −th smallest value of ϵ^* in the jth bin where $\epsilon^* = \epsilon(\mathcal{X}_{i[j]}, \eta_i, U_{it})$. Notice since U_{it} in the jth bin is still independent (conditional on X_{it} and η_i), both $\mathcal{E}_{ij:\tau}$ and $\mathcal{E}_{ij:\tau}^*$ are still conditionally independent. Furthermore since for $\epsilon^* = \epsilon(\mathcal{X}_{i[j]}, \eta_i, U_{it})$ being the j-th bin for a given i with the constants $\mathcal{X}_{i[j]}$ and η_i, ϵ^* is conditionally i.i.d. which allows for applying the quantile coupling results stated below.

By the definition of $\mathcal{E}_{ij:\tau}$ and $\mathcal{E}_{ij:\tau}^*$, where the difference is that for $\mathcal{E}_{ij:\tau}^*$, we keep the the first argument of $\epsilon(\cdot, \eta_i, U_{it})$ to be constant within each bin, so that the only variation of $\mathcal{E}_{ij:\tau}^*$ comes from the variation in the U_{it} component. Thus by the following Lemma C1, we have

$$\left|\zeta_{ij}^A\right| = \sqrt{m}\left|\mathcal{E}_{ij:\tau} - \mathcal{E}_{ij:\tau}^*\right| \leq C\frac{\sqrt{m}}{J}.$$

In order to bound ζ_j^B, we need Lemma C2; then we have for any fixed $l > 0$,

$$E\left|\zeta_{ij}^B\right|^l = O_p(m^{-l}).$$

The independence of Z_{ij} stems from being constructed by $\Phi^{-1}\left[F_\varepsilon(U_{it})\right]$, where U_{it} is independent for $1 \leq j \leq J$.

Q.E.D.

Proof for Lemma 3.2: Denote the set of indices in the jth bin as I_j, then $\mathcal{Y}_{ij:\tau} \leq \max_{k\in I_j}\left[\mathcal{X}_{i[j]}^\top\beta(\tau)+\eta_i(\tau)\right]+\mathcal{E}_{ij:\tau}$ and $\mathcal{Y}_{ij:\tau} \geq \min_{k\in I_j}\left[\mathcal{X}_{i[j]}^\top\beta(\tau)+\eta_i(\tau)\right]+\mathcal{E}_{ij:\tau}$. Then, according to Assumption 4, $\theta_{i,j}\left(\beta(\tau),\eta_i(\tau)\right) - \mathcal{X}_{i[j]}^\top\beta(\tau)-\eta_i(\tau) = O_p\left(\frac{1}{J}\right)$.

Proof for Proposition 4.1: Decompose ξ_{ij} as

$$\begin{aligned}\xi_{ij} &\equiv \sqrt{m}\mathcal{Y}_{ij:\tau} - \sqrt{m}\left[\mathcal{X}_{i[j]}^\top\beta(\tau)+\eta_i(\tau)\right] - \frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{ij}\\ &= \zeta_j^A + \zeta_j^B + \zeta_j^C\end{aligned}$$

where $\zeta_j^C = \sqrt{m}\mathcal{Y}_{ij:\tau} - \sqrt{m}\left[\mathcal{X}_{i[j]}^\top\beta(\tau)+\eta_i(\tau)\right] - \sqrt{m}\mathcal{E}_{ij:\tau} = \sqrt{m}\theta_{i,j}\left(\beta(\tau),\eta_i(\tau)\right) - \sqrt{m}\left[\mathcal{X}_{i[j]}^\top\beta(\tau)+\eta_i(\tau)\right]$, and ζ_j^A and ζ_j^B are the same as in Equations (2) and (3). Following Lemma 3.1 and Lemma 3.2, the proposition is proved.

Appendix B: Asymptotics of Within-group Estimators

Proof for Theorem (sequential asymptotic):

$$\begin{aligned}\widehat{\beta}(\tau) &= \beta(\tau) + \\ & \frac{1}{\sqrt{m}}\left\{\sum_{i=1}^{N}\sum_{j=1}^{J}(\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top}\right\}^{-1} \cdot \\ & \left\{\sum_{i=1}^{N}\sum_{j=1}^{J}\left[\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{ij} - \frac{1}{J}\sum_{s=1}^{J}\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{is}\right]\left[\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}\right\} \\ & + \frac{1}{\sqrt{m}}\left\{\sum_{i=1}^{N}\sum_{j=1}^{J}(\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top}\right\}^{-1} \cdot \\ & \left\{\sum_{i=1}^{N}\sum_{j=1}^{J}\left[\xi_{ij} - \frac{1}{J}\sum_{s=1}^{J}\xi_{is}\right]\left[\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]\right\}^{\top}.\end{aligned}$$

Hence for a given i, where N is finite and $J \to \infty$ (since $T \to \infty$) and we condition on the design $\{X_{ij}\}_{i=1,t=1}^{N,T}$:

$$\begin{aligned}& \frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left[\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{ij} - \frac{1}{J}\sum_{s=1}^{J}\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{is}\right]\left[\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top} \\ & \xrightarrow{d} N\left[0, \lim_{J\to\infty}\frac{1}{J}\sum_{j=1}^{J}\left[(\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top}\right] Var\left(\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{ij} - \frac{1}{J}\sum_{s=1}^{J}\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{is}\right)^{2}\right] \\ & \xrightarrow{d} N\left[0, \lim_{J\to\infty}\frac{1}{J}\sum_{j=1}^{J}\frac{\tau(1-\tau)}{[f_{\epsilon,ij}(0)]^{2}}\left[(\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top}\right]\right]\end{aligned}$$

Since conditional on $\{X_{ij}\}_{i=1,t=1}^{N,T}$, we have

$$E\left[\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{ij} - \frac{1}{J}\sum_{s=1}^{J}\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{is}\right] = 0$$

and

$$\begin{aligned}& Var\left[\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{ij} - \frac{1}{J}\sum_{s=1}^{J}\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{is}\right] \\ =\ & \left(1 - \frac{2}{J}\right)E\left[\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{ij}\right]^{2} + \frac{1}{J^{2}}\sum_{j=1}^{J}E\left[\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{ij}\right]^{2} \\ & \xrightarrow{p} \frac{\tau(1-\tau)}{[f_{\epsilon,ij}(0)]^{2}},\end{aligned}$$

then by the Lyapunov condition, we have the first equality.

