Estimating Quantile Treatment Effects for Panel Data[*,†,‡]

Zongwu Cai[a], Ying Fang[b,c], Ming Lin[b,c], and Mingfeng Zhan[b]

[a]Department of Economics, University of Kansas, Lawrence, KS 66045, USA.

[b]Wang Yanan Institute for Studies in Economics and Fujian Key Laboratory of Statistical Sciences, Xiamen University, Xiamen, Fujian 361005, China.

[c]Department of Statistics and Data Science, School of Economics, Xiamen University, Xiamen, Fujian 361005, China.

February 4, 2022

Abstract: Motivated by the paper by Hsiao, Ching and Wan (2012), which proposed a factor-based model to estimate the average treatment effect with panel data, this paper proposes a quantile treatment effect model for panel data to characterize the distributional effect of a treatment. We propose to estimate the counterfactual quantile for the treated unit by using the relationship between conditional and unconditional distributions. Also, the asymptotic properties for the proposed quantile treatment effect estimator are established, together with discussing the choice of control units and covariates. A simulation study is conducted to illustrate our method. Finally, the proposed method is applied to estimate the quantile treatment effects of introducing CSI 300 index futures trading on both the log-return and volatility of the stock market in China.

Keywords: LASSO method; Panel data; Nonparametric estimation; Quantile regression; Treatment effects.

JEL classification: C13; C14; C33; C52; C54

[*]The authors acknowledge the partially financial supports from the National Science Fund of China (NSFC) for Distinguished Scholars with the grant number 71625001, the NSFC key projects with grant numbers 71631004 and 72033008, and the Science Foundation of Ministry of Education of China with the grant number 19YJA910003.

[†]*Corresponding author:* Y. Fang (E-mail: yifst1@xmu.edu.cn)

[‡]E-mail addresses: caiz@ku.edu (Z. Cai), yifst1@xmu.edu.cn (Y. Fang), linming50@xmu.edu.cn (M. Lin), zhanmf@stu.xmu.edu.cn (M. Zhan).

1 Introduction

Since the seminal work by Hsiao, Ching and Wan (2012, hereafter HCW), estimating average treatment effects (ATE) using panel data approach (PDA) has become widespread in empirical economics. HCW proposed to construct the counterfactual outcome of a treated unit by exploiting the cross-sectional dependence driven by a common factor structure for all units. Under a commonly adopted identification assumption that a policy intervention does not affect outcomes in the control group, HCW estimated the counterfactual outcome by a simple linear regression method without estimating factors and loadings. Moreover, the HCW's method has an advantage of least data demand which is a very desirable virtue in some empirical applications. The counterfactual outcome can be estimated using outcome variables of cross-sectional units and not necessary to use other covariates. For example, HCW estimated the impact on economic growth of political and economic integration of Hong Kong with mainland China only using the quarterly real GDP growth rates of 22 countries and regions. Recently, Bai, Li and Ouyang (2014), Ouyang and Peng (2015), and Li and Bell (2017) relaxed some assumptions made implicitly in HCW and instead of using AIC type approach to select control units, Li and Bell (2017) and Carvalho, Masini and Medeiros (2018) suggested using a LASSO type method to select control units when the number of control units is large with sparsity. Also, Li and Bell (2017) derived the asymptotic distribution of $\hat{\Delta}_1$ which facilitates inference and Carvalho, Masini and Medeiros (2018) provided a theoretical justification for the LASSO estimate.

By virtue of the aforementioned desirable properties, such as weak assumptions, least data demand and easy-to-implementation, the HCW's PDA has received increasing interests in policy evaluation literature. For example, some influential applications include, but not limit to, the papers by Chen et al. (2013) for investigating the effect of introducing index futures trading on the spot price volatility in the Chinese stock market, Bai, Li and Ouyang (2014) for exploring the influence of property taxes on home prices, Fujiki and Hsiao (2015) for measuring the net economic impact of the 1995 great Hanshin-Awaji earthquake, Du and Zhang (2015) for evaluating the effects of home-purchase restrictions and the trial property taxes on housing prices in China, Ouyang and Peng (2015) for studying the macroeconomic effect of the 2008 Chinese Economic Stimulus Program., Bove, Elia and Smith (2017) for estimating the economic effect of civil war, Ke, Hong and Hsiao (2017) for exploring the effect of high speed rail projects on the economic growth of targeted city with high speed rail nodes in China, Li and Long (2018) for examining the effect of the justice reform enacted on January 1, 1995 in Virginia, Carvalho, Masini and Medeiros (2018) for evaluating the impacts on inflation and other macroeconomic variables of an anti tax-evasion program implemented in Brazil. Recently, Ke and Hsiao (2021) applied the HCW's method to assess the evolution of economic consequences of the drastic lockdown policy in the epicenter of COVID-19 — the Hubei Province of China during worldwide curbs on economic activity.

However, the HCW's method is designed for estimating ATEs and it might not be sufficient in some applications for characterizing policy effects particularly when the outcome distribution is either asymmetric, heterogeneous or heavy-tailed. For example, how an introduction of futures trading affects spot stock volatility (VIX) is an important policy issue but still in a big debate. It is well documented in the literature that the distribution of spot stock VIX is skewed and heavily tailed. To address this issue, Chen et al. (2013) estimated the impact of introducing index futures trading on stock VIX by employing the HCW's PDA. But, the ATE would be distorted by the fact that VIX is usually asymmetric and heavily tailed. To gauge this phenomenon, let us look at the stock VIX of the monthly VIX of the stock market in China from January 2002 to February 2021, displayed in Figure 1 given in Section 4 for the estimated density of the pre-treatment, post-treatment and whole sample VIX of CSI 300 index[1]. Clearly, one can see from Figure 1 that the distribution of VIX is evidently asymmetric and heavily tailed and these phenomena can also be strongly supported from Table 5 in Section 4. To properly assess the impact of introducing futures trading on stock VIX, we will investigate distributional rather than mean effects on outcome variables of interest.

Motivated by the aforementioned empirical issues, this paper considers quantile treatment effects (QTE) to characterize the impact of a policy at any quantile corresponding to whole

[1]The CSI 300 is a capitalization-weighted stock market index designed to replicate the performance of the top 300 stocks traded on the Shanghai Stock Exchange and the Shenzhen Stock Exchange.

distributions of the observed outcomes and the counterfactual outcomes. Although there is a growing literature on identification and estimation of QTEs, see, for example, Firpo (2007), Firpo, Fortin and Lemieux (2009), Rothe (2010), Chernozhukov, Fernández-Val and Galichon (2010), Chernozhukov, Fernández-Val and Melly (2013), Cai (2021), Hsu, Lai and Lieli (2022), and among others, there are very limited efforts on the estimation of QTE with panel data, as argued by Cai (2021). Perhaps the most recent work includes the papers by Callaway, Li and Oka (2018) and Callaway and Li (2019), but both considered the estimation of QTE with panel data in the difference-in-differences setting with fixed time (two or three) periods. Therefore, to the best of our knowledge, there is still no literature on considering the estimation of QTE with large panel data as in HCW, as addressed by Cai (2021). Thus, the main contribution of this paper is to fill this gap. In other words, the novelty of our paper is to generalize the HCW's approach to the QTE setting for panel data to obtain a comprehensive examination of treatment effects. To estimate the counterfactual quantiles for the treated unit, different from the HCW's approach and its extensions, we introduce the conditional cumulative distributional function invariance assumption and propose a simple method to utilize the relationship between the conditional and unconditional cumulative distribution functions (CDF). In such a way, both nonparametric and parametric methods are considered to estimate the conditional CDF. The asymptotic properties for the proposed QTE estimators are also derived, together with an easily implemented method to construct a confidence interval based on a blockwise Bootstrap.

As argued in Xiao and Koenker (2009), the quantile estimation is an essential ingredient in modern risk management in financial applications so that the QTE technique can be used to evaluate if some new financial policy has a significant impact on the VIX and volatility-in-volatility[2], termed as VVIX in the finance literature, of financial markets or institutes. Actually, the motivation of this paper comes from investigating whether introducing CSI 300 index futures trading, formally introduced by the China Financial Futures Exchange on April 16, 2010, has an impact on spot market VIX and its VVIX or not, through the QTE

[2]The definition of volatility-in-volatility can found in Hollstein and Prokopczuk (2018), Huang et al. (2019), and Chen et al. (2021). Indeed, similar to VIX, Hollstein and Prokopczuk (2018) and Chen et al. (2021) provided some empirical evidences that the market VVIX can predict market returns and drive the time-varying volatility risk.

analysis. After the introduction, some criticize that the introduction of the index futures trading may shake the spot market due to the excessive speculation while others believe that the index futures market can improve the speed and quality of the information flows and make financial markets more complete. By using the proposed modeling approach, we will estimate the impact of introducing CSI 300 index futures trading on both VIX and VVIX of the stock market in China. The detailed analysis results are reported in Section 4.

The rest of the paper is organized as follows. Section 2 first reviews briefly the literature about studies on treatment effects for panel data and then describes the general setting for the proposed model. Also, in the same section, we discuss the method of estimating QTE for panel data, and establish the asymptotic property of the estimator, together with an easily implemented method to construct a confidence interval based on a blockwise Bootstrap. Furthermore, we discuss the estimation of the conditional CDF with many control units based on quantile regression method and dimension reduction approach to choose control units and covariates. Section 3 provides Monte Carlo simulation studies for accessing the finite sample performance for the proposed estimators. Section 4 presents an empirical application which estimates the QTEs of introducing CSI 300 index futures trading on both the log-return and the volatility of the Chinese stock market. Section 5 concludes. All proofs are given in the Appendix.

2 QTE for Panel Data

2.1 A Primer on the HCW's Method

This section is devoted to introducing the framework of the HCW's method. To this end, we first define some important notations. Let y_{it}^1 denote the outcome for the ith unit in period t with treatment, and y_{it}^0 denote the outcome for the ith unit in period t with the absence of treatment, $1 \leq i \leq N$ and $1 \leq t \leq T$. The treatment effect for the ith unit at time t is defined as

$$\Delta_{it} = y_{it}^1 - y_{it}^0.$$

Since y_{it}^1 and y_{it}^0 cannot be simultaneously observed, the observed outcome is given by

$$y_{it} = d_{it}y_{it}^1 + (1 - d_{it})y_{it}^0,$$

where $d_{it} = 1$ if the ith unit is under the treatment at time t, and $d_{it} = 0$ otherwise.

The HCW's model focuses on the case where there is only one treated unit that receives a treatment at time $T_1 + 1$ for some $1 < T_1 < T$, and thereafter. Without loss of generality, it is assumed that it is the first unit. In other words, there is no treatment for y_{jt} with units $j = 2, \cdots, N$ and all $t = 1, \cdots, T$, while for the first unit, there is no treatment for y_{1t} with $t = 1, \cdots, T_1$ either, and the treatment only occurs for y_{1t} with $t = T_1 + 1, \cdots, T$. HCW used the following factor structure to model the cross-sectional dependence across all units

$$y_{it}^0 = \alpha_i + b_i^\top f_t + u_{it}, \quad i = 1, \cdots, N; \quad t = 1, \cdots, T,$$

where α_i is the ith individual specific intercept, b_i is a $K \times 1$ vector of factor loadings, f_t is a $K \times 1$ vector (unobservable) common factors and u_{it} is a (mean zero) weakly dependent and stationary error term. If both T and N are large, the method of Bai and Ng (2002) or the approach in Pesaran (2006) can be adopted to estimate the common factors f_t's. In the case that neither T nor N is large, HCW suggested a novel method by using $\tilde{y}_t = (y_{2t}, \cdots, y_{Nt})^\top$, which is called the control units, in lieu of f_t, to predict the counterfactual outcome y_{1t}^0 for post-treatment periods.

