Testing Financial Hierarchy Based on A PDQ-CRE Model*†

Zongwu Cai[a], Meng Shi[b], Wuqing Wu[c,†], Yue Zhao[c]

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

[b]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, Beijing 100190, and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, Beijing 100049, China

[c]School of Business, Renmin University of China, Beijing, Beijing 100872, China

---

Abstract

This paper investigates the relative importance of internal and external sources of funds in financing activities across different levels of investment activities by proposing a *panel data quantile* regression model with *correlated random effects* (PDQ-CRE), which accounts for heteroscedasticity in both firm-specific individuals and distribution of investment. A new estimation method is proposed by using the quasi-likelihood function for conditional quantile model and Laplace approximation. We show that the proposed estimator is consistent and normally distributed. A Monte Carlo simulation is conducted to examine the performance of the estimator in finite samples. Finally, empirical results find a strong evidence that the financing hierarchy of U.S. firms is in accordance with the first rung of the pecking order theory across all levels of investments from 10% to 90%.

Keywords: Correlated Random Effects; Panel Data; Pecking Order Theory; Quantile Regression Model; Quasi-Maximum Likelihood Estimator

---

*The authors gratefully acknowledge the financial support, in part, by the National Natural Science Foundation of China grants #71631004 and #71871216, and the Social Science Foundation of Beijing grant #17GLB022 as well as the Research Funds of Renmin University of China #6XNB025.

†*Corresponding author: W. Wu (wuwuqing@rmbs.ruc.edu.cn).*

1 Introduction

The pecking order theory in finance is a widely held belief that firms prefer internal (retained earnings) to external funds, and prefer debt to equity if internal funds are exhausted and external funds are sought (Myers, 1984; Myers and Majluf, 1984). For more than thirty years, a substantial empirical literature has examined whether and when there is a departure of observed corporate financing behaviors from the pecking order theory in practice (Brennan and Kraus, 1987; Jung et al., 1996; Shyam-Sunder and Myers, 1999; Frank and Goyal, 2003; Fama and French, 2005; Bharath et al., 2009; Brown et al., 2009; Chay et al., 2015; Brown et al., 2019). Unfortunately, no consensus has been reached.

Indeed, most of these studies employ a mean model, of which the results and conclusions are not warranted for two main reasons. On the one hand, the traditional mean regression (e.g., OLS) assumes an unchanging association at different levels of investment scales. However, this paper predicts that the relationship between financing sources and investments varies across different investment scales. On the other hand, the estimation of linear model or fixed effect model is unbiased only when the dependent variable is normally distributed or with a central tendency. In contrast, the distribution of investment amount is usually highly skewed and heavily tailed, see Figure 1 later for investment amount in year 2016, which shows clearly that the distribution of investment amount is highly right-tailed skewed. Thus, the mean regression estimates are likely to be biased and probably not truly reflect the behavior of data in tail regions. To the best of our knowledge, Chay et al. (2015) is among the first and the latest study to investigate the responsiveness of investments to internal funds, debt issues and equity issues taking a quantile regression based on panel data. However, limited by the difficulties in the theoretical modeling and computational challenges to control firm and year fixed effects, their findings may suer from endogeneity problems while this paper takes a further step to introduce a linear quantile model for panel data with fixed effects.

Recently, there is growing literature to study panel quantile models with individual effects. The main reason is that the commonly used method to eliminate unobservable individual effects in the conditional mean model, taking difference, is not available in the conditional quantile regression model. Among this literature, Koenker (2004) first introduced a location-shift quantile regression model for panel data with fixed effects which assumes that the fixed effect is purely a location shift and time period T goes to infinity. The fixed effects are viewed as nuisance parameters and regularized by L_1-penalty to shrinkage them to common values. Furthermore, Lamarche (2010) discussed the choice of the tuning parameter λ in Koenker (2004), and proposed an optimal selection method. Galvao et al. (2013) made an extension to a class of censored quantile regression models for panel data with fixed effects by applying fixed effects quantile regression to subsets of observations selected either parametrically or nonparametrically. For more on the most recent work, the reader is referred to the papers by Powell (2016), Gu and Volgushev (2019), Machado and Santos Silva (2019), and the references therein. However, the aforementioned studies rely on the critical assumption that the time period T is large or goes to infinity.