Hence for fixed N and $T \to \infty$, we have

$$
\begin{aligned}
& \frac{\sqrt{T}}{\sqrt{m}} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{J} (\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})^{\top} \right\}^{-1} \cdot \\
& \sum_{i=1}^{N} \sum_{j=1}^{J} \left[\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)} Z_{ij} - \frac{1}{J} \sum_{s=1}^{J} \frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)} Z_{is} \right] \left[\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]} \right]^{\top} \\
= & \left\{ \sum_{i=1}^{N} \frac{\sum_{j=1}^{J} (\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})^{\top}}{J} \right\}^{-1} \cdot \\
& \sum_{i=1}^{N} \frac{1}{\sqrt{J}} \sum_{j=1}^{J} \left[\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)} Z_{ij} - \frac{1}{J} \sum_{s=1}^{J} \frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)} Z_{is} \right] \left[\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]} \right]^{\top} \\
& \xrightarrow{d} \sum_{i=1}^{N} \left\{ \lim_{J \to \infty} \frac{1}{J} \frac{\sum_{j=1}^{J} (\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})^{\top}}{J} \right\}^{-1} \cdot \\
& \sum_{i}^{N} N \left[0, \lim_{J \to \infty} \frac{1}{J} \sum_{j=1}^{J} \frac{\tau(1-\tau)}{[f_{\epsilon,ij}(0)]^2} (\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})^{\top} \right].
\end{aligned}
$$

Next, we will allow $N \to \infty$. Then by the Slutsky theorem we have

$$
\begin{aligned}
& \sum_{i=1}^{N} \left\{ \lim_{J \to \infty} \frac{1}{J} \frac{\sum_{j=1}^{J} (\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})^{\top}}{J} \right\}^{-1} \cdot \\
& \sqrt{N} \sum_{i=1}^{N} N \left[0, \lim_{J \to \infty} \frac{1}{J} \sum_{j=1}^{J} \frac{\tau(1-\tau)}{[f_{\epsilon,ij}(0)]^2} (\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})^{\top} \right] \\
& \xrightarrow{d} \left\{ \lim_{N \to \infty} \frac{1}{N} \left[\lim_{J \to \infty} \frac{\sum_{j=1}^{J} (\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})^{\top}}{J} \right] \right\}^{-1} \cdot \\
& N \left[0, \lim_{N \to \infty} \frac{1}{N} \left\{ \lim_{J \to \infty} \frac{1}{J} \sum_{j=1}^{J} \frac{\tau(1-\tau)}{[f_{\epsilon,ij}(0)]^2} (\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]})^{\top} \right\} \right].
\end{aligned}
$$

We bound $\frac{1}{\sqrt{J}} \sum_{j=1}^{J} \left[\xi_{ij} - \frac{1}{J} \sum_{s=1}^{J} \xi_{is} \right] \left[\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]} \right]^{\top}$ for fixed N and $T \to \infty$. By Chebyshev's inequality,

$$
\begin{aligned}
& \Pr \left[\left| \xi_{ij} \left(\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]} \right)^{\top} - E\xi_{ij} \left(\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]} \right)^{\top} \right| \geq C_1 \right] \\
\leq & \frac{Var\left(\xi_{ij}\right)}{C_{11}^2} \\
\leq & \frac{E\left(\xi_{ij}\right)^2}{C_{11}^2} = O\left(\frac{m}{J^2} + \frac{1}{m^2} \right).
\end{aligned}
$$

Since $m = T/J$, then when $J = T^a$ where $a \in (1/3, 1)$, we have

$$
\begin{aligned}
& \lim_{J \to \infty} \Pr \left[\left| \xi_{ij} \left(\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]} \right)^{\top} - E\xi_{ij} \left(\mathcal{X}_{i[j]} - \frac{1}{J} \sum_{s=1}^{J} \mathcal{X}_{i[s]} \right)^{\top} \right| \leq C_1 \right] \qquad (4) \\
= & \ 1.
\end{aligned}
$$

Applying Lemma C3 (Bernstein inequality), we have

$$\begin{aligned}
&\Pr\left[\left|\sum_{j=1}^{J}\left[\xi_{ij}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}-\sum_{j=1}^{J}E\left[\xi_{ij}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}\right|\geq\sqrt{J}C_2\right]\\
\leq\;& 2\exp\left(-\frac{JC_2^2}{2\sum_{j=1}^{J}Var\left[\xi_{ij}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}+\frac{2}{3}C_1\sqrt{J}C_2}\right)\\
=\;& 2\exp\left(-\frac{JC_2^2}{2\sum_{j=1}^{J}(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top}Var\left[\xi_{ij}\right]+\frac{2}{3}C_1\sqrt{J}C_2}\right)\\
=\;& \begin{array}{c} O\left[\exp\left(-\frac{JC_2^2}{2\sum_{j=1}^{J}(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top}Var\left[\xi_{ij}\right]}\right)\right]\\ \text{when } 2\sum_{j=1}^{J}(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top}Var\left[\xi_{ij}\right]\geq\frac{2}{3}C_1\sqrt{J}C_2,\\ \text{which is equivalent to } J=T^a,\ a\in(0,2/5)\cup(4/5,1)\,.\end{array}\\
&\text{or}\\
=\;& \begin{array}{c} O\left[\exp\left(-\frac{JC_2^2}{\frac{2}{3}C_1\sqrt{J}C_2}\right)\right]\\ \text{when } 2\sum_{j=1}^{J}(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top}Var\left[\xi_{ij}\right]\leq\frac{2}{3}C_1\sqrt{J}C_2,\\ \text{which is equivalent to } J=T^a,\ a\in(2/5,4/5)\,.\end{array}
\end{aligned}$$