Specifically, as in HCW, it is assumed the correlations among cross-sectional units are due to some common factors that drive all cross-sectional units, although their impacts on each cross-sectional unit may be different. Therefore, by the correlation between y_{1t} and $\{y_{jt}\}_{j=2}^N$, HCW estimated the counterfactual outcome y_{1t}^0 based on the following regression model

$$y_{1t}^0 = \beta^\top x_t + u_{1t}, \qquad t = 1, \cdots, T_1,$$

where $x_t = (1, y_{2t}, \cdots, y_{Nt})^\top$, $\beta = (\beta_1, \cdots, \beta_N)^\top$ is a vector of unknown coefficients, and u_{1t} is a zero mean and finite variance idiosyncratic error term. The estimation of the coefficient β is then obtained by the ordinary least squared regression method

$$\hat{\beta} = \arg\min_{\beta} \sum_{t=1}^{T_1} (y_{1t} - \beta^\top x_t)^2.$$

By assuming that the data structure remains the same before and after the treatment and other assumptions; see, for example, the assumptions listed in HCW for details, the predicted counterfactual outcome of y_{1t}^0 for $t = T_1 + 1, \cdots, T$ can be estimated by

$$\hat{y}_{1t}^0 = \hat{\beta}^\top x_t, \qquad t = T_1 + 1, \cdots, T.$$

Finally, the estimator of the average treatment effect for the first unit

$$\Delta_1 = E(\Delta_{1t}) = E(y_{1t}^1 - y_{1t}^0)$$

can be constructed by averaging the difference between the observed outcomes and estimated counterfactual outcomes over the post-treatment period

$$\hat{\Delta}_1 = \frac{1}{T_2} \sum_{t=T_1+1}^{T} \left(y_{1t}^1 - \hat{y}_{1t}^0 \right),$$

where $T_2 = T - T_1$. Also, HCW suggested using the AIC of Akaike (1974) and the AICC in Hurvich and Tsai (1989) model selection methods to see if there is a need to use all cross-sectional units in their approach, which can be improved by a LASSO type method as proposed and studied by Li and Bell (2017) and Carvalho, Masini and Medeiros (2018).

2.2 Model Setup For QTE

It is well known that if the distributions of potential outcomes poorly concentrate around the mean or the heterogeneity of outcomes exists, the ATE may not be an ideal representative for the effect of a treatment. To grab the global effect of a treatment, it is natural to study the distributional treatment effects. In this section, our focus is on the estimation of the QTE with panel data.

The setting of panel data is the same as that in HCW. Suppose that we have a panel data set $\{(y_{it}, z_t); 1 \leq i \leq N, 1 \leq t \leq T\}$, where z_t is a $d_z \times 1$ vector of covariates[3]. Also, it is assumed that the time series in the panel data is strictly stationary. For the first unit, a treatment is implemented from $t = T_1 + 1$ to $t = T$ and no treatment occurs before $t = T_1 + 1$.

[3]For example, in our empirical study in Section 4, it includes 3 macroeconomic variables such as the monthly CPI growth rate, the monthly M1 growth rate, and the monthly M2 growth rate.

The remaining $N-1$ units as a control group remain untreated throughout the time. Denote $T_2 = T - T_1$. For simplicity, let $\lim_{T\to\infty} T_2/T_1 = c$ throughout the paper, where $c > 0$ is a constant, so that $\lim_{T\to\infty} T_1/T = \lambda$ and $\lim_{T\to\infty} T_2/T = 1-\lambda$ with $\lambda = 1/(1+c)$. To ease notation, the panel data are divided into four parts as follows:

$$\begin{array}{c|c} Y_1 & X_1 \\ \hline Y_2 & X_2 \end{array},$$

where $Y_1 = \{y_{1t}, t = 1, \cdots, T_1\}$ represents the first unit from $t = 1$ to $t = T_1$, $X_1 = \{(y_{it}, z_t); i = 2, \cdots, N, \text{ and } t = 1, \cdots, T_1\}$ stands for the remaining units from $t = 1$ to $t = T_1$, $Y_2 = \{y_{1t}, t = T_1+1, \cdots, T\}$ is the information for the first unit from $t = T_1+1$ to $t = T$, and $X_2 = \{(y_{it}, z_t); i = 2, \cdots, N, \text{ and } t = T_1+1, \cdots, T\}$ is for the remaining units from $t = T_1 + 1$ to $t = T$. Let $Y_2^0 = \{y_{1t}^0, t = T_1 + 1, \cdots, T\}$ be the counterfactual outcome of Y_2. For convenience, the observed outcome Y_2 is also denoted as $Y_2^1 = \{y_{1t}^1, t = T_1 + 1, \cdots, T\}$. Then, the QTE for the first unit after $t = T_1$ is defined as

$$\Delta_\tau = q_{1\tau}^1 - q_{1\tau}^0, \tag{1}$$

where $q_{1\tau}^j$ is the τth quantile of $F_j(y) = P(y_{1t}^j \le y)$ for $j = 0$ and 1 and $\tau \in (0, 1)$. The next subsection describes details on how to estimate Δ_τ.

2.3 Estimation Procedures

To estimate the QTE, Δ_τ, it suffices to estimate $q_{1\tau}^1$ and $q_{1\tau}^0$, respectively. Since the outcome for the first unit under treatment with $t > T_1$ is observable, the sample quantile of Y_2 is simply used, denoted as $\hat{q}_{1\tau}^1$, to estimate $q_{1\tau}^1$. The difficulty in estimating QTE is due to the fact that Y_2^0 is not observable, so that it is not straightforward to estimate $q_{1\tau}^0$. To estimate the counterfactual quantile for the first unit, different from the factor augmented idea in HCW, a new method is proposed by utilizing the relationship between the conditional and unconditional CDFs, described as follows.

To be specific, $q_{1\tau}^0$ is written as

$$q_{1\tau}^0 = \inf\left\{y : F_{Y_2^0}(y) \ge \tau\right\} = \inf\left\{y : E[F_{Y_2^0|X_2}(y|X_{2t})] \ge \tau\right\},$$

where $F_{Y_2^0}(\cdot)$ is the CDF of y_{1t}^0 for $t > T_1$ and $F_{Y_2^0|X_2}(\cdot|\cdot)$ denotes the conditional CDF of y_{1t}^0 given X_2 for $t > T_1$, which leads to an estimator of $q_{1\tau}^0$ as

$$\overline{q}_{1\tau}^0 = \inf\left\{ y : \frac{1}{T_2} \sum_{t=T_1+1}^{T} F_{Y_2^0|X_2}(y|X_{2t}) \geq \tau \right\}. \tag{2}$$

Generally speaking, the conditional CDF $F_{Y_2^0|X_2}(y|x)$ is unknown, so that the above estimator of $q_{1\tau}^0$ is infeasible. To get a feasible estimate of $q_{1\tau}^0$ from the observed data, it needs to estimate $F_{Y_2^0|X_2}(y|x)$ first. To this end, the following assumption is needed, which indeed, is similar to the assumption in a mean setting imposed in HCW and the aforementioned references and the quantile setting as in Callaway, Li and Oka (2018).

Assumption 1: (Conditional CDF Invariance) The structures of the conditional CDFs of $Y_1|X_1$ and $Y_2^0|X_2$ are the same; that is, $F_{Y_1|X_1}(\cdot|\cdot) = F_{Y_2^0|X_2}(\cdot|\cdot) \equiv F(\cdot|\cdot)$.

Assumption 1 postulates some kind of structure invariance, which ensures that given the outcomes of the control group and the covariates, the conditional distribution of the potential outcome of the treated unit without treatment remains the same before and after the treatment. With Assumption 1, it is then possible to estimate the counterfactual conditional CDF in the treated group by the observed data before treatment. This assumption corresponds to Assumption 1 in Rothe (2010) for a nonparametric structural model. Also, Hsu, Lai and Lieli (2022) adopted the same kind of assumption (see their Assumption 2.3) in the counterfactual treatment effects settings.

Clearly, under Assumption 1, a kernel method such as Nadaraya-Watson estimation method or other procedures, can be used to estimate $F(y|x)$ if $d_x = N - 1 + d_z$ is not very large. Specifically, $F(y|x)$ can be estimated using the observed data before treatment as follows

$$\tilde{F}(y|x) = \frac{\sum_{t=1}^{T_1} I\left(Y_{1t} \leq y\right) K_h\left(X_{1t} - x\right)}{\sum_{t=1}^{T_1} K_h\left(X_{1t} - x\right)}, \tag{3}$$

where $I(\cdot)$ is the indicator function, $K_h(X_{1t} - x) = h^{-d_x} K((X_{1t} - x)/h)$, $K(\cdot)$ is a higher-order kernel function as defined in Assumption 4[4], and h is bandwidth. Then, we plug the

[4]See Gasser, Müller and Mammitzsch (1985) for details on the definition of higher-order kernel.

estimated conditional CDF into (2) to obtain

$$\tilde{q}_{1\tau}^0 = \inf\left\{y : \frac{1}{T_2}\sum_{t=T_1+1}^{T} \tilde{F}_{Y_2^0|X_2}(y|X_{2t}) \geq \tau\right\} = \inf\left\{y : \tilde{F}_{Y_2^0}(y) \geq \tau\right\},$$

where $\tilde{F}_{Y_2^0}(y) := \frac{1}{T_2}\sum_{t=T_1+1}^{T} \tilde{F}_{Y_2^0|X_2}(y|X_{2t})$. Note that when high-order kernels are used in (3), $\tilde{F}_{Y_2^0}(y)$ could be non-monotonic or take values outside the $[0,1]$ interval. One can use the re-weighting method in Rothe (2010) or the monotonization method in Hsu, Lieli and Lai (2022) to turn $\tilde{F}_{Y_2^0}(y)$ into a monotonically nondecreasing CDF. Here, we follow Hsu, Lieli and Lai (2022) and let

$$\hat{F}_{Y_2^0}(y) = \sup_{u \leq y} \tilde{F}_{Y_2^0}(u) \ / \sup_{-\infty < u < \infty} \tilde{F}_{Y_2^0}(u).$$

Obviously, $\hat{F}_{Y_2^0}(y)$ is a CDF with probability one. Then, the estimator of the QTE for the first unit is given by

$$\hat{\Delta}_\tau = \hat{q}_{1\tau}^1 - \hat{q}_{1\tau}^0, \tag{4}$$

where $\hat{q}_{1\tau}^0 = \inf\left\{y : \hat{F}_{Y_2^0}(y) \geq \tau\right\}$.

2.4 Asymptotic Theory

In this subsection, we derive the asymptotic results for the proposed estimator of the QTE defined in (4). To establish the asymptotic results, it is common to impose a time series structure to the panel data such as strictly stationary and α-mixing. To this end, some extra assumptions are needed and listed as follows:

Assumption 2: The time series in the panel data is strictly stationary with α-mixing coefficient satisfying $\alpha(s) = O(s^{-\varepsilon_0})$ for some $\varepsilon_0 > 5/2$. Also, assume that $\lim_{T\to\infty} T_2/T_1 = c$ with $0 < c < \infty$ so that $\lambda = \lim_{T\to\infty} T_2/T \in (0,1)$.

Assumption 3: (1) The supports of X_{1t} and X_{2t}, denoted by $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively, are compact and they satisfy that $\mathcal{X}_2 \subset \mathcal{X}_1$. (ii) The density function of X_{1t} is denoted by $f_{X_1}(x)$. Assume that it is uniformly continuous and bounded away from zero on $\mathcal{X}_1$. (iii) For each $x \in \mathcal{X}_2$, $F(y|x)$ is continuous in y and its conditional density function $f(y|x)$ is bounded.

Assumption 4: The kernel function $K(\cdot)$ satisfies (i) $K(u) = 0$ if $|u| > 1$. (ii) $\int K(u)du = 1$. (iii) For some $\ell > d_x$, $\int u^j K(u)du = 0$ for $1 \leq j \leq \ell$ and $\int \left|u^\ell K(u)\right| du < \infty$. (iv) $K(u)$ is ℓ-times differentiable and the derivatives are uniformly continuous and bounded.

Assumption 5: (i) For $j = 0$ and 1, the stationary distribution of $\{y_{1t}^j, t > T_1\}$ has density function $f_{Y_2^j}(y)$ that is bounded away from zero on the support of the distribution. (ii) The density function $f_{Y_2^j}(y)$ is twice differentiable.

Assumption 6: (i) The density function $f_{X_1}(x)$ is ℓ-times differentiable on the interior of $\mathcal{X}_1$ and the derivatives are uniformly continuous and bounded. (ii) The density function $f_{X_2}(x)$ is ℓ-times differentiable on the interior of $\mathcal{X}_2$ and the derivatives are uniformly continuous and bounded. (iii) The conditional CDF $F(y|x)$ is ℓ-times differentiable with respect to x on the interior of $\mathcal{X}_1$ and the derivatives are uniformly continuous and bounded.