Alternative method of estimating individual effects in the panel quantile model is to view individual effects as correlated random effects when T is small or finite. In the framework of Chamberlain (1982, 1984), Abrevaya and Dahl (2008) employed a correlated random-effects quantile model to estimate the effects of mother smoking on the entire birthweight distribution. The correlated random effects are assumed to be a linear projection on observable independent variables plus a disturbance, while Harding and Lamarche (2017) relaxed the assumption in Abrevaya and Dahl (2008), by allowing the unobserved α_i to be arbitrarily related to observable variables, to increase the flexibility in the model specification.

Our contribution is that motivated by controlling the firm-specific heterogeneity in testing the pecking order theory under the quantile framework, we propose a panel data quantile regression model with individual effects, which is the usual model specification in conditional mean model for panel data. We view the individual effects as correlated random effects in the framework of Chamberlain (1982) and Abrevaya and Dahl (2008). Therefore, we do not require the time period T going to infinity, as is the main assumption in many models (Koenker, 2004; Lamarche, 2010; Galvao et al., 2013; Powell, 2016; Gu and Volgushev, 2019; Machado and Santos Silva, 2019). Actually, $T \geq 2$ is only needed for model identification. Also, our model uses a more general objective function, a quasi-likelihood function for conditional quantile models proposed by Komunjer (2005), compared with Abrevaya and Dahl (2008). Finally, our empirical study finds a strong evidence that the financing hierarchy of U.S. firms is in accordance with the first rung of the pecking order theory across all levels of investments from 10% to 90%.

The remainder of this paper is organized as follows. Section 2 introduces the proposed quantile regression model for panel data with correlated random effects. The estimation method and asymptotic properties of our estimator are also provided. Section 3 examines the performance of the proposed estimator with finite samples via Monte Carlo simulations and the simulation results are consistent with the asymptotic theory in Section 2. In Section 4, we apply the model to testing the pecking order theory using non-financial and non-utility U.S. firms from *Compustat* database. Section 5 offers conclusions and suggests possible extensions for future research. All technical proofs are relegated to the Appendix together with the list of assumptions needed for the asymptotic properties for the proposed estimator.

2 Econometric Modeling

2.1 Model Setup

We first introduce the econometric model studied in this paper. Given $\tau \in (0,1)$, we consider the following *panel data quantile* regression model with *correlated random effects* (PDQ-CRE). Let y_{it}, a scalar dependent variable, be the observation on ith individual at time t for $1 \leq i \leq N$ and $1 \leq t \leq T$. The conditional quantile model is given by

$$q_\tau(y_{it}|\boldsymbol{x}_{it}, \alpha_{i,\tau}) = \boldsymbol{x}_{it}'\boldsymbol{\delta}_\tau + \alpha_{i,\tau}, \tag{1}$$

where $q_\tau(y_{it}|\boldsymbol{x}_{it}, \alpha_{i,\tau})$ is the τth quantile of y_{it} given $\boldsymbol{x}_{it}$ and $\alpha_{i,\tau}$, $\boldsymbol{x}_{it}$ is regressors with $K \times 1$ dimensions, $\boldsymbol{\delta}_\tau$ denotes a $K \times 1$ vector of constant coefficients, and $\alpha_{i,\tau}$ is an individual effect dependent on τ. In this model, we assume both coefficients $\boldsymbol{\delta}_\tau$ and individual effect $\alpha_{i,\tau}$ can be dependent on τ. When T goes to infinity, one can use estimation method studied by Koenker (2004). But in this paper, we consider the case that T is a fixed finite number or small. Thus, we follow the idea in Chamberlain (1982, 1984), similar to Abrevaya and Dahl (2008) and Harding and Lamarche (2017), and view the individual effect $\alpha_{i,\tau}$ as a correlated random effect which is allowed to be correlated with covariates $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}', \cdots, \boldsymbol{x}_{iT}')$; that is,