Hence (a) when $J=T^a$, $a\in(0,2/5)\cup(4/5,1)$, we have

$$\begin{aligned}
&\exp\left(-\frac{JC_2^2}{2\sum_{j=1}^{J}\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^2Var\left[\xi_{ij}\right]}\right)\\
=\;&\exp\left(-\frac{C_3}{\frac{m}{J^2}+\frac{1}{m^2}}\right)\\
\to\;& 0 \text{ when } J=T^a, a\in(1/3,2/5)\,;\ \text{and}
\end{aligned}$$

(b) when $J=T^a$, $a\in(2/5,4/5)$, we have

$$\begin{aligned}
&\exp\left(-\frac{JC_2^2}{\frac{2}{3}C_1\sqrt{J}C_2}\right)\\
=\;&\exp\left(-\sqrt{J}C_3\right)\\
\to\;& 0 \text{ when } J=T^a, a\in(2/5,4/5)\,.
\end{aligned}$$

When $J=T^a$, $a\in(1/3,4/5)$,

$$\Pr\left[\left|\sum_{j=1}^{J}\left[\xi_{ij}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}-\sum_{j=1}^{J}E\left[\xi_{ij}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}\right|\geq\sqrt{J}C_2\right]\to 0,$$

which means that

$$\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left[\xi_{ij}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}-\frac{1}{\sqrt{J}}\sum_{j=1}^{J}E\left[\xi_{ij}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}\xrightarrow{p}0.$$

In the end, when $J=T^a$ with $a\in(1/2,2/3)$, we have

$$\frac{1}{\sqrt{J}}\sum_{j=1}^{J}E\left[\xi_{ij}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}\to 0,$$

which implies

$$\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left[\xi_{ij}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}\xrightarrow{p}0.$$

Similarly we can prove

$$\frac{1}{\sqrt{J}}\left[\sum_{s=1}^{J}\xi_{is}\right]\sum_{j=1}^{J}\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}\xrightarrow{p}0.$$

In the end, we have

$$\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left[\xi_{ij}-\frac{1}{J}\sum_{s=1}^{J}\xi_{is}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}\xrightarrow{p}0.$$

Q.E.D.

Proof for Theorem (joint asymptotic):

$$\begin{aligned}
\sqrt{NT}\left[\widehat{\beta}(\tau)-\beta(\tau)\right] &= \left\{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{J}\sum_{j=1}^{J}(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top}\right\}^{-1}\cdot \qquad (5)\\
&\quad \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left[\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{ij}-\frac{1}{J}\sum_{s=1}^{J}\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{is}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top} \qquad (6)\\
&\quad +\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{J}\sum_{j=1}^{J}(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top}\right\}^{-1}\\
&\quad \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left[\xi_{ij}-\frac{1}{J}\sum_{s=1}^{J}\xi_{is}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}. \qquad (7)
\end{aligned}$$

For the numerator of the first term on the right hand side, we have

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left[\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{ij}-\frac{1}{J}\sum_{s=1}^{J}\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{is}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}\\
\equiv\ &\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Pi_{iN}
\end{aligned}$$

where

$$\Pi_{iN}\equiv\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left[\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{ij}-\frac{1}{J}\sum_{s=1}^{J}\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)}Z_{is}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}.$$

Being conditional on the design $\{X_{ij}\}_{i=1,t=1}^{N,T}$,

$$\begin{aligned}
E(\Pi_{iN}) &= 0\\
Var(\Pi_{iN}) &= \frac{1}{J}\sum_{j=1}^{J}\left[(1-\frac{1}{J})\frac{\tau(1-\tau)}{[f_{\epsilon,ij}(0)]^2}\right](\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top}.
\end{aligned}$$

Hence by Lyapunov central limit theorem,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Pi_{iN}\xrightarrow{d}N\left[0,\lim_{N\to\infty}\frac{1}{N}\left\{\sum_{i=1}^{N}\frac{1}{J}\sum_{j=1}^{J}\frac{\tau(1-\tau)}{[f_{\epsilon,ij}(0)]^2}(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top}\right\}\right].$$

Next, for the numerator of the second term of Equation (5), we have

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left[\xi_{ij}-\frac{1}{J}\sum_{s=1}^{J}\xi_{is}\right]\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top} \\
\equiv\quad &\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\Lambda_{i1N}+\Lambda_{i2N}\right)
\end{aligned}$$

where

$$\begin{aligned}
\Lambda_{i1N} &\equiv \frac{1}{\sqrt{J}}\sum_{j=1}^{J}\xi_{ij}\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top} \\
.\Lambda_{i2N} &\equiv \frac{1}{\sqrt{J}}\sum_{s=1}^{J}\xi_{is}\sum_{j=1}^{J}\left[\mathcal{X}_{i[j]}-\frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}.
\end{aligned}$$

In order to prove the limiting normal distribution under the joint asymptotics, we need to show

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\Lambda_{i1N}+\Lambda_{i2N}\right)\xrightarrow{p} o_p(1)$$

which is sufficient to show

$$\max_{1\le i\le N}\left(\Lambda_{i1N}+\Lambda_{i2N}\right)\xrightarrow{p} o_p(\frac{1}{\sqrt{N}}).$$