Assumption 7:[5] $T^{1/2}h^{d_x}/\log(T) \to \infty$ and $T^{1/2}h^\ell \to 0$ as $T \to \infty$.

The asymptotic normality of the proposed estimator $\hat{\Delta}_\tau$ is presented in the following theorem with its detailed proof given in the Appendix. Now, define

$$\sigma_\tau^2 = \sum_{h=-\infty}^{\infty} \text{Cov}(\xi_t, \xi_{t-h}) + \sum_{h=-\infty}^{\infty} \text{Cov}(\eta_t, \eta_{t-h})$$

with $\xi_t = \psi_{1\tau}(Y_{2t}) - \psi_{0\tau}(X_{2t})$, $\eta_t = \psi_{2\tau}(X_{1t}, Y_{1t})$, $\psi_{0\tau}(X_{2t}) = [\tau - F(q_{1\tau}^0|X_{2t})]/f_{Y_2^0}(q_{1\tau}^0)$, $\psi_{1\tau}(Y_{2t}^1) = [\tau - I(Y_{2t}^1 \leq q_{1\tau}^1)]/f_{Y_2^1}(q_{1\tau}^1)$, and $\psi_{2\tau}(X_{1t}, Y_{1t}) = \sqrt{c}\, f_{X_2}(X_{1t})[I(Y_{1t} \leq q_{1\tau}^0) - F(q_{1\tau}^0|X_{1t})]/[f_{Y_2^0}(q_{1\tau}^0) f_{X_1}(X_{1t})]$. According to Davydov's inequality for α-mixing, which can be found in the book by Hall and Heyde (1980) and by Assumption 2, one can show easily that σ_τ^2 exists.

Theorem 1. *Under Assumptions 1-7, then, we have*

$$\sqrt{T_2}(\hat{\Delta}_\tau - \Delta_\tau) \xrightarrow{d} N\left(0, \sigma_\tau^2\right),$$

where σ_τ^2 is the asymptotic variance.

[5]In practice, h might be taken to be under-smoothed so that this assumption is satisfied, as seen in our Monte Carlo simulation study in Section 3.

Consequently, Theorem 1 implies that $\hat{\Delta}_\tau = \Delta_\tau + O_p(T_2^{-1/2})$ so that it is consistent. Furthermore, it shows that, although the kernel method is used to estimate the conditional CDF, the proposed QTE estimators can still achieve the $\sqrt{T_2}$ convergence rate. Also, Theorem 1 gives clearly that it would be easy to construct $(1-\alpha)100\%$ confidence interval (CI) for Δ_τ for given τ as $\hat{\Delta}_\tau \pm z_{\alpha/2}/\sqrt{T_2}\sigma_\tau$ if σ_τ^2 would be known, where $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$ is the critical value. One way to estimate consistently σ_τ^2 is to employ the heteroskedasticity and autocorrelation consistent (HAC) estimation of Newey and West (1987) by using $\hat{\xi}_t$ and $\hat{\eta}_t$. However, due to the complicated structure of σ_τ^2, it might not be easy to obtain a consistent estimate of σ_τ^2. Therefore, one might use a Bootstrap approach instead, described in the next subsection.

2.5 A Bootstrap Inference

The blockwise Bootstrap method as in Künsch (1989) is applied here to construct a CI for Δ_τ, described as follows. First, for the sample $W_{1t} = (Y_{1t}, X_{1t})$, $t = 1, 2, \cdots, T_1$, each block is constructed as $V_{1t} = \{W_{1t}, W_{1,t+1}, \cdots, W_{1,t+b_1-1}\}$, where $t = 1, 2, \cdots, T_1 - b_1 + 1$ and $b_1 = \lfloor \sqrt[3]{T_1} \rfloor$, where $\lfloor x \rfloor$ denotes the maximum integer less than or equal to x, is the length of each block, and for the sample $W_{2t} = (Y_{2t}, X_{2t})$, $t = T_1+1, T_1+2, \cdots, T$, each block is constructed as $V_{2t} = \{W_{2t}, W_{2,t+1}, \cdots, W_{2,t+b_2-1}\}$, where $t = T_1+1, T_1+2, \cdots, T-b_2+1$ and $b_2 = \lfloor \sqrt[3]{T_2} \rfloor$ is the length of each block. Second, draw $l_1 = \lfloor T_1/b_1 \rfloor$ sample with replacement from $V_1 = \{V_{11}, V_{12}, \cdots, V_{1,T_1-b+1}\}$ and denote the re-sampling sample as $V_1^* = \{V_{11}^*, V_{12}^*, \cdots, V_{1l_1}^*\}$; draw $l_2 = \lfloor T_2/b_2 \rfloor$ sample with replacement from $V_2 = \{V_{2,T_1+1}, V_{2,T_1+2}, \cdots, V_{2,T-b_2+1}\}$ and denote the re-sampling sample as $V_2^* = \{V_{21}^*, V_{22}^*, \cdots, V_{2l_2}^*\}$. Third, define $W_1^* = \{W_{11}^*, W_{12}^*, \cdots, W_{1,b_1l_1}^*\} = \{V_{11}^*, V_{12}^*, \cdots, V_{1l_1}^*\}$ and $W_2^* = \{W_{21}^*, W_{22}^*, \cdots, W_{2,b_2l_2}^*\} = \{V_{21}^*, V_{22}^*, \cdots, V_{2l_2}^*\}$, where $b_1l_1 \approx T_1$ and $b_2l_2 \approx T_2$, based on the re-sampling sample W_1^* and W_2^*, the QTE can be estimated by the proposed method. Finally, repeat the procedures above for B (say, $B = 1000$) times, then the CI can be calculated by the resulting estimators. Actually, this blockwise Bootstrap procedure is implemented in Section 4 for our empirical study.

2.6 Choosing Control Units and Covariates

In this subsection, our attempt is paid to considering the case where there are many control units and covariates, that is, d_x is large. For such a case, it is not ideal to estimate $F(\cdot|x)$ by kernel regression method due to the so-called curse of dimensionality. To avoid this problem, one can follow the idea in Aït-Shahalia and Brant (2001) and Hall and Yao (2005) to adopt the index approach as $\beta^\top x$ to estimate $F(\cdot|\beta^\top x)$. As argued in Aït-Shahalia and Brant (2001), indeed, from a statistical perspective, the index avoids the curse of dimensionality because it allows us to reduce the multivariate problem to one where we can implement the nonparametric approach described above in a univariate setting since $\beta^\top x$ is univariate. Another approach is to estimate the conditional CDF via quantile regression as in Koenker and Bassett (1978), which is described below in detail.

Let $q_\tau(x)$ be the τth conditional quantile of $F(y|x)$ so that $q_\tau(x) = F^{-1}(\tau|x) \equiv q(\tau, x)$. Then, we can use $q(\tau, x)$ to estimate $F(y|x)$ if $q(\tau, x)$ is estimable. Indeed, a simple calculation leads to the following relationship between conditional CDF and conditional quantile,

$$F(y|x) = \int_0^1 I(q(u,x) \le y)du \approx \varepsilon + \int_\varepsilon^{1-\varepsilon} I(q(u,x) \le y)du \tag{5}$$

for some small constant $\varepsilon > 0$. Note that the approximation above is for computational convenience. By assuming that the τth conditional quantile function of Y_1 given X_1 is $q_\tau(x) = \beta_\tau^\top x$, which includes the model in HCW as a special case. Then, the conditional quantile is estimated as $\hat{q}_\tau(x) = \hat{\beta}_\tau^\top x$, where

$$\hat{\beta}_\tau = \arg\min_{\beta_\tau} \sum_{t=1}^{T_1} \rho_\tau(y_{1t} - \beta_\tau^\top X_{1t}) \tag{6}$$

and $\rho_\tau(v) = v[\tau - I(v < 0)]$. Therefore, in view of (5), the estimated conditional CDF, $\hat{F}(y|x)$ becomes to

$$\hat{F}(y|x) = \varepsilon + \int_\varepsilon^{1-\varepsilon} I(\hat{\beta}_u^\top x \le y)du \approx \varepsilon + \sum_{j=1}^{m} \delta_j I(\hat{\beta}_{\tau_j}^\top x \le y), \tag{7}$$

where $\hat{\beta}_{\tau_j}$ can be obtained from (6) for any $\varepsilon \le \tau_0 < \cdots < \tau_m \le 1 - \varepsilon$ and $\delta_j = \tau_j - \tau_{j-1} \to 0$ as $m \to \infty$. Note that the last approximation in (7) might be sensitive to the choice of $\{\tau_j\}_{j=0}^m$ in real applications. Also, the above idea is used in Chernozhukov, Fernández-Val and Galichon (2010) and Chernozhukov, Fernández-Val and Melly (2013), respectively. Therefore, if the time series in the panel data is stationary and α-mixing, similar to Proposition 5 of Chernozhukov, Fernández-Val and Galichon (2010), one can show that $\hat{F}(y|x)$ is consistent and enjoys the asymptotic normality as follows:

$$\sqrt{T_1}\left[\hat{F}(y|x) - F(y|x)\right] \xrightarrow{d} N(0, V(y|x)),$$

where $V(y|x) > 0$ is the asymptotic variance, which might depend on x and y.

In real applications, if the dimension of cross-section and covariates is large and a sparsity exists, using all of the control units and covariates may result in unstable estimation. For such a case, we suggest using the method of quantile regression with a penalty similar to that in Li and Zhu (2008) and Wu and Liu (2009) to estimate the conditional quantile function. To be specific, a penalty term is added into (6)

$$\hat{\beta}_{\text{pen},\tau} = \arg\min_{\beta_\tau} \sum_{t=1}^{T_1} \rho_\tau(y_{1t} - \beta_\tau^\top X_{1t}) + \sum_{j=1}^{d_x} \psi_{\lambda^*}(\beta_{\tau,j}) \tag{8}$$

for some penalty function $\psi_{\lambda^*}(\cdot)$, say the absolute function as in Li and Zhu (2008) or the smoothly clipped absolute deviation penalty function as in Wu and Liu (2009). Note that one can easily implement (8) by using the *rqPen* package in R in Sherwood and Maidman (2016) or *hreg* package in Yi and Huang (2017). Then, the conditional distribution with penalty is then estimated as

$$\hat{F}_{\text{pen}}(y|x) = \varepsilon + \int_{\varepsilon}^{1-\varepsilon} I(\hat{\beta}_{\text{pen},\tau}^\top x \le y) d\tau$$

for some small constant $\varepsilon > 0$. Analog to the result in Li and Zhu (2008), the asymptotic normality may be obtained for dependent data, in particular for sequences that satisfy sufficiently some strong mixing conditions. Actually, via a Monte Carlo simulation study in Section 3, we explore the effective of the proposed estimator based on the penalized quantile regression method.

Finally, one extra advantage of using the penalized quantile regression method given in (8) is that by some simple extensions, one can even deal with the case when the ultra high dimension of cross-section and covariates is possibly larger than the dimension of time series ($d_x > T_1$); see Wang, Wu and Li (2012), Sherwood and Maidman (2016), and Yi and Huang (2017) for details. Such an extension for ultra high dimensional cases is warranted as future research.

3 Monte Carlo Simulation Studies

In this section, a series of Monte Carlo experiments are conducted to study the finite sample performances of the proposed QTE estimators for panel data. The first simulation is designed for low dimensional case in which conditional CDF can be estimated by the kernel method. The second simulation illustrates the method of parametric quantile regression in the estimation of the conditional CDF when the number of the control units is moderate. The last simulation is conducted to show the performance of the quantile regression method with penalty when facing the high dimension case with sparsity. For the first two simulations, the sample sizes are set to be $T = 200$, 400 and 800 with $\lambda = T_1/T = 1/2$, while for the last simulation, the sample sizes are set to be $T = 100$, 200 and 400 with $\lambda = T_1/T = 1/2$, and 1000 Monte Carlo simulations are carried out for each setting. To assess the performance of the proposed estimators, we calculate the median of the 1000 absolute errors (MAE), which is MAE$= |\hat{\Delta}_\tau - \Delta_\tau|$, and its standard deviation (SD, presented in the parentheses) for the QTE estimator $\hat{\Delta}_\tau$ with $\tau = 0.25$, 0.5 and 0.75. In what follows, we use the AR(1) model with coefficient φ in $\xi_t = \varphi\xi_{t-1} + \upsilon_t$ specified later, where $\upsilon_t = \zeta_t - 1$, and $\zeta_t \sim \mathit{Exponential}(1)$, such that υ_t is a sequence of white noises with mean 0 and variance 1.