$$\alpha_{i,\tau} = \alpha_\tau(\boldsymbol{x}_i) + v_{i,\tau},$$

where $\alpha_\tau(\cdot)$ is an unknown function of $\boldsymbol{x}_i$, and $v_{i,\tau}$ is an unobserved random error. The unknown function $\alpha_\tau(\boldsymbol{x}_i)$ is approximated by a linear projection on $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT})'$ such that

$$\alpha_\tau(\boldsymbol{x}_i) = \sum_{s=1}^{T} \boldsymbol{x}'_{is}\boldsymbol{\eta}_{s,\tau} + \eta_{0,\tau},$$

where $\eta_{0,\tau}$ is an unknown constant and $\{\boldsymbol{\eta}_{t,\tau}\}_{t=1}^{T}$ is a sequence of $K \times 1$ vector of unknown constant coefficients or nuisance parameters. Therefore, model (1) can be rewritten as

$$q_\tau(y_{it}|\boldsymbol{x}_i, v_{i,\tau}) = \boldsymbol{x}'_{it}\boldsymbol{\delta}_\tau + \sum_{s=1}^{T} \boldsymbol{x}'_{is}\boldsymbol{\eta}_{s,\tau} + \eta_{0,\tau} + v_{i,\tau}. \tag{2}$$

Although the linear approximation may seem to be restrictive, it is the usual specification employed by empirical researchers in cross sectional applications and can be seen as reduced-form approximation to the true conditional quantile, as discussed in Abrevaya and Dahl (2008). Furthermore, similar to Cai et al. (2018), a semiparametric form of $\alpha(\boldsymbol{x}_i)$ might be considered. Also, it is interesting to notice that model (2) happens to be a special case of the semiparametric panel data model proposed in Cai et al. (2018) if the nonparametric part of the model is not considered. Finally, model (2) covers the following conditional quantile regression model with error in the dependent variable (EIDV) for cross-sectional data as a special case when T is set by 1,

$$q_\tau(y_i|\boldsymbol{x}_i, v_i) = q_\tau(\boldsymbol{x}_i) + v_i,$$

where $q_\tau(\boldsymbol{x}_i)$ is the τth quantile of true y_i^* with $y_i = y_i^* + v_i$ and v_i is the measurement error. Indeed, Hausman et al. (2019) considered the case that $q_\tau(\boldsymbol{x}_i)$ is a linear function of $\boldsymbol{x}_i$. For the most recent research about EIDV in quantile models, the reader is referred to the paper by Hausman et al. (2019) and the references therein.

2.2 Estimation Procedures

2.2.1 Pooled Regression Method

We consider the following equations:

$$q_\tau(y_{it}|\boldsymbol{x}_{it}, v_{i,\tau}) = \boldsymbol{x}'_{it}\boldsymbol{\delta}_\tau + \sum_{s=1}^{T} \boldsymbol{x}'_{is}\boldsymbol{\eta}_{s,\tau} + \eta_{0,\tau} + v_{i,\tau} = \boldsymbol{x}'_{it}(\boldsymbol{\delta}_\tau + \boldsymbol{\eta}_{t,\tau}) + \sum_{\substack{s=1 \\ s\neq t}}^{T} \boldsymbol{x}'_{is}\boldsymbol{\eta}_{s,\tau} + \eta_{0,\tau} + v_{i,\tau},$$

and

$$q_\tau(y_{ir}|\boldsymbol{x}_{ir}, v_{i,\tau}) = \boldsymbol{x}_{ir}'\boldsymbol{\delta}_\tau + \sum_{s=1}^{T} \boldsymbol{x}_{is}'\boldsymbol{\eta}_{s,\tau} + \eta_{0,\tau} + v_{i,\tau} = \boldsymbol{x}_{ir}'(\boldsymbol{\delta}_\tau + \boldsymbol{\eta}_{r,\tau}) + \sum_{\substack{s=1\\ s\neq r}}^{T} \boldsymbol{x}_{is}'\boldsymbol{\eta}_{s,\tau} + \eta_{0,\tau} + v_{i,\tau},$$