To this end, we only need to show that for any $C_1>0$,

$$\Pr\left(\left|\max_{1\le i\le N}\left(\Lambda_{i1N}+\Lambda_{i2N}\right)\right|>\frac{C_1}{\sqrt{N}}\right)=o_p(1) \tag{8}$$

which is equivalent to

$$\max_{1\le i\le N}\Pr\left(\left|\Lambda_{i1N}+\Lambda_{i2N}\right|>\frac{C_1}{\sqrt{N}}\right)=o_p(\frac{1}{N}).$$

We have

$$
\begin{aligned}
& \Pr\left(|\Lambda_{i1N}| > \frac{C_1}{\sqrt{N}}\right) \\
= & \Pr\left(\left|\sum_{j=1}^{J} \xi_{ij}\left[\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}\right| > \frac{C_1\sqrt{J}}{\sqrt{N}}\right) \\
\leq & \Pr\left(\left|\sum_{j=1}^{J} \xi_{ij}\left[\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top} - E\sum_{j=1}^{J} \xi_{ij}\left[\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}\right| > \frac{C_1\sqrt{J}}{\sqrt{N}} - E\sum_{j=1}^{J} \xi_{ij}\left[\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}\right) \\
= & \Pr\left(\left|\sum_{j=1}^{J} \xi_{ij}\left[\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top} - E\sum_{j=1}^{J} \xi_{ij}\left[\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]^{\top}\right| > C_2\left(\frac{\sqrt{J}}{\sqrt{N}} + \sqrt{m} + \frac{J}{m}\right)\right) \\
\leq & 2\exp\left(-\frac{C_2^2\left(\frac{\sqrt{J}}{\sqrt{N}} + \sqrt{m} + \frac{J}{m}\right)^2}{2\sum_{j=1}^{J} Var\left[\xi_{ij}\right]\left[\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right] + \frac{2}{3}C_3\left(\frac{\sqrt{J}}{\sqrt{N}} + \sqrt{m} + \frac{J}{m}\right)}\right) \\
= & \quad O\left[\exp\left(-\frac{C_2^2\left(\frac{\sqrt{J}}{\sqrt{N}} + \sqrt{m} + \frac{J}{m}\right)^2}{2\sum_{j=1}^{J}(\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top} Var\left[\xi_{ij}\right]}\right)\right] = o\left[\exp\left(-\log N\right)\right] \\
& \quad \text{when } 2\sum_{j=1}^{J}(\mathcal{X}_{i[j]} - \tfrac{1}{J}\textstyle\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \tfrac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top} Var\left[\xi_{ij}\right] \geq \tfrac{2}{3}C_3\left(\tfrac{\sqrt{J}}{\sqrt{N}} + \sqrt{m} + \tfrac{J}{m}\right). \\
& \text{or} \\
= & \quad O\left[\exp\left(-\frac{C_2^2\left(\frac{\sqrt{J}}{\sqrt{N}} + \sqrt{m} + \frac{J}{m}\right)^2}{\frac{2}{3}C_3\left(\frac{\sqrt{J}}{\sqrt{N}} + \sqrt{m} + \frac{J}{m}\right)}\right)\right] = o\left[\exp\left(-\log N\right)\right] \\
& \quad \text{when } 2\sum_{j=1}^{J}(\mathcal{X}_{i[j]} - \tfrac{1}{J}\textstyle\sum_{s=1}^{J}\mathcal{X}_{i[s]})(\mathcal{X}_{i[j]} - \tfrac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]})^{\top} Var\left[\xi_{ij}\right] < \tfrac{2}{3}C_3\left(\tfrac{\sqrt{J}}{\sqrt{N}} + \sqrt{m} + \tfrac{J}{m}\right).
\end{aligned}
$$

where the second equality is by Proposition 4.1, and the second inequality is coming from Lemma C3 (Bernstein inequality) where conditions are satisfied by Equation (4) (when $J = T^a$ where $a \in (1/3, 1)$) and $\left\{\xi_{ij}\left[\mathcal{X}_{i[j]} - \frac{1}{J}\sum_{s=1}^{J}\mathcal{X}_{i[s]}\right]\right\}_{j=1}^{J}$ are independent conditional on the $\{X_{ij}\}_{i=1,t=1}^{N,T}$.

Hence we need to show

(a) under both conditions

$$
\frac{\frac{\sqrt{J}}{\sqrt{N}} + \sqrt{m} + \frac{J}{m}}{\frac{m}{J} + \frac{J}{m^2}} \to 0,
$$

and

$$
\frac{\left(\frac{m}{J} + \frac{J}{m^2}\right)\log N}{\left(\frac{\sqrt{J}}{\sqrt{N}} + \sqrt{m} + \frac{J}{m}\right)^2} \to 0
$$

then we have $\max_{1\leq i\leq N}\Pr\left(|\Lambda_{i1N}| > \frac{C_4}{\sqrt{N}}\right) = o_p(\frac{1}{N})$.

(b) under both conditions

$$
\frac{\frac{m}{J} + \frac{J}{m^2}}{\frac{\sqrt{J}}{\sqrt{N}} + \sqrt{m} + \frac{J}{m}} \to 0,
$$

and

$$
\frac{\log N}{\frac{\sqrt{J}}{\sqrt{N}} + \sqrt{m} + \frac{J}{m}} \to 0
$$

we will have $\max_{1\leq i\leq N}\Pr\left(|\Lambda_{i1N}| > \frac{C_4}{\sqrt{N}}\right) = o_p(\frac{1}{N})$.