Example 1: Suppose $N = 3$. That is, there are two control units so that $d_x = 2$. Let

$$y_{1t} = \frac{1}{\sqrt{5}} y_{2t} + \frac{2}{\sqrt{5}} \sin(y_{3t}) + \sqrt{y_{2t}^2 + y_{3t}^2} \cdot \varepsilon_t, \quad 1 \le t \le T_1,$$

$$y_{1t}^0 = \frac{1}{\sqrt{5}} y_{2t} + \frac{2}{\sqrt{5}} \sin(y_{3t}) + \sqrt{y_{2t}^2 + y_{3t}^2} \cdot \varepsilon_t, \quad T_1 + 1 \le t \le T,$$

and $y_{1t}^1 = \rho_t + y_{1t}^0$ for $T_1 + 1 \le t \le T$, where $\{y_{2t}\}_{t=1}^T$, $\{y_{3t}\}_{t=1}^T$, $\{\rho_t\}_{t=T_1+1}^T$ and $\{\varepsilon_t\}_{t=1}^T$ are independent, $\{y_{2t}\}_{t=1}^{T_1}$ and $\{y_{3t}\}_{t=1}^{T_1}$ are generated from an AR(1) model with coefficient $\varphi = 0.6$, $\{y_{2t}\}_{t=T_1+1}^T$ and $\{y_{3t}\}_{t=T_1+1}^T$ are generated from an AR(1) model with coefficient

$\varphi = 0.4$, $\{\rho_t\}_{t=T_1+1}^T$ is generated from an AR(1) model with coefficient $\varphi = 0.5$, and $\{\varepsilon_t\}_{t=1}^T$ is a sequence of white noises with the same distribution as $\{v_t\}_{t=1}^T$. Then, $Y_1|X_1$ and $Y_2^0|X_2$ have the same conditional CDF. The true QTE can be calculated numerically by simulation with a very large sample size for $\tau \in (0,1)$. Since the dimension $d_x = 2$ is low in this simulation, the conditional CDF of $Y_1|X_1$ is estimated by the nonparametric method described in Section 2.3. To construct a higher order kernel $K(\cdot)$ satisfying Assumption 4, let $K(u) = \prod_{j=1}^{N-1} e_1^\top S^{-1}(1, u_j, \cdots, u_j^p)^\top k(u_j)$, where $e_1 = (1, 0, \cdots, 0)^\top$ is the unit vector, $S = (\mu_{i+r})_{0\le i,r\le p}$ is a matrix with the element $\mu_j = \int u^j k(u)du$, $p = \ell - 2$, and $k(u)$ is a univariate Epanechnikov kernel function, so that $K(u)$ is a fourth-order kernel and $p = 2$. The bandwidth satisfying Assumption 7 is given by $h = 3.12\sigma_{X_1} T_1^{-1/6}$, where 3.12 is the rule-of-thumb bandwidth constant for fourth-order Epanechnikov kernel with 2-dimensional covariates X_1, and σ_{X_1} is the sample standard deviation of X_1. Table 1 represents the results for the simulation. From Table 1, it can be seen that the finite sample performance of the proposed QTE estimators is well-behaved in the sense that both the MAE and the SD are generally small and decrease rapidly with the sample size. Also note that by increasing the sample size T_2 from 100 to 400, the MAE values decrease by almost a half, indicating that the convergence of the estimators is indeed at the $\sqrt{T_2}$ rate.

Table 1: Simulation Results for Example 1

(T_1, T_2)	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
(100, 100)	0.189 (0.180)	0.194 (0.184)	0.263 (0.239)
(200, 200)	0.141 (0.135)	0.145 (0.137)	0.192 (0.176)
(400, 400)	0.095 (0.089)	0.096 (0.094)	0.134 (0.130)

Example 2: Suppose $N = 8$. That is we have 7 control units so that $d_x = 7$ without sparsity. Let

$$y_{1t} = \frac{1}{\sqrt{7}} \sum_{i=2}^{8} y_{it} + \varepsilon_t,\ 1 \le t \le T_1, \quad y_{1t}^0 = \frac{1}{\sqrt{7}} \sum_{i=2}^{8} y_{it} + \varepsilon_t,\ T_1 + 1 \le t \le T,$$

and $y_{1t}^1 = \rho_t + y_{1t}^0$ for $T_1 + 1 \le t \le T$, where $\{y_{2t}\}_{t=1}^T, \cdots, \{y_{8t}\}_{t=1}^T$, $\{\rho_t\}_{t=T_1+1}^T$ and $\{\varepsilon_t\}_{t=1}^T$ are independent, $\{y_{2t}\}_{t=1}^{T_1}, \cdots, \{y_{8t}\}_{t=1}^{T_1}$ are generated from an AR(1) model with coefficient

$\varphi = 0.6$, $\{y_{2t}\}_{t=T_1+1}^{T}, \cdots, \{y_{8t}\}_{t=T_1+1}^{T}$ are generated from an AR(1) model with coefficient $\varphi = 0.4$, $\{\rho_t\}_{t=T_1+1}^{T}$ is generated from an AR(1) model with coefficient $\varphi = 0.5$, and $\{\varepsilon_t\}_{t=1}^{T}$ is a sequence of white noises with the same distribution as $\{\upsilon_t\}_{t=1}^{T}$. It can be seen that the conditional CDFs of $Y_1|X_1$ and $Y_2^0|X_2$ remain the same. The true QTE can be calculated numerically by simulation with a very large sample size for $\tau \in (0,1)$. Due to the moderate dimension of the covariates ($d_x = 7$) and limmited sample size, a nonparametric method might not work well so that the quantile regression (QR) method is used to estimate the conditional CDF of $Y_1|X_1$, as in (7). Next, we calculate the MAE and its SD for the QTE estimator $\hat{\Delta}_\tau$ among the 1000 replications. The results are summarized in Table 2, from which one can see that the MAE and the SD are generally small and become smaller with larger sample size, suggesting that the finite sample performance of the proposed QTE estimators is satisfactory. By increasing the sample size from T_2 from 100 to 400, the MAE values decrease by nearly a half, indicating that the convergence of the estimators is at the $\sqrt{T_2}$ rate.

Table 2: Simulation Results for Example 2

(T_1, T_2)	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
(100, 100)	0.290 (0.240)	0.307 (0.254)	0.357 (0.311)
(200, 200)	0.194 (0.179)	0.214 (0.180)	0.267 (0.219)
(400, 400)	0.146 (0.131)	0.147 (0.137)	0.191 (0.176)

Example 3: In this example, we illustrate the performance of the penalized quantile regression method in the proposed QTE estimators. Suppose there are $d_x = 40$ control units, which is large, with a huge of sparsities. Let

$$y_{1t} = \frac{5}{2} \cdot y_{2t} + \frac{7}{2} \cdot y_{3t} + 0 \cdot \sum_{i=4}^{41} y_{it} + \varepsilon_t, \; 1 \le t \le T_1, \quad y_{1t}^0 = \frac{5}{2} \cdot y_{2t} + \frac{7}{2} \cdot y_{3t} + 0 \cdot \sum_{i=4}^{41} y_{it} + \varepsilon_t, \; T_1+1 \le t \le T,$$

and $y_{1t}^1 = \rho_t + y_{1t}^0$ for $T_1 + 1 \le t \le T$, where $\{y_{2t}\}_{t=1}^{T}, \cdots, \{y_{41,t}\}_{t=1}^{T}$, $\{\rho_t\}_{t=T_1+1}^{T}$ and $\{\varepsilon_t\}_{t=1}^{T}$ are independent, $\{y_{2t}\}_{t=1}^{T_1}, \cdots, \{y_{41,t}\}_{t=1}^{T_1}$ are generated from an AR(1) model with coefficient $\varphi = 0.6$, $\{y_{2t}\}_{t=T_1+1}^{T}, \cdots, \{y_{41,t}\}_{t=T_1+1}^{T}$ are generated from an AR(1) model with coefficient $\varphi = 0.4$, $\{\rho_t\}_{t=T_1+1}^{T}$ is generated from an AR(1) model with coefficient $\varphi = 0.5$ and $\{\varepsilon_t\}_{t=1}^{T}$ is a sequence of white noises with the same distribution as $\{\upsilon_t\}_{t=1}^{T}$. Then, the conditional CDFs of $Y_1|X_1$ and $Y_2^0|X_2$ are indeed the same. The true QTE can be calculated numerically by simulation with a very large sample size for $\tau \in (0,1)$. Next, we consider both the quantile regression (QR) approach as in (7) and the penalized quantile regression (PQR) method with absolute penalty function as in (8) to calculate the MAE and SD value for the QTE estimator $\hat{\Delta}_\tau$ with $\tau = 0.25$, 0.5 and 0.75, respectively. The results are depicted in Table 3, from which one can observe that when N is large and T is small, the MAE and the SD calculated by PQR method are smaller than those computed based on the QR procedure, which, as expected, are in line with our theory. In addition, as T_2 increases from 50 to 200, the MAE based on the PQR method decreases by roughly a half, which suggests that the convergence of the estimator is at the $\sqrt{T_2}$ rate.

Table 3: Simulation Results for Example 3

Method	(T_1, T_2)	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
QR	(50,50)	0.976 (0.978)	0.868 (0.826)	1.009 (1.130)
	(100,100)	0.296 (0.270)	0.314 (0.269)	0.373 (0.340)
	(200,200)	0.185 (0.163)	0.195 (0.171)	0.242 (0.220)
PQR	(50,50)	0.357 (0.340)	0.378 (0.351)	0.467 (0.432)
	(100,100)	0.242 (0.222)	0.262 (0.225)	0.304 (0.298)
	(200,200)	0.181 (0.153)	0.187 (0.165)	0.233 (0.205)

4 Empirical Analysis

4.1 Data and Descriptive Statistics

In this section, our method is illustrated by estimating the QTEs of introducing CSI 300 index futures trading on the spot price volatility of the Chinese stock market and its volatility, respectively. As part of financial reform, the CSI 300 index futures contracts were formally introduced by the China Financial Futures Exchange on April 16, 2010. Since then, China opens her own futures market. Whether the introduction of the futures trading has a positive impact on the stock market in China or not is a controversial issue in finance literature. Some criticized that the introduction of the index futures trading may shake the spot market due to the excessive speculation while others believed that the index futures market can improve the speed and quality of the information flows and make the financial markets more complete. To identify the impact of introducing CSI 300 index futures trading on the volatility of the stock market in China, we take the geographical tie and trade relations into account, such that 13 major international market indices are selected as the control units, which include the Hang Seng Index (HSI), the Hang Seng China Affiliated Corporation Index (HSCCI), Korean Composite Stock Price Index, Japanese Nikkei 225 Index, Singaporean Strait Times Index, Taiwanese Composite Index, the FTSE 100 Index in KU, the S&P 500 Index in USA, Franch CAC 40 Index, German Frankfurt DAX Index, Brazilian Bovespa Index, Canadian S&P/TSX Composite Index, and Australian All Ordinaries Index. In addition, 3 macroeconomic variables are also included: the monthly CPI growth rate, the monthly M1 growth rate, and the monthly M2 growth rate. The time period of the data is from January 2002 to February 2021. All the market indices are collected from the Resset Financial Research Database[6] and the macroeconomic data are from the CEIC Database[7]. The monthly stock log-returns[8] of the 14 market indices are calculated by the difference of the log-returns between the last day and the first day in a month. Therefore, the total sample size $T = 230$. The descriptive statistics for the log-returns of the 14 indices and the 3 macroeconomic variables are reported in Table 4, from which one can see that for most of market indices, their distribution of the log-return is almost symmetric.