where $t \neq r$. Obviously, the effects of $\boldsymbol{x}_{it}$ upon the conditional quantile $q_\tau(y_{it}|\boldsymbol{x}_{it}, v_{i,\tau})$ are through two channels: (1) a direct effect $\boldsymbol{x}_{it}'\boldsymbol{\delta}_\tau$ and (2) an indirect effect $\boldsymbol{x}_{it}'\boldsymbol{\eta}_{t,\tau}$. In contrast, $\boldsymbol{x}_{it}$ affects $q_\tau(y_{ir}|\boldsymbol{x}_{ir}, v_{i,\tau})$ only through the effect of $\boldsymbol{x}_{it}'\boldsymbol{\eta}_{t,\tau}$, a part of the unobservable $\alpha_{i,\tau}$. Hence, $\boldsymbol{\delta}_\tau$ is given by the following equation:

$$\boldsymbol{\delta}_\tau = \frac{\partial q_\tau(y_{it}|\boldsymbol{x}_{it}, v_{i,\tau})}{\partial \boldsymbol{x}_{it}} - \frac{\partial q_\tau(y_{ir}|\boldsymbol{x}_{ir}, v_{i,\tau})}{\partial \boldsymbol{x}_{it}} = \frac{\partial q_\tau(y_{ir}|\boldsymbol{x}_{ir}, v_{i,\tau})}{\partial \boldsymbol{x}_{ir}} - \frac{\partial q_\tau(y_{it}|\boldsymbol{x}_{it}, v_{i,\tau})}{\partial \boldsymbol{x}_{ir}}.$$

That is, $\boldsymbol{\delta}_\tau$ tells us how much $\boldsymbol{x}_{it}$ affects $q_\tau(y_{it}|\boldsymbol{x}_{it}, v_{i,\tau})$ above the effect that goes through the unobservable $\alpha_{i,\tau}$.

We adopt the pooled regression as in Abrevaya and Dahl (2008) by stacking $\boldsymbol{x}_{it}$ at different time t together. Let

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{i1} \\ \vdots \\ y_{iT} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & \boldsymbol{x}_{11}' & \boldsymbol{x}_{11}' & \cdots & \boldsymbol{x}_{1T}' \\ \vdots & \vdots & \cdots & \vdots & \\ 1 & \boldsymbol{x}_{1T}' & \boldsymbol{x}_{11}' & \cdots & \boldsymbol{x}_{1T}' \\ \vdots & \vdots & \cdots & \vdots & \\ 1 & \boldsymbol{x}_{i1}' & \boldsymbol{x}_{i1}' & \cdots & \boldsymbol{x}_{iT}' \\ \vdots & \vdots & \cdots & \vdots & \\ 1 & \boldsymbol{x}_{iT}' & \boldsymbol{x}_{i1}' & \cdots & \boldsymbol{x}_{iT}' \\ \vdots & \vdots & \cdots & \vdots & \\ 1 & \boldsymbol{x}_{N1}' & \boldsymbol{x}_{N1}' & \cdots & \boldsymbol{x}_{NT}' \\ \vdots & \vdots & \cdots & \vdots & \\ 1 & \boldsymbol{x}_{NT}' & \boldsymbol{x}_{N1}' & \cdots & \boldsymbol{x}_{NT}' \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{g}_{11}' \\ \vdots \\ \boldsymbol{g}_{1T}' \\ \vdots \\ \boldsymbol{g}_{i1}' \\ \vdots \\ \boldsymbol{g}_{iT}' \\ \vdots \\ \boldsymbol{g}_{N1}' \\ \vdots \\ \boldsymbol{g}_{NT}' \end{pmatrix} \equiv \boldsymbol{G}.$$

be the dependent and the explanatory variables of the quantile regression, respectively, where $\boldsymbol{g}_{11}'$ denotes the first row vector in the above design matrix $\boldsymbol{G}$, $\boldsymbol{g}_{12}'$ denotes the second row vector, etc. Then, we now consider the following transformed model from (2),

$$q_\tau(\boldsymbol{g}_{it}, v_{i,\tau}) = \boldsymbol{g}_{it}'\boldsymbol{\xi}_\tau + v_{i,\tau}, \tag{3}$$

where $\boldsymbol{\xi}_\tau$ includes $\boldsymbol{\delta}_\tau$ and $\boldsymbol{\eta}_\tau$'s.