Similarly we can prove $\max_{1\leq i\leq N}\Pr\left(|\Lambda_{i2N}| > \frac{C_5}{\sqrt{N}}\right) = o_p(\frac{1}{N})$ under the above (a) and (b) cases. Thus Equation (8) is proved.

Q.E.D.

Appendix C: Lemmas used in the Appendices A and B

Lemma C1: Let I be a closed interval and $\|f-g\|_{L^\infty(I)} \leq c$; then

$$\begin{aligned} & |\lceil \tau m \rceil \text{ order statistics of } f \text{ in the } I - \lceil \tau m \rceil \text{ order statistics of } g \text{ in the } I| \\ \leq \quad & c. \end{aligned}$$

Proof: $\lceil \tau m \rceil$order statistics of f and g in the I are monotone functional: $f \leq g \Longrightarrow \lceil \tau m \rceil$order statistics of f in the $I \leq \lceil \tau m \rceil$order statistics of g in the I. Thus if $\|f-g\|_{L^\infty(I)} \leq c$, then $f \leq g + c$ and

$$\begin{aligned} & \lceil \tau m \rceil \text{ order statistics of } f \text{ in the } I \\ \leq \quad & \lceil \tau m \rceil \text{ order statistics of } (f+c) \text{ in the } I \\ \leq \quad & \lceil \tau m \rceil \text{ order statistics of } g \text{ in the } I + c. \end{aligned}$$

By symmetry, we also get

$$\begin{aligned} & \lceil \tau m \rceil \text{ order statistics of } g \text{ in the } I \\ \leq \quad & \lceil \tau m \rceil \text{ order statistics of } f \text{ in the } I + c. \end{aligned}$$

Q.E.D.

Lemma C2: (a) Under Assumption QC, we have

$$\left| \sqrt{\frac{m}{\tau(1-\tau)}} f_{\epsilon,ij}(0) \mathcal{E}^*_{ij:\tau} - Z_{ij} \right| \leq \frac{C}{m} \left(1 + |Z_{ij}|^3 \right)$$

for $|Z_{ij}| \leq \kappa\sqrt{m}$, where constants $C, \kappa > 0$ do not depend on m.

(b) When Assumptions QC hold and constants $C', C'' > 0$ do not depend on m, we have

$$E \left| \sqrt{m} \mathcal{E}^*_{ij:\tau} - \frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon,ij}(0)} Z_{ij} \right|^l = O(m^{-l})$$

for any fixed $l > 0$.

Proof:

The main steps of proving QC inequality follow Brown, Cai and Zhou (2008) and Mason and Zhou (2012). Here our paper extends to the arbitrary $\lceil \tau m \rceil$th sample quantile with independent, not identically distributed distribution. Although the proof is elementary, only requiring the basic Taylor expansions and various exponential approximations, it is not simple at all. Therefore we collect and reorganize these procedures so that an expository proof is present.

In sum, the proof consists of:

Step 1: Exponentialize the density of the $\lceil \tau m \rceil$ $-$th sample quantile (or an asymptotic expansion of the density of the $\lceil \tau m \rceil$ $-$th sample quantile in terms of exponential approximations).

Step 2: Exponentialize the distribution of the $\lceil \tau m \rceil$ $-$th sample quantile (or an asymptotic expansion of the distribution of the $\lambda-$th sample quantile in terms of exponential approximations);

Step 3: Exponentialize the ratio of tails' standard normal distributions, for example, $\log \frac{1-\Phi(y+u)}{1-\Phi(y)}$ for some u;

Step 4: Use Proposition 3 in Mason and Zhou (2012) to finish the proof for the quantile coupling inequality. This finishes the proof for part (a) .

Step 5: In order to prove the bounds of quantile coupled moments, check the finiteness of order statistics.

Step 1 is the fundamental part of the proof, and other steps come directly from Theorem 6 in Cai and Zhou (2009). See Mason and Zhou (2012) for an excellent expository summary.

(Part a) Assume τm is an integer for simplification. Let $g(x)$ and $G(x)$ be the density and cumulative distribution function of $\mathcal{E}^*$, and $\phi(z)$ and $\Phi(z)$ denote respectively the density and cumulative distribution function of a standard normal random variable. Since in the j-th bin, all $f_{\epsilon,ij}(x)$ and $F_{\epsilon,ij}(x)$ have the same density and distribution functions, we suppress the i,j subscripts for $f_{\epsilon,ij}$ and $F_{\epsilon,ij}(x)$, and only denote them as f_ϵ and $F_\epsilon(x)$. The density of $\mathcal{E}^*$ is

$$g(x) = \frac{m!}{(\tau m)!\,(m-\tau m)!} F_{\epsilon^*}^{\tau m-1}(x)\left[1-F_{\epsilon^*}(x)\right]^{m-\tau m} f_{\epsilon^*}(x).$$

From Stirling's formula, $p! = \sqrt{2\pi} p^{p+1/2}\exp(-p+q)$ with $q = O(1/p)$, gives