Following Chen et al. (2013), the monthly stock volatilities of the 14 market indices are calculated as the standard deviation of daily index returns multiplied by the square root of the number of trading days in that month. The descriptive statistics for the volatilities of the 14 indices are reported in Table 5, from which it can be observed that the distributions of volatility for all 14 market indices are asymmetric (see Column 8 in Table 5) and heavy-tailed (see Column 7 in Table 5), which are strongly supported by observing Figure 1.

For the sample period from January 2002 to June 2011, Chen et al. (2013) employed the panel data policy evaluation approach by HCW to construct counterfactuals of the spot market volatility, mainly based on the correlations between China and international

[6] http://www.resset.cn/endatabases

[7] https://www.ceicdata.com

[8] The monthly log-return is computed as $r_t = \log p_t - \log p_{t-1}$, where p_t is the closing price at the last day of the t month, p_{t-1} is the closing price at the first day of the t month.

Table 4: Descriptive Statistics of Monthly Return

Index	Mean	Std. Dev.	Median	Min.	Max.	Kurt.	Skew.
CSI 300	-0.052	1.056	-0.064	-3.807	4.544	6.350	0.424
HSI	-0.088	0.867	-0.065	-2.535	2.303	3.295	0.106
HSCCI	-0.113	1.063	-0.100	-3.426	2.993	3.317	-0.022
Korea	-0.051	0.827	-0.065	-2.730	2.498	3.982	-0.189
Japan	-0.077	0.949	-0.097	-2.650	3.146	4.128	0.464
Singapore	-0.023	0.800	-0.090	-2.340	5.368	13.436	1.760
Taiwan	0.037	0.737	-0.037	-2.452	2.921	4.171	0.211
UK	-0.160	0.813	-0.254	-1.686	4.224	7.182	1.325
US	-0.085	0.821	-0.148	-2.656	4.673	9.084	1.371
France	-0.075	0.927	-0.106	-2.262	4.199	5.728	0.972
Germany	-0.106	0.943	-0.122	-2.423	4.337	6.298	1.023
Brazil	-0.141	1.083	-0.280	-3.380	3.600	4.030	0.430
Canada	0.009	0.767	-0.006	-1.513	5.016	12.538	2.042
Australia	-0.012	0.603	-0.012	-1.933	2.187	4.533	0.190
CPI growth rate	0.001	0.028	0.002	-0.095	0.058	4.106	-0.615
M1 growth rate	0.134	0.075	0.127	0.000	0.390	3.053	0.554
M2 growth rate	0.149	0.047	0.142	0.080	0.297	3.861	0.800

The monthly stock log-return is calculated as 100 multiplied by the difference of the log-returns between the last day and the first day in a month. CPI, M1 and M2 growth rates denote the monthly growth rate compared to those in the same month of the previous year.

Table 5: Descriptive Statistics of Monthly Volatility

Index	Mean	Std. Dev.	Median	Min.	Max.	Kurt.	Skew.
CSI 300	0.066	0.033	0.057	0.013	0.184	4.568	1.334
HSI	0.056	0.032	0.047	0.020	0.325	26.522	3.777
HSCCI	0.068	0.034	0.060	0.026	0.317	15.987	2.712
Korea	0.053	0.030	0.046	0.017	0.249	12.208	2.381
Japan	0.059	0.031	0.053	0.019	0.318	24.820	3.284
Singapore	0.044	0.030	0.037	0.013	0.256	18.776	3.284
Taiwan	0.049	0.025	0.042	0.016	0.142	4.812	1.401
UK	0.047	0.029	0.039	0.012	0.231	13.735	2.711
USA	0.046	0.034	0.037	0.011	0.276	18.382	3.287
France	0.058	0.034	0.049	0.017	0.248	10.028	2.240
Germany	0.059	0.034	0.050	0.018	0.239	9.239	2.165
Brazil	0.073	0.037	0.065	0.028	0.360	28.565	4.116
Canada	0.040	0.031	0.032	0.011	0.290	29.700	4.398
Australia	0.040	0.024	0.035	0.012	0.222	20.896	3.370

The monthly stock index volatility is calculated as the standard deviation of daily index returns multiplied by the square root of the number of trading days in that month.

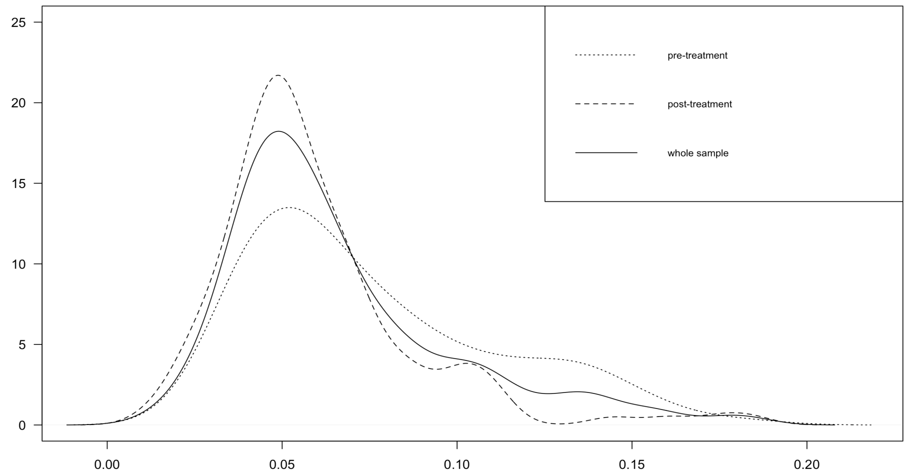

Figure 1: The plot of the estimated density for the pre-treatment, post-treatment and whole sample VIX of CSI 300 index.

stock markets, and draw the conclusion that the introduction of index futures trading can significantly reduce the volatility of the Chinese stock market.

However, different from Chen et al. (2013), we consider the QTEs of the index futures trading on both the log-return and volatility of the Chinese stock market with similar datasets but with different time periods. According to the introduction date of the CSI 300 index futures, the whole time period is divided into two sections: the pre-treatment period from January 2002 to April 2010 which consists of the sample size of $T_1 = 100$ observations and the post-treatment period from May 2010 to February 2021 which consists of the sample size of $T_2 = 130$ observations so that $\lambda = 10/23$. Finally, considering the sample size and the number of the control units, the conditional CDF in this application is estimated by the quantile regression method since $d_x = 16$, which is moderate.

4.2 QTE of Futures Trading on Stock Returns

First, we study the QTE of introducing the index futures trading on the monthly log-return of the stock market in China. Now, we implement the method proposed in this paper to calculate the estimated QTEs of CSI 300 index futures trading on the log-return (y_{1t}) of the Chinese stock market. Figure 2 presents the estimated QTEs of the CSI 300 index futures trading on the log-return of the Chinese stock market, together with 95% CI (the red shaded area) for each quantile based on the blockwise Bootstrap with $B = 1000$ replications. Also, the ATE $\hat{\Delta}_1$, calculated by the HCW's approach is plotted by the horizontal (blue) line, together with its 95% CI (the blue shaded area).

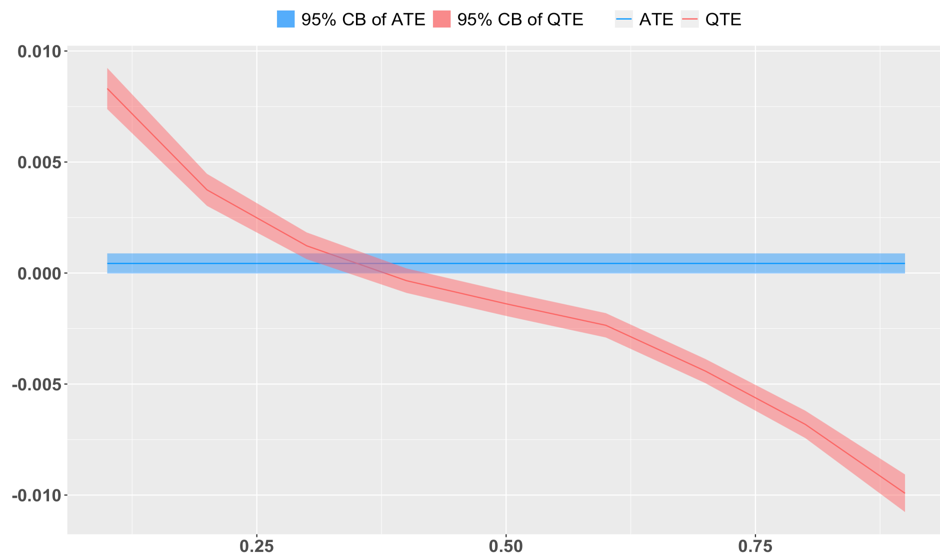

Figure 2: The plot of the estimated QTE is in the red line, $\hat{\Delta}_\tau$ versus τ, together with its 95% CI (the shaded area) based on the blockwise Bootstrap. The horizontal (blue) line is $\hat{\Delta}_1$, the ATE calculated by the HCW's approach.

From Figure 2, it can be seen that first, the 95% CI for $\hat{\Delta}_1$ contains basically zero, which implies that Δ_1 should be zero so that the average return is not affected by the introduction of the CSI 300 index futures trading, which, as expected, is not surprising. Second, the estimated QTEs changes (decreases almost linearly) with τ and are significantly positive at the lower quantiles (about 0.008 at the 10% quantile), while significantly negative at the higher quantiles (about -0.01 at the 90% quantile), which indicates that the introduction of the CSI 300 index futures trading has different impacts on the log-return of the Chinese stock market at different quantiles. Indeed, a quantile of log-return can be used to characterize the risk of log-return, as argued by Xiao and Koenker (2009). For example, a lower quantile corresponds to the Value-at-Risk (VaR), a well-known downside risk measure in finance literature. The positive QTEs at the lower quantiles indicate that the introduction of the CSI 300 futures trading can reduce the VaR by making the negative log-return less negative. Meanwhile, the negative QTEs at the higher quantiles suggest that the introduction of the

CSI 300 futures trading can also reduce the VaR by making the positive log-return less positive. In conclusion, similar to Chen et al. (2013), introducing the CSI 300 futures trading makes the stock market in China more stable in terms of the VaR. However, Chen et al. (2013) did not find such an asymmetric effect and then it is hard for them to empirically interpret why introducing the futures trading can stabilize the spot stock market.

4.3 QTE of Futures Trading on Volatility of Stock Markets

In this subsection, we study the QTEs of introducing the index futures trading on the volatility of the stock market in China. Different from the previous section, y_{1t} in this section is volatility instead of log-return. As discussed in Section 1, the QTE can be used to evaluate the impact of volatility of volatility, see, for example, Huang et al. (2019) for details. Next, we implement the proposed method to calculate the QTEs of CSI 300 index futures trading on the volatility of the Chinese stock market. Figure 3 depicts the estimated QTEs for the CSI 300 index futures trading on the volatility of the Chinese stock market, together with 95% CI (the red shaded area) for each quantile, which is obtained via the blockwise Bootstrap with $B = 1000$ replications. Also, for a comparison purpose, the ATE $\hat{\Delta}_1$ calculated by the HCW's approach is plotted by the horizontal (blue) line, together with its 95% CI (the blue shaded area).

From Figure 3, first, it is not surprising to see that the estimated median effect $\hat{\Delta}_{\tau=1/2}$ is not the same as the ATE $\hat{\Delta}_1$, because the distribution of volatility is asymmetric and heavily tailed, as seen in Figure 1. Second, it is clear to see that the estimated QTEs decreases in two phases (two piecewise linear) and are significantly negative and the volatility at the higher quantile is much negative than that at the lower quantile, which decreases about 0.054 at the 90% quantile compared to only 0.01 at the 10% quantile. The ATE by the HCW's approach gives a general effect of the CSI 300 index futures trading on the volatility variation while the QTE by the proposed method can offer more details for the effect of the CSI 300 index futures trading on the variation of the volatility. Overall, introducing the CSI 300 futures trading can reduce the volatility of the stock market and the higher the volatility is, the more significant the treatment effect demonstrates. The results are consistent with the findings in

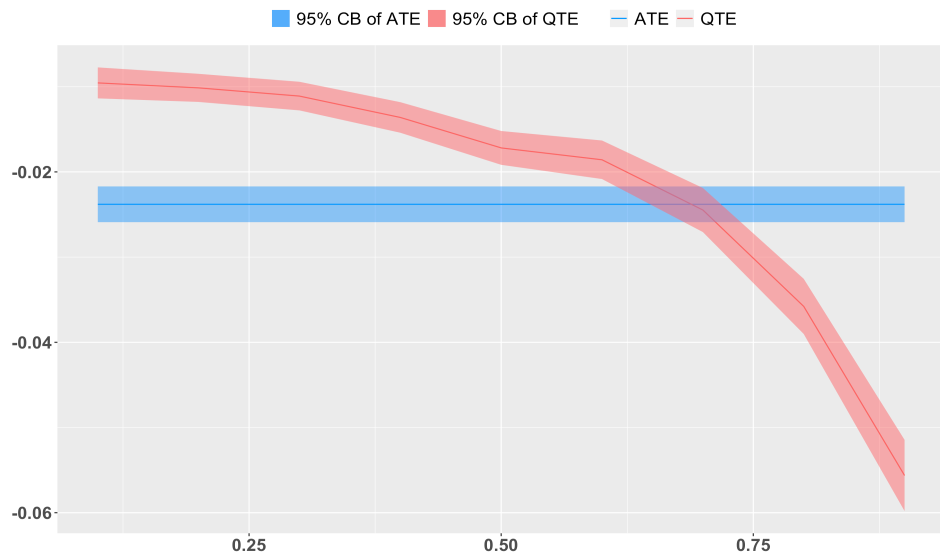

Figure 3: The plot of the estimated QTE is in the red line, $\hat{\Delta}_\tau$ versus τ, together with its 95% CI (the red shaded area) based on the blockwise Bootstrap. The horizontal (blue) line is $\hat{\Delta}_1$, the ATE calculated by the HCW's approach.

Subsection 4.2 in terms of the Value-at-Risk and both suggest that introducing the CSI 300 futures market can make the stock market more stable in China.

5 Conclusion

To grab a more comprehensive effect of a treatment, this paper considers the estimation of QTE with panel data which generalizes HCW's approach from a mean setting to a quantile framework. To make it possible to use the observational data of control units to estimate the distributional characteristic of the unobserved counterfactual outcome variable for the treated units, the invariance conditional CDF assumption is adopted. With this invariance assumption, a simple method is proposed to estimate the QTE for the treated unit in panel data. The main issue of the proposed method is the estimation of the conditional CDF. To estimate the conditional CDF, both nonparametric and parametric models are discussed in the paper. Furthermore, when the number of control units is greater than the time periods in panel data, the LASSO type quantile regression method is suggested to estimate the conditional CDF. In the empirical analysis, the proposed QTE estimator for panel data is applied to estimate the QTE of the introduction of the CSI 300 index futures market on the log-return and the volatility of the stock market in China.

Finally, note that as mentioned earlier, some extensions are warranted as future research.

Appendix: Mathematical Proofs

To prove Theorem 1, we first prove the following two lemmas.

Lemma 1. *Under the assumptions in the theorem, we have*

$$\begin{aligned}\sqrt{T_2}\big[\tilde{F}_{Y_2^0}(y) - F_{Y_2^0}(y)\big] &= \frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T}\big[F(y|X_{2t}) - F_{Y_2^0}(y)\big] \\ &\quad + \frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\sqrt{\frac{T_2}{T_1}}\cdot\frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})}\cdot[I(Y_{1t}\le y) - F(y|X_{1t})] + o_p(1).\end{aligned}$$

for $-\infty < y < \infty$.

Proof of Lemma 1: Note that

$$\begin{aligned}&\sqrt{T_2}\big[\tilde{F}_{Y_2^0}(y) - F_{Y_2^0}(y)\big] \\ =&\ \sqrt{T_2}\left\{\frac{1}{T_2}\sum_{t=T_1+1}^{T}\big[\tilde{F}(y|X_{2t}) - F(y|X_{2t})\big] + \frac{1}{T_2}\sum_{t=T_1+1}^{T}\big[F(y|X_{2t}) - F_{Y_2^0}(y)\big]\right\} \\ =&\ \sqrt{T_2}\left\{\frac{1}{T_2}\sum_{t=T_1+1}^{T}\big[\tilde{F}(y|X_{2t}) - F(y|X_{2t})\big] - \int\big[\tilde{F}(y|x) - F(y|x)\big]\,dF_{X_2}(x)\right\} \\ &+\sqrt{T_2}\Big[\int\tilde{F}(y|x)dF_{X_2}(x) - \int F(y|x)dF_{X_2}(x)\Big] + \frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T}\big[F(y|X_{2t}) - F_{Y_2^0}(y)\big] \\ =&\ S_1 + S_2 + \frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T}\big[F(y|X_{2t}) - F_{Y_2^0}(y)\big].\end{aligned}$$

Following the proof of Lemma 1 in Rothe (2010), it can be shown easily that $S_1 = o_p(1)$. Now, our focus is on S_2, which can be written as

$$\begin{aligned}S_2 &= \sqrt{T_2}\Big[\int\tilde{F}(y|x)dF_{X_2}(x) - \int F(y|x)dF_{X_2}(x)\Big] \\ &= \sqrt{T_2}\Big[\int\frac{\frac{1}{T_1}\sum_{t=1}^{T_1}K_h(X_{1t}-x)I(Y_{1t}\le y)}{\tilde{f}_{X_1}(x)}\,dF_{X_2}(x) - \int F(y|x)dF_{X_2}(x)\Big] \\ &= \sqrt{T_2}\Big[\int\frac{\frac{1}{T_1}\sum_{t=1}^{T_1}K_h(X_{1t}-x)I(Y_{1t}\le y)}{\tilde{f}_{X_1}(x)}\,dF_{X_2}(x)\end{aligned}$$

$$
\begin{aligned}
& \quad - \int \frac{\frac{1}{T_1}\sum_{t=1}^{T_1} K_h(X_{1t}-x)F(y|x)}{\tilde{f}_{X_1}(x)}\, dF_{X_2}(x)\Big] \\
&= \sqrt{T_2}\Big\{ \int \frac{1}{T_1}\sum_{t=1}^{T_1} \frac{K_h(X_{1t}-x)}{\tilde{f}_{X_1}(x)}\big[I(Y_{1t}\le y) - F(y|X_{1t})\big]\, dF_{X_2}(x) \\
& \quad + \int \frac{1}{T_1}\sum_{t=1}^{T_1} \frac{K_h(X_{1t}-x)}{\tilde{f}_{X_1}(x)}\big[F(y|X_{1t}) - F(y|x)\big]\Big\}\, dF_{X_2}(x) \\
&= \sqrt{T_2}(A+B), \qquad \text{(A.1)}
\end{aligned}
$$

where $\tilde{f}_{X_1}(x) = \frac{1}{T_1}\sum_{t=1}^{T_1} K_h(X_{1t}-x)$. By Theorem 2 in Masry (1996),

$$
\sup_{x\in\mathcal{X}_2} |\tilde{f}_{X_1}(x) - f_{X_1}(x)| = O_p\left\{ \left(\frac{\log T_1}{T_1 h^{d_x}}\right)^{1/2} + h^{\ell} \right\}.
$$

Then, under Assumption 7, $\sup_{x\in\mathcal{X}_2} |\tilde{f}_{X_1}(x) - f_{X_1}(x)| = o_p(T_1^{-1/4})$. Thus, by applying a second order Taylor expansion of $1/\tilde{f}_{X_1}(x)$ around $1/f_{X_1}(x)$, one obtains that

$$
\begin{aligned}
A &= \frac{1}{T_1}\sum_{t=1}^{T_1}\big[I(Y_{1t}\le y) - F(y|X_{1t})\big]\int \frac{f_{X_2}(x)}{\tilde{f}_{X_1}(x)} K_h(X_{1t}-x)\, dx \\
&= \frac{1}{T_1}\sum_{t=1}^{T_1}\big[I(Y_{1t}\le y) - F(y|X_{1t})\big]\int \frac{f_{X_2}(x)}{f_{X_1}(x)} K_h(X_{1t}-x)\, dx \\
& \quad - \frac{1}{T_1}\sum_{t=1}^{T_1}\big[I(Y_{1t}\le y) - F(y|X_{1t})\big]\int \frac{f_{X_2}(x)}{f^2_{X_1}(x)}\big[\tilde{f}_{X_1}(x) - f_{X_1}(x)\big] K_h(X_{1t}-x)\, dx + o_p(T_1^{-1/2}) \\
&= A_1 - A_2 + o_p(T_1^{-1/2}).
\end{aligned}
$$

Since $K(\cdot)$ is a high order kernel,

$$
\begin{aligned}
A_1 &= \frac{1}{T_1}\sum_{t=1}^{T_1}\big[I(Y_{1t}\le y) - F(y|X_{1t})\big]\frac{f_{X_2}(X_t)}{f_{X_1}(X_t)} + O_p(h^{\ell}) \\
&= \frac{1}{T_1}\sum_{t=1}^{T_1}\big[I(Y_{1t}\le y) - F(y|X_{1t})\big]\frac{f_{X_2}(X_t)}{f_{X_1}(X_t)} + o_p(T^{-1/2})
\end{aligned}
$$

and

$$
\begin{aligned}
A_2 &= \frac{1}{T_1}\sum_{t=1}^{T_1}\big[I(Y_{1t}\le y) - F(y|X_{1t})\big]\int \frac{f_{X_2}(x)}{f^2_{X_1}(x)}\big[\tilde{f}_{X_1}(x) - f_{X_1}(x)\big] K_h(X_{1t}-x)\, dx \\
&= \frac{1}{T_1}\sum_{t=1}^{T_1}\big[I(Y_{1t}\le y) - F(y|X_{1t})\big]\int \left[\frac{1}{T_1}\sum_{s=1}^{T_1} K_h(X_{1s}-x) - f_{X_1}(x)\right] K_h(X_{1t}-x)\, \frac{f_{X_2}(x)}{f^2_{X_1}(x)}\, dx
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{T_1^2}\sum_{s,t=1}^{T_1}\big[I(Y_{1t}\le y)-F(y|X_{1t})\big]\bigg[\int K_h(X_{1s}-x)K_h(X_{1t}-x)\frac{f_{X_2}(x)}{f_{X_1}^2(x)}\,dx \\
&\qquad\qquad -\int K_h(X_{1t}-x)\frac{f_{X_2}(x)}{f_{X_1}(x)}\,dx\bigg] \\
&= \frac{1}{T_1^2}\sum_{s,t=1}^{T_1}\big[I(Y_{1t}\le y)-F(y|X_{1t})\big]\bigg[K_h(X_{1s}-X_{1t})\frac{f_{X_2}(X_{1t})}{f_{X_1}^2(X_{1t})}-\frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})}\bigg]+o_p(T_1^{-1/2}) \\
&= \frac{1}{T_1^2}\sum_{s\neq t}\big[I(Y_{1t}\le y)-F(y|X_{1t})\big]\bigg[K_h(X_{1s}-X_{1t})\frac{f_{X_2}(X_{1t})}{f_{X_1}^2(X_{1t})}-\frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})}\bigg]+o_p(T_1^{-1/2}).
\end{aligned}
$$

Note that when $T_1\to\infty$ and $|s-t|\to\infty$,

$$
\begin{aligned}
& E\bigg\{\big[I(Y_{1t}\le y)-F(y|X_{1t})\big]\bigg[K_h(X_{1s}-X_{1t})\frac{f_{X_2}(X_{1t})}{f_{X_1}^2(X_{1t})}-\frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})}\bigg]\bigg|\,X_{1t},Y_{1t}\bigg\} \\
\to\ & \big[I(Y_{1t}\le y)-F(y|X_{1t})\big]\bigg[\frac{f_{X_2}(X_{1t})}{f_{X_1}^2(X_{1t})}\cdot\int K_h(x-X_{1t})f_{X_1}(x)\,dx-\frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})}\bigg]=o_p(T_1^{-1/2}).
\end{aligned}
$$

So the leading term in A_2 is a degenerate second-order U-statistic and

$$A_2=O_p(T_1^{-1}h^{-d_x})=o_p(T_1^{-1/2}).$$

Therefore,

$$A=A_1-A_2+o_p(T_1^{-1/2})=\frac{1}{T_1}\sum_{t=1}^{T_1}\big[I(Y_{1t}\le y)-F(y|X_{1t})\big]\frac{f_{X_2}(X_t)}{f_{X_1}(X_t)}+o_p(T^{-1/2}). \quad (A.2)$$