$$\begin{aligned}
g(x) &= \frac{1}{\sqrt{2\pi}} \frac{m^{m+\frac{1}{2}}}{(\tau m-1)^{\tau m-1+\frac{1}{2}}(m-\tau m)^{m-\tau m+\frac{1}{2}}} \exp\left[-1+O(\frac{1}{m})\right] \\
&\quad \cdot F_\epsilon^{\tau m-1}(x)\left[1-F_\epsilon(x)\right]^{m-\tau m} f_\epsilon(x) \\
&= \frac{1}{\sqrt{2\pi}} \frac{m^{\frac{3}{2}}}{(\tau m-1)^{\frac{1}{2}}(m-\tau m)^{\frac{1}{2}}} \left(\frac{\tau m}{\tau m-1}\right)^{\tau m} \frac{\tau m-1}{\tau m} \exp\left[-1+O(\frac{1}{m})\right] \\
&\quad \cdot \left\{\left[\frac{F_\epsilon(x)}{\tau}\right]^{\frac{\tau m-1}{\tau m}} \left[\frac{1-F_\epsilon(x)}{1-\tau}\right]^{\frac{1-\tau}{\tau}}\right\}^{\tau m} f_\epsilon(x) \\
&= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{m}{\tau(1-\tau)}} O(1)\exp\left[1+O(\frac{1}{m})\right] O(1)\exp\left[-1+O(\frac{1}{m})\right] \\
&\quad \cdot \left\{\left[\frac{F_\epsilon(x)}{\tau}\right]^{\frac{\tau m-1}{\tau m}} \left[\frac{1-F_\epsilon(x)}{1-\tau}\right]^{\frac{1-\tau}{\tau}}\right\}^{\tau m} f_\epsilon(x) \\
&= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{m}{\tau(1-\tau)}} \left\{\left[\frac{F_\epsilon(x)}{\tau}\right]^{\frac{\tau m-1}{\tau m}} \left[\frac{1-F_\epsilon(x)}{1-\tau}\right]^{\frac{1-\tau}{\tau}}\right\}^{\tau m} f_\epsilon(x)\exp\left[O(\frac{1}{m})\right],
\end{aligned}$$

where the third equality comes from

$$\begin{aligned}
\frac{m^{\frac{3}{2}}}{(\tau m-1)^{\frac{1}{2}}(m-\tau m)^{\frac{1}{2}}} &= \sqrt{\frac{m}{\tau(1-\tau)}} O(1), \\
\left(\frac{\tau m}{\tau m-1}\right)^{\tau m} &= \exp\left[1+O(\frac{1}{m})\right], \\
\frac{\tau m-1}{\tau m} &= O(1).
\end{aligned}$$

Thus

$$\begin{aligned}
g(x) &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{m}{\tau(1-\tau)}} \left\{\left[\frac{F_{\epsilon^*}(x)}{\tau}\right]^{\frac{\tau m-1}{\tau m}} \left[\frac{1-F_{\epsilon^*}(x)}{1-\tau}\right]^{\frac{1-\tau}{\tau}}\right\}^{\tau m} f_{\epsilon^*}(x)\exp\left[O(\frac{1}{m})\right] \\
&= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{m}{\tau(1-\tau)}} (\frac{F_{\epsilon^*}(x)}{\tau})^{-1} \left\{\left[\frac{F_{\epsilon^*}(x)}{\tau}\right]^{\tau} \left[\frac{1-F_{\epsilon^*}(x)}{1-\tau}\right]^{1-\tau}\right\}^{m} f_{\epsilon^*}(x)\exp\left[O(\frac{1}{m})\right] \\
&= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{m}{\tau(1-\tau)}} \left\{\left[\frac{F_{\epsilon^*}(x)}{\tau}\right]^{\tau} \left[\frac{1-F_{\epsilon^*}(x)}{1-\tau}\right]^{1-\tau}\right\}^{m} f_{\epsilon^*}(x)\exp\left[O(\frac{1}{m})\right].
\end{aligned}$$

For the term $\left\{ \left[\frac{F_\epsilon(x)}{\tau}\right]^\tau \left[\frac{1-F_\epsilon(x)}{1-\tau}\right]^{1-\tau} \right\}^m$, we have

$$\begin{aligned}
& \left[\frac{F_\epsilon(x)}{\tau}\right]^{\tau m} \left[\frac{1-F_\epsilon(x)}{1-\tau}\right]^{(1-\tau)m} \\
= & \left[1+\frac{1}{F_\epsilon(0)}\left(F_\epsilon(x)-F_\epsilon(0)\right)\right]^{\tau m} \left[1-\frac{1}{1-F_\epsilon(0)}\left(F_\epsilon(x)-F_\epsilon(0)\right)\right]^{(1-\tau)m} \\
= & \left\{ \sum_{k=0}^{\tau m} \binom{\tau m}{k} \left[\frac{1}{\tau}\left[F_\epsilon(x)-F_\epsilon(0)\right]\right]^k \right\} \left\{ \sum_{k=0}^{m-\tau m} \binom{m-\tau m}{k} \left[-\frac{1}{1-\tau)}\left[F_\epsilon(x)-F_\epsilon(0)\right]\right]^k \right\} \\
= & 1+\left[\binom{\tau m}{2}\frac{1}{\tau^2}+\binom{m-\tau m}{2}\frac{1}{(1-\tau)^2}-m^2\right]\left[F_\epsilon(x)-F_\epsilon(0)\right]^2 \\
& +O\left(\left[\binom{\tau m}{2}\frac{1}{\tau^2}+\binom{m-\tau m}{2}\frac{1}{(1-\tau)^2}-m^2\right]\left[F_\epsilon(x)-F_\epsilon(0)\right]^2\right) \\
= & 1-\frac{m}{2\tau(1-\tau)}\left[F_\epsilon(x)-F_\epsilon(0)\right]^2+O\left(\left[\binom{\tau m}{2}\frac{1}{\tau^2}+\binom{m-\tau m}{2}\frac{1}{(1-\tau)^2}-m^2\right]\left[F_\epsilon(x)-F_\epsilon(0)\right]^2\right) \\
= & 1-\frac{m}{2\tau(1-\tau)}\left[\int_0^x \left[f_\epsilon(t)-f_\epsilon(0)\right]dt+xf_\epsilon(0)\right]^2+O\left(\left[\binom{\tau m}{2}\frac{1}{\tau^2}+\binom{m-\tau m}{2}\frac{1}{(1-\tau)^2}-m^2\right]\left[F_\epsilon(x)-F_\epsilon(0)\right]^2\right) \\
= & 1-\frac{m}{2\tau(1-\tau)}f_\epsilon^2(0)x^2+O(m\,|x|^4)
\end{aligned}$$

where the last equality is coming from $|f_\epsilon(u)-f_\epsilon(0)| \le Cu^2$.