Similarly, we can show that the term B in (A.1) is $o_p(T^{-1/2})$, together with (A.1), (A.2) and $\lim_{T\to\infty}T_2/T_1=c$, one has that

$$
\begin{aligned}
S_2 &= \sqrt{T_2}\big[A_1-A_2+B_1-B_2+o_p(T_1^{-1/2})\big] \\
&= \frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\sqrt{\frac{T_2}{T_1}}\frac{f_{X_2}(X_t)}{f_{X_1}(X_t)}\cdot\big[I(Y_{1t}\le y)-F(y|X_{1t})\big]+o_p(1).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\sqrt{T_2}\big[\tilde F_{Y_2^0}(y)-F_{Y_2^0}(y)\big] &= S_1+S_2+\frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T}\big[F(y|X_{2t})-F_{Y_2^0}(y)\big] \\
&= \frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\sqrt{\frac{T_2}{T_1}}\frac{f_{X_2}(X_t)}{f_{X_1}(X_t)}\cdot\big[I(Y_{1t}\le y)-F(y|X_{1t})\big] \\
&\quad +\frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T}\big[F(y|X_{2t})-F_{Y_2^0}(y)\big]+o_p(1).
\end{aligned}
$$

This completes the proof of Lemma 1. $\square$

Lemma 2. *Under the assumptions in the theorem, we have*

$$\hat{F}_{Y_2^0}(y) - \tilde{F}_{Y_2^0}(y) = o_p\big(T_2^{-1/2}\big)$$

and

$$\begin{aligned}\sqrt{T_2}\big[\hat{F}_{Y_2^0}(y) - F_{Y_2^0}(y)\big] \;&=\; \frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T}\big[F(y|X_{2t}) - F_{Y_2^0}(y)\big] \\ &\quad + \frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\sqrt{\frac{T_2}{T_1}}\cdot\frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})}\cdot[I(Y_{1t}\le y) - F(y|X_{1t})] + o_p(1)\end{aligned}$$

for $-\infty < y < \infty$.

Proof of Lemma 2: We first show that

$$\sup_{u\le y}\{\tilde{F}_{Y_2^0}(u)\} - \tilde{F}_{Y_2^0}(y) = o_p\big(T_2^{-1/2}\big).$$

By Lemma 1, for $r > 0$,

$$\begin{aligned}&\sqrt{T_2}\big[\bar{F}_{Y_2^0}(y-r) - \tilde{F}_{Y_2^0}(y-r)\big] - \sqrt{T_2}\big[\bar{F}_{Y_2^0}(y) - \tilde{F}_{Y_2^0}(y)\big] \\ =\;& -\frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T}\big[P\big(y-r < Y_{2t}^0 \le y|X_{2t}\big) - P\big(y-r < Y_{2t}^0 \le y\big)\big] \\ & -\frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\sqrt{\frac{T_2}{T_1}}\cdot\frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})}\cdot\big[I\big(y-r<Y_{1t}\le y\big) - P\big(y-r<Y_{1t}\le y|X_{1t}\big)\big] + o_p(1) \\ =\;& G(r) + o_p(1). \end{aligned} \tag{A.3}$$

It is easy to find that $E\big[G(r)\big] = 0$ and $\lim_{r\to 0^+}\operatorname{Var}\big[G(r)\big] = 0$. By the Markov inequality, for any $\varepsilon > 0$ and $\epsilon > 0$, there exists a small $\delta > 0$ such that $P\big(\sup_{0\le r<\delta}\{G(r)\} > \varepsilon\big) < \epsilon$. Then, (A.3) implies that for any $\varepsilon > 0$ and $\epsilon > 0$, there exists a small $\delta > 0$ and an $N_1 > 0$ such that for all $T_2 > N_1$

$$P\Big(\sqrt{T_2}\Big\{\sup_{y-\delta<u\le y}\big\{\tilde{F}_{Y_2^0}(u) - F_{Y_2^0}(u)\big\} - \big[\tilde{F}_{Y_2^0}(y) - F_{Y_2^0}(y)\big]\Big\} > \varepsilon\Big) < \epsilon. \tag{A.4}$$

Note that for any $y-\delta < u \le y$, $F_{Y_2^0}(u) \le F_{Y_2^0}(y)$. Hence, (A.4) leads to

$$P\Big(\sqrt{T_2}\Big\{\sup_{y-\delta<u\le y}\big\{\tilde{F}_{Y_2^0}(u)\big\} - \tilde{F}_{Y_2^0}(y)\Big\} > \varepsilon\Big) < \epsilon. \tag{A.5}$$

Also, we have that

$$
\begin{aligned}
& P\Big(\sqrt{T_2}\Big[\sup_{u\le y-\delta}\{\tilde F_{Y_2^0}(u)\}-\tilde F_{Y_2^0}(y)\Big]>\varepsilon\Big)\\
\le\ & P\Big(\sup_{u\le y-\delta}\{\tilde F_{Y_2^0}(u)\}-\tilde F_{Y_2^0}(y)>0\Big)\\
=\ & P\Big(\sup_{u\le y-\delta}\{\tilde F_{Y_2^0}(u)-F_{Y_2^0}(u)+F_{Y_2^0}(u)\}-\big[\tilde F_{Y_2^0}(y)-F_{Y_2^0}(y)\big]-F_{Y_2^0}(y)>0\Big)\\
\le\ & P\Big(\sup_{u\le y-\delta}\{\tilde F_{Y_2^0}(u)-F_{Y_2^0}(u)\}+F_{Y_2^0}(y-\delta)-\big[\tilde F_{Y_2^0}(y)-F_{Y_2^0}(y)\big]-F_{Y_2^0}(y)>0\Big)\\
=\ & P\Big(\sup_{u\le y-\delta}\{\tilde F_{Y_2^0}(u)-F_{Y_2^0}(u)\}-\big[\tilde F_{Y_2^0}(y)-F_{Y_2^0}(y)\big]>F_{Y_2^0}(y)-F_{Y_2^0}(y-\delta)\Big).
\end{aligned}
$$

Note that $\sup_{u\le y-\delta}\{\tilde F_{Y_2^0}(u)-F_{Y_2^0}(u)\}-\big[\tilde F_{Y_2^0}(y)-F_{Y_2^0}(y)\big]\xrightarrow{p}0$ by Lemma 1 and $F_{Y_2^0}(y)-F_{Y_2^0}(y-\delta)>0$. Then, $P\Big(\sqrt{T_2}\Big[\sup_{u\le y-\delta}\{\tilde F_{Y_2^0}(u)\}-\tilde F_{Y_2^0}(y)\Big]>\varepsilon\Big)\to 0$, which implies that there exist an N_2 such that for all $T_2>N_2$,

$$
P\Big(\sqrt{T_2}\Big\{\sup_{u\le y-\delta}\{\tilde F_{Y_2^0}(u)\}-\tilde F_{Y_2^0}(y)\Big\}>\varepsilon\Big)<\epsilon. \tag{A.6}
$$

Combing (A.5) and (A.6) leads to

$$
\begin{aligned}
& P\Big(\sqrt{T_2}\Big\{\sup_{u\le y}\{\tilde F_{Y_2^0}(u)\}-\tilde F_{Y_2^0}(y)\Big\}>\varepsilon\Big)\\
\le\ & P\Big(\sqrt{T_2}\Big\{\sup_{y-\delta<u\le y}\{\tilde F_{Y_2^0}(u)\}-\tilde F_{Y_2^0}(y)\Big\}>\varepsilon\Big)+P\Big(\sqrt{T_2}\Big\{\sup_{u\le y-\delta}\{\tilde F_{Y_2^0}(u)\}-\tilde F_{Y_2^0}(y)\Big\}>\varepsilon\Big)\\
\le\ & 2\epsilon
\end{aligned}
$$

for $T_2>\max\{N_1,N_2\}$, which implies that

$$
\sup_{u\le y}\{\tilde F_{Y_2^0}(u)\}-\tilde F_{Y_2^0}(y)=o_p\big(T_2^{-1/2}\big). \tag{A.7}
$$

Next, it is easy to show that

$$
\sup_{u}\{\tilde F_{Y_2^0}(u)\}=\lim_{y\to\infty}\sup_{u\le y}\{\tilde F_{Y_2^0}(u)\}=\lim_{y\to}\tilde F_{Y_2^0}(y)+o_p\big(T_2^{-1/2}\big)=1+o_p\big(T_2^{-1/2}\big). \tag{A.8}
$$

Combing (A.7) and (A.8) leads to

$$
\begin{aligned}
\hat F_{Y_2^0}(y)-\tilde F_{Y_2^0}(y) &= \frac{\sup_{u\le y}\{\tilde F_{Y_2^0}(u)\}}{\sup_u\{\tilde F_{Y_2^0}(u)\}}-\tilde F_{Y_2^0}(y)\\
&= \sup_{u\le y}\{\tilde F_{Y_2^0}(u)\}\big[1+o_p\big(T_2^{-1/2}\big)\big]-\tilde F_{Y_2^0}(y)=o_p\big(T_2^{-1/2}\big).
\end{aligned}
$$

Then, by Lemma 1, we have

$$
\sqrt{T_2}\big[\hat F_{Y_2^0}(y)-F_{Y_2^0}(y)\big] = \sqrt{T_2}\big[\hat F_{Y_2^0}(y)-\tilde F_{Y_2^0}(y)\big]+\sqrt{T_2}\big[\tilde F_{Y_2^0}(y)-F_{Y_2^0}(y)\big]
$$

$$
\begin{aligned}
&= \sqrt{T_2}\big[\tilde{F}_{Y_2^0}(y) - F_{Y_2^0}(y)\big] + o_p(1) \\
&= \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^{T} \big[F(y|X_{2t}) - F_{Y_2^0}(y)\big] \\
&\quad + \frac{1}{\sqrt{T_1}} \sum_{t=1}^{T_1} \sqrt{\frac{T_2}{T_1}} \cdot \frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})} \cdot [I(Y_{1t} \le y) - F(y|X_{1t})] + o_p(1).
\end{aligned}
$$

This completes the proof of Lemma 2. □

Proof of Theorem 1: First, by Theorem 2 of Yoshihara (1995) and Assumption 2, the following the Bahadur representation for $\hat{q}_{1\tau}^1$ is given by

$$
\sqrt{T_2}(\hat{q}_{1\tau}^1 - q_{1\tau}^1) = \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^{T} \psi_{1\tau}(Y_{2t}^1) + o_p(1), \tag{A.9}
$$

To prove the theorem, it remains to find the representation for $\hat{q}_{1\tau}^0$. Define

$$
Z_{T_2} = \sqrt{T_2}[\tau - \hat{F}_{Y_2^0}(q_{1\tau}^0)]/f_{Y_2^0}(q_{1\tau}^0).
$$

Note that $\hat{F}_{Y_2^0}(q)$ is a CDF. For any $-\infty < u < \infty$, $\hat{q}_{1\tau}^0 \le q_{1\tau}^0 + u/\sqrt{T_2}$ implies $\hat{F}_{Y_2^0}\big(q_{1\tau}^0 + \frac{u}{\sqrt{T_2}}\big) \ge \tau$. Then, for any $\epsilon > 0$,

$$
\begin{aligned}
& P\big(\sqrt{T_2}(\hat{q}_{1\tau}^0 - q_{1\tau}^0) \le u,\, Z_{T_2} > u + \epsilon\big) \\
&= P\big(\hat{q}_{1\tau}^0 \le q_{1\tau}^0 + u/\sqrt{T_2},\, Z_{T_2} > u + \epsilon\big) \\
&\le P\left(\tau - \hat{F}_{Y_2^0}\big(q_{1\tau}^0 + u/\sqrt{T_2}\big) \le 0,\, Z_{T_2} > u + \epsilon\right) \\
&= P\left(\frac{\sqrt{T_2}[\tau - \hat{F}_{Y_2^0}(q_{1\tau}^0 + \frac{u}{\sqrt{T_2}})]}{f_{Y_2^0}(q_{1\tau}^0)} \le 0,\, Z_{T_2} > u + \epsilon\right) \\
&= P\left(Z_{T_2} + R_{T_2} \le u,\, Z_{T_2} > u + \epsilon\right),
\end{aligned} \tag{A.10}
$$

where

$$
\begin{aligned}
R_{T_2} &= \left[\frac{\hat{F}_{Y_2^0}(q_{1\tau}^0) - \hat{F}_{Y_2^0}\big(q_{1\tau}^0 + \frac{u}{\sqrt{T_2}}\big)}{f_{Y_2^0}(q_{1\tau}^0)/\sqrt{T_2}} - \frac{F_{Y_2^0}(q_{1\tau}^0) - F_{Y_2^0}\big(q_{1\tau}^0 + \frac{u}{\sqrt{T_2}}\big)}{f_{Y_2^0}(q_{1\tau}^0)/\sqrt{T_2}}\right] \\
&\quad + \left[\frac{F_{Y_2^0}(q_{1\tau}^0) - F_{Y_2^0}\big(q_{1\tau}^0 + \frac{u}{\sqrt{T_2}}\big)}{f_{Y_2^0}(q_{1\tau}^0)/\sqrt{T_2}} + u\right] \\
&= R_{T_2,1} + R_{T_2,2}.
\end{aligned}
$$