Hence when x is chosen small enough, we have

$$\begin{aligned}
& \log\left\{ \left[\frac{F_\epsilon(x)}{\tau}\right]^{\tau m} \left[\frac{1-F_\epsilon(x)}{1-\tau}\right]^{(1-\tau)m} \right\} \\
= & -\frac{m}{2\tau(1-\tau)}f_\epsilon^2(0)x^2+O(m\,|x|^4).
\end{aligned}$$

In addition, Assumption QC (ii) implies that $f_\epsilon(x)/f_\epsilon(0)=1+O(|x|^2)=\exp(|x|^2)$.
Thus

$$\begin{aligned}
g(x) &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{m}{\tau(1-\tau)}}f_\epsilon(0)\exp\left[-\frac{m}{2\tau(1-\tau)}f_\epsilon^2(0)x^2+O(m\,|x|^4+x^2+\frac{1}{m})\right] \\
&= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{m}{\tau(1-\tau)}}f_\epsilon(0)\exp\left[-\frac{m}{2\tau(1-\tau)}f_\epsilon^2(0)x^2+O(m\,|x|^4+\frac{1}{m})\right].
\end{aligned}$$

According to this, we have our quantile coupling inequality following Theorem 6 in Cai and Zhou (2009).

(Part b) For the bound on the moment, and let $l>1$ be any finite integer, we have

$$\begin{aligned}
& E\left|\sqrt{m}\mathcal{E}_{ij:\tau}^*-\frac{\sqrt{\tau(1-\tau)}}{f_{\epsilon^*}(0)}Z_{ij}\right|^l \\
= & E\left\{\left|\sqrt{m}\mathcal{E}_{ij:\tau}^*-\frac{\sqrt{\tau(1-\tau)}}{f_\epsilon(0)}Z_{ij}\right|^l I\left\{|Z_{ij}|\le c\sqrt{m}\right\}\right\} \\
& +E\left\{\left|\sqrt{m}\mathcal{E}_{ij:\tau}^*-\frac{\sqrt{\tau(1-\tau)}}{f_\epsilon(0)}Z_{ij}\right|^l I\left\{|Z_{ij}|> c\sqrt{m}\right\}\right\} \\
= & \Xi_A+\Xi_B
\end{aligned}$$

where

$$\Xi_A \equiv E\left\{\left|\sqrt{m}\mathcal{E}^*_{ij:\tau} - \frac{\sqrt{\tau(1-\tau)}}{f_\epsilon(0)}Z_{ij}\right|^l I\left\{|Z_{ij}| \le c\sqrt{m}\right\}\right\}$$

$$\Xi_B \equiv E\left\{\left|\sqrt{m}\mathcal{E}^*_{ij:\tau} - \frac{\sqrt{\tau(1-\tau)}}{f_\epsilon(0)}Z_{ij}\right|^l I\left\{|Z_{ij}| > c\sqrt{m}\right\}\right\}.$$

By part a, we have

$$\begin{aligned}\Xi_A &= Cm^{-l}E\left\{\left(\left|1+|Z_{ij}|^3\right|\right)^l I\left\{|Z_{ij}| \le c\sqrt{m}\right\}\right\} \\ &= O\left(m^{-l}\right)\end{aligned}$$

where the equality comes from the finiteness of the absolute moments of the standard normal random variable.

For term Ξ_B,

$$\begin{aligned}\Xi_B &\le E\left\{\left|\sqrt{m}\mathcal{E}^*_{ij:\tau} - \frac{\sqrt{\tau(1-\tau)}}{f_\epsilon(0)}Z_{ij}\right|^{2l}\right\}^{1/2} \left\{\Pr(|Z_{ij}| > c\sqrt{m})\right\}^{1/2} \\ &\le E\left\{\left|\sqrt{m}\mathcal{E}^*_{ij:\tau} - \frac{\sqrt{\tau(1-\tau)}}{f_\epsilon(0)}Z_{ij}\right|^{2l}\right\}^{1/2} \left\{\frac{1}{2}\exp\left(-\frac{c^2}{2}m\right)\right\}^{1/2} \\ &= \left\{E\sum_{k=0}^{2l}\binom{2l}{k}\left[\sqrt{m}\mathcal{E}^*_{ij:\tau}\right]^{2l-k}\left(-\frac{\sqrt{\tau(1-\tau)}}{f_\epsilon(0)}Z_{i,j}\right)^k\right\}^{1/2} \left\{\frac{1}{2}\exp\left(-\frac{c^2}{2}m\right)\right\}^{1/2} \\ &\le C''m^{l/2}\left\{\frac{1}{2}\exp\left(-\frac{c^2}{2}m\right)\right\}^{1/2} \\ &= o(m^{-l})\end{aligned}$$

where the first equality is by the Cauchy–Schwarz inequality; the second inequality is from Mill's ratio inequality $\frac{\phi(x)}{1-\Phi(x)} > \max\left\{x, \frac{2}{\sqrt{2\pi}}\right\}$ for $x > 0$; the third inequality is due to the $l-$th finite moments of $\mathcal{E}^*_{i,j:\tau}$ for any integer $l > 1$ by Assumption QC (ii) from Cramer et al. (2002). The first equality is from Binomial theorem, while the last equality is because $\lim_{m\to\infty}\frac{m^{3l/2}}{e^m} = 0$ for a large l.

Q.E.D.

Lemma C3: (Bernstein in Serfling, 1980) Let $Y_1, \cdots, Y_n$ be independent random variables satisfying $\Pr\left(|Y_i - E(Y_i)| \le m\right) = 1$, each i, where $m < \infty$. Then, for $t > 0$,

$$\Pr\left(\left|\sum_{i=1}^n Y_i - \sum_{i=1}^n E(Y_i)\right| \ge nt\right) \le 2\exp\left(-\frac{n^2t^2}{2\sum_{i=1}^n Var(Y_i) + \frac{2}{3}mnt}\right),$$

for all $n = 1, 2, \cdots$.

References

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

[2] Angrist, Joshua, Victor Chernozhukov, and Iván Fernández-Val. "Quantile regression under misspecification, with an application to the US wage structure." Econometrica 74.2 (2006): 539-563.

[3] Belloni, A., Chernozhukov, V., & Fernández-Val, I. (2011). Conditional quantile processes based on series or many regressors.

[4] Breiman, Leo, and Jerome H. Friedman. "Estimating optimal transformations for multiple regression and correlation." Journal of the American statistical Association 80.391 (1985): 580-598.

[5] Brown, L. D., Cai, T. T., & Zhou, H. H. (2008). Robust nonparametric estimation via wavelet median regression. The Annals of Statistics, 2055-2084.

[6] Cai, T. T., & Zhou, H. H. (2009). Asymptotic equivalence and adaptive estimation for robust nonparametric regression. The Annals of Statistics, 37(6A), 3204-3235.

[7] Canay, I. A. (2011). A simple approach to quantile regression for panel data. The Econometrics Journal, 14(3), 368-386.

[8] Chamberlain, G. (1994). Quantile regression, censoring, and the structure of wages. In Advances in Econometrics: Sixth World Congress (Vol. 2, pp. 171-209).

[9] Chen, H. (2014). Sheep in Wolf's clothing: Using the least squares criterion for quantile estimation. Economics Letters, 125(3), 426-431.

[10] Cramer, E., Kamps, U., Rychlik, T., 2002. On the Existence of Moments of Generalized Order Statistics. *Statistics & Probability Letters*. 59, 397–404.

[11] Fernández-Val, I. (2005). Bias correction in panel data models with individual specific parameters. Available at SSRN 869104.

[12] Galvao, A. F., Lamarche, C., & Lima, L. R. (2013). Estimation of censored quantile regression for panel data with fixed effects. Journal of the American Statistical Association, 108(503), 1075-1089.

[13] Galvao, A. F., & Kato, K. (2014). Estimation and inference for linear panel data models under misspecification when both n and t are large. Journal of Business & Economic Statistics, 32(2), 285-309.

[14] Galvao, A. F., & Wang, L. (2015). Efficient minimum distance estimator for quantile regression fixed effects panel data. Journal of Multivariate Analysis, 133, 1-26.

[15] Gasser, T., Muller, H. G., 1979. Kernel Estimation of Regression Functions. Smoothing Techniques for Curve Estimation: Lecture Notes in Mathematics 757, 23–68.

[16] Graham, B. S., Hahn, J., & Powell, J. L. (2009). The incidental parameter problem in a non-differentiable panel data model. Economics Letters, 105(2), 181-182.

[17] Hahn, J., & Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. Econometrica, 72(4), 1295-1319.

[18] Hastie, T. J., & Tibshirani, R. J. (1990). Generalized additive models (Vol. 43). CRC Press.

[19] Kato, K., Galvao, A. F., & Montes-Rojas, G. V. (2012). Asymptotics for panel quantile regression models with individual effects. Journal of Econometrics, 170(1), 76-91.

[20] Kato, K., & Galvao, A. F. (2010). Smoothed quantile regression for panel data. mimeo, Hiroshima.

[21] Knight, K. (2001). Comparing conditional quantile estimators: rst and second order considerations. University of Toronto working paper.

[22] Koenker, R. (2004). Quantile regression for longitudinal data. Journal of Multivariate Analysis, 91(1), 74-89.

[23] Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. Econometrica: journal of the Econometric Society, 33-50.

[24] Koenker, R., & Hallock, K. F. (2001). Quantile Regression. The Journal of Economic Perspectives, 15(4), 143-156.

[25] Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. Journal of Econometrics, 157(2), 396-408.

[26] Lancaster, T. (2000). The incidental parameter problem since 1948. Journal of econometrics, 95(2), 391-413.

[27] Mammen, E., Linton, O., & Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. The Annals of Statistics, 27(5), 1443-1490.

[28] Mason, D. M., & Zhou, H. H. (2012). Quantile coupling inequalities and their applications. Probability Surveys, 9, 439-479.

[29] Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. Econometrica: Journal of the Econometric Society, 1-32.

[30] Nussbaum, M. (1996), "Asymptotic Equivalence of Density Estimation and Gaussian White Noise," *Annals of Statistics*, p2399-2430.

[31] Phillips, P. C., & Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. Econometrica, 67(5), 1057-1111.

[32] Phillips, P. C., & Moon, H. R. (2000). Nonstationary panel data analysis: an overview of some recent developments. Econometric Reviews, 19(3), 263-286.

[33] Rosen, A. M. (2012). Set identification via quantile restrictions in short panels. Journal of Econometrics, 166(1), 127-137.

[34] Sen, B., & Chaudhuri, P. (2011). Mahalanobis's fractile graphs: Some history and new developments. International Journal of Statistical Sciences, 11, 17-35.

[35] Serfling, R. J. (1980). Approximation theorems of mathematical statistics. Wiley series in probability and mathematical statistics.

[36] Zhou, H. H. (2006, June). A note on quantile coupling inequalities and their applications. In Submitted. Available from www. stat. yale. edu/■ hz68.

[37] Zheng, Z. G. and Yang, Y. (1998). Cross-validation and median criterion. Statistics Sinica, 8, 907-921.