By Lemma 2, we have that

$$
R_{T_2,1} = -\frac{1}{f_{Y_2^0}(q_{1\tau}^0)} \Big\{ \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1}^{T} \big[P\big(q_{1\tau}^0 < Y_{2t}^0 \le q_{1\tau}^0 + u/\sqrt{T_2}|X_{2t}\big) - P\big(q_{1\tau}^0 < Y_{2t}^0 \le q_{1\tau}^0 + u/\sqrt{T_2}\big)\big]
$$

$$+\frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\sqrt{\frac{T_2}{T_1}}\cdot\frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})}\cdot\big[I\big(q_{1\tau}^0<Y_{1t}\le q_{1\tau}^0+u/\sqrt{T_2}\big)-P\big(q_{1\tau}^0<Y_{1t}\le q_{1\tau}^0+u/\sqrt{T_2}|X_{1t}\big)\big]\Big\}$$
$$+o_p(1).$$

It is easy to find that the leading term in $R_{T_2,1}$ has a zero mean and its variance goes to zero as T_2 goes to infinity, so $R_{T_2,1}=o_p(1)$. We also have

$$R_{T_2,2}\to\frac{-uf_{Y_2^0}(q_{1\tau}^0)}{f_{Y_2^0}(q_{1\tau}^0)}+u=0,$$

so that $R_{T_2}=o_p(1)$. Then, by (A.10),

$$P\big(\sqrt{T_2}(\hat{q}_{1\tau}^0-q_{1\tau}^0)\le u,\,Z_{T_2}>u+\epsilon\big)\le P\left(Z_{T_2}+R_{T_2}\le u,\,Z_{T_2}>u+\epsilon\right)\to 0. \quad \text{(A.11)}$$

Similarly, we can prove that for any $-\infty<u<\infty$ and $\epsilon>0$,

$$P\big(\sqrt{T_2}(\hat{q}_{1\tau}^0-q_{1\tau}^0)\ge u,\,Z_{T_2}<u-\epsilon\big)\to 0. \quad \text{(A.12)}$$

Combining (A.11) and (A.12) leads to $\sqrt{T_2}(\hat{q}_{1\tau}^0-q_{1\tau}^0)=Z_{T_2}+o_p(1)$. Then, by Lemma 2,

$$\begin{aligned}\sqrt{T_2}(\hat{q}_{1\tau}^0-q_{1\tau}^0) &= -\sqrt{T_2}[\hat{F}_{Y_2^0}(q_{1\tau}^0)-\tau]/f_{Y_2^0}(q_{1\tau}^0)+o_p(1)\\ &= \frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T}\frac{-[F(q_{1\tau}^0|X_{2t})-\tau]}{f_{Y_2^0}(q_{1\tau}^0)}\\ &\quad -\frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\frac{1}{f_{Y_2^0}(q_{1\tau}^0)}\cdot\sqrt{\frac{T_2}{T_1}}\cdot\frac{f_{X_2}(X_{1t})}{f_{X_1}(X_{1t})}\cdot[I(Y_{1t}\le q_{1\tau}^0)-F(q_{1\tau}^0|X_{1t})]+o_p(1)\\ &= \frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T}\psi_{0\tau}(X_{2t})-\frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\psi_{2\tau}(X_{1t},Y_{1t})+o_p(1). \qquad \text{(A.13)}\end{aligned}$$

By (A.9) and (A.13), we have

$$\begin{aligned}\sqrt{T_2}(\hat{\Delta}_\tau-\Delta_\tau) &= \frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T}[\psi_{1\tau}(Y_{2t})-\psi_{0\tau}(X_{2t})]+\frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\psi_{2\tau}(X_{1t},Y_{1t})+o_p(1)\\ &= \frac{1}{\sqrt{T_2}}\sum_{t=T_1+1}^{T}\xi_t+\frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\eta_t+o_p(1)=K_1+K_2+o_p(1),\end{aligned}$$

where K_1 and K_2 are clearly defined. By the stationarity, we define $C(s)=\mathrm{Cov}\big(\eta_t,\xi_{t+s}\big)$ for $s>0$. A simple algebra gives

$$\mathrm{Cov}(K_1,K_2)=\sqrt{\frac{T_1}{T_2}}\frac{1}{T_1}\sum_{s=1}^{T_1}\min\{s,T_2\}\,C(s)+\sqrt{\frac{T_2}{T_1}}\frac{1}{T_2}\sum_{s=T_1+1}^{T-1}\min\{T_1,T-s\}\,C(s).$$

By Kronecker's lemma, Davydov's inequality, and Assumption 2, the first term in the right hand side of the above equation converges to zero and the second term is bounded by $\sum_{s=T_1+1}^{\infty} |C(s)| \to 0$ as $T_1 \to \infty$ by Assumption 2. Therefore, by the central limit theorem for α-mixing as in Theorem 2 of Yoshihara (1995), one has

$$\sqrt{T_2}(\hat{\Delta}_\tau - \Delta_\tau) \xrightarrow{d} N\left(0, \sigma_\tau^2\right).$$

This completes the proof of Theorem 1. □

References

Aït-Shahalia, Y. and M.W. Brant (2001). Variable selection for portfolio choice. *Journal of Finance*, **56**(4), 1297-1351.

Akaike H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**(6), 716-723.

Bai, C., Q. Li and M. Ouyang (2014). Property taxes and home prices: A tale of two cities. *Journal of Econometrics*, **180**(1), 1-15.

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

Bove, V., L. Elia and R.P Smith (2017). On the heterogeneous consequences of civil war. *Oxford Economic Papers*, **69**(3), 550-568.

Cai, Z. (2021). Recent developments in estimating treatment effects for panel data. *China Journal of Econometrics*, **1**(2), 232-249.

Callaway, B., T. Li and T. Oka (2018). Quantile treatment effects in difference in differences models under dependence restrictions and with only two time periods. *Journal of Econometrics*, **206**(2), 395-413.

Callaway, B. and T. Li (2019). Quantile treatment effects in difference in differences models with panel data. *Quantitative Economics*, **10**(4), 1579-1618.

Carvalho, C., R. Masini and M.C. Medeiros (2018). ArCo: An artificial counterfactual approach for high-dimensional panel time-series data. *Journal of Econometrics*, **207**(2), 352-380.

Chen, H., Q. Han, Y. Li, and K. Wu (2013). Does index futures trading reduce volatility in the Chinese stock market? A panel data evaluation approach. *Journal of Futures Markets*, **33**(12), 1167-1190.

Chen, T.F., T. Chordia, S.L. Chung and J. Lin (2021). Volatility-of-volatility risk in asset pricing. *Review of Asset Pricing Studies* with DOI https://doi.org/10.1093/rapstu/raab018.

Chernozhukov, V., I. I. Fernández-Val and A. Galichon (2010). Quantile and probability curves without crossing. *Econometrica*, **78**(3), 1093-1125.

Chernozhukov, V., I. Fernández-Val and B. Melly (2013). Inference on counterfactual distributions. *Econometrica*, **81**(6), 2205-2268.

Du, Z. and L. Zhang (2015). Home-purchase restriction, property tax and housing price in China: A counterfactual analysis. *Journal of Econometrics*, **188**(2), 558-568.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, **75**(1), 259-276.

Firpo, S., N.M. Fortin and T. Lemieux (2009). Unconditional quantile regressions. *Econometrica*, **77**(3), 953-973.

Fujiki, H. and C. Hsiao (2015). Disentangling the effects of multiple treatments – Measuring the net economic impact of the 1995 great Hanshin-Awaji earthquake. *Journal of Econometrics*, **186**(1), 66-73.

Gasser, T., H. Müller and V. Mammitzsch (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society, Series B*, **47**(2), 238-252.

Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and its Applications*. Academic Press, New York

Hall, P. and Q. Yao (2005). Approximating conditional distribution functions using dimension reduction. *Annals of Statistics*, **33**(3), 1404-1421.

Hollstein, F. and M. Prokopczuk (2018). How aggregate volatility-of-volatility affects stock returns. *Review of Asset Pricing Studies*, **8**(2), 253-292.

Hsiao, C., S.H. Ching and K.S. Wan (2012). A panel data approach for program evaluation: Measuring the benefits of political and economic integration of Hong kong with mainland China.*Journal of Applied Econometrics*, **27**(5), 705-740.

Hsu, Y.C., T.C. Lai and R.P. Lieli (2022). Counterfactual treatment effects: Estimation and inference. *Journal of Business & Economic Statistics*, **40**(1), 240-255.

Huang, D., C. Schlag, I. Shaliastovich, and J. Thimme (2019). Volatility-of-volatility risk. *Journal of Financial and Quantitative Analysis*, **54**(6), 2432-2452.

Hurvich, C.M. and C.-L. Tsai (1989). Regression and time series model selection in small samples. *Biometrika*, **76**(2), 297-307.

Ke, X., H. Chen, Y. Hong and C. Hsiao (2017). Do China's high-speed-rail projects promote local economy? – New evidence from a panel data approach. *China Economic Review*, **44**(1), 203-226.

Ke, X. and C. Hsiao (2021). Economic impact of the most drastic lockdown during COVID-19 pandemic — the experience of Hubei, China. *Journal of Applied Econometrics* with DOI: https://doi.org/10.1002/jae.2871.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica*, **46**(1), 33-50.

Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, **17**(3), 1217-1241.

Li, K.T. and D.R. Bell (2017). Estimation of average treatment effects with panel data: Asymptotic theory and implementation. *Journal of Econometrics*, **197**(1), 65-75.

Li, Q. and W. Long (2018). Do parole abolition and Truth-in-Sentencing deter violent crimes in Virginia? *Empirical Economics*, **55**(4), 2027-2045.

Li, Y. and J. Zhu (2008). L_1-norm quantile regression. *Journal of Computational and Graphical Statistics*, **17**(1), 163-185.

Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis*, **17**(6), 571-599.

Newey, W.K. and K.D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**(3), 703-708.

Ouyang, M. and Y. Peng (2015). The treatment-effect estimation: A case study of the 2008 economic stimulus package of China. *Journal of Econometrics*, **188**(2), 545-557.

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

Rothe, C. (2010). Nonparametric estimation of distributional policy effects. *Journal of Econometrics*, **155**(1), 56-70.

Sherwood, B. and A. Maidman (2016). rqPen: Penalized quantile Regression. R package version 1.4, 2016.

Wang, L., Y. Wu and R. Li (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, **107**(497), 214-222.

Wu, Y. and Y. Liu (2009). Variable selection in quantile regression. *Statistica Sinica*, **19**(2), 801-817.

Xiao, Z. and R. Koenker (2009). Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of American Statistical Association*, **104**(488), 1696-1712.

Yi, C. and J. Huang (2017). Semismooth Newton coordinate descent algorithm for elastic-net penalized Huber loss regression and quantile regression. *Journal of Computational and Graphical Statistics*, **26**(3), 547-557.

Yoshihara, K. (1995). The Bahadur representation of sample quantiles for sequences of strongly mixing random variables. *Statistics & Probability Letters*, **24**(4), 299-304.