2.2.2 Quasi-Likelihood Approach

For a conditional quantile regression model proposed by Koenker and Bassett (1978), the estimation of parameters can be obtained by minimizing the following objective (loss) function

$$\hat{\boldsymbol{\xi}}_\tau = \arg\min_{\boldsymbol{\xi}_\tau} \sum_{t=1}^{T} \rho_\tau(y_t - q_\tau(y_t|\boldsymbol{x}_t, \boldsymbol{\xi}_\tau)),$$

where $\rho_\tau(x) = x(\tau - I_{x<0})$ is the check function, $q_\tau(y_t|\boldsymbol{x}_t, \boldsymbol{\xi}_\tau)$ is the conditional quantile regression function of y_t given $\boldsymbol{x}_t$, and satisfies $P(y_t \leq q_\tau(y_t|\boldsymbol{x}_t, \boldsymbol{\xi}_\tau)) = \tau$. Komunjer (2005) generalized the estimation of Koenker and Bassett (1978) by proposing a class of quasi-maximum likelihood estimators (QMLE), obtained by solving

$$\hat{\boldsymbol{\xi}}_\tau = \arg\max_{\boldsymbol{\xi}_\tau} \sum_{t=1}^{T} \ln l_t(y_t, q_\tau(y_t|\boldsymbol{x}_t, \boldsymbol{\xi}_\tau)),$$

where $l_t(\cdot)$ is the conditional quasi-likelihood function y_t given $\boldsymbol{x}_t$ at time t.

Using the aforementioned idea, we consider a class of QMLEs for the conditional quantile regression model defined in (2), obtained by solving the maximization of a quasi-likelihood function for the τth conditional quantile

$$\hat{\boldsymbol{\xi}}_\tau = \arg\max_{\boldsymbol{\xi}_\tau} \sum_{i=1}^{N} \sum_{t=1}^{T} \ln l(y_{it}, q_\tau(y_{it}|\boldsymbol{g}_{it}, \boldsymbol{\xi}_\tau)), \tag{4}$$

where $l(y_{it}, q_\tau(y_{it}, \boldsymbol{g}_{it}, \boldsymbol{\xi}_\tau))$ is the quasi-likelihood function for the τth conditional quantile on individual i at time t. Indeed, model (2) can be expressed as follows:

$$q_\tau(y_{it}|\boldsymbol{x}_{it}, v_{i,\tau}) = q_{-v}(y_{it}|\boldsymbol{x}_{it}) + v_{i,\tau},$$

where $q_{-v}(y_{it}|\boldsymbol{x}_{it}) = \boldsymbol{x}'_{it}\boldsymbol{\delta}_\tau + \sum_{s=1}^{T} \boldsymbol{x}'_{is}\boldsymbol{\eta}_{s,\tau} + \eta_{0,\tau}$. Hence, $v_{i,\tau}$ in the quasi-likelihood function $l(y_{it}, q_\tau(y_{it}|\boldsymbol{x}_{it}))$ can be integrated out so that we obtain the integrated quasi-likelihood function

$$l(y_{it}, q_{-v}(y_{it}|\boldsymbol{x}_{it})) = \int l(y_{it}, q_\tau(y_{it}|\boldsymbol{x}_{it}, v_{i,\tau}))f(v_{i,\tau})dv_{i,\tau}, \tag{5}$$

by assuming that $\{v_{i,\tau}\}$ are independent and identically distributed (iid) and independent of $\{\boldsymbol{x}_{it}, y_{it}\}$, where $f(v)$ is the density function of $v_{i,\tau}$, which is assumed to be free of parameter $\boldsymbol{\delta}_\tau$.

Generally speaking, the integral in (5) does not have a close form unless it is assumed that $v_{i,\tau}$ has some particular density, such as the normality assumption as in Cai et al. (2018). Therefore, without the normality assumption of $v_{i,\tau}$, we propose using the Laplace approximation[1] method to approximate the above integral as in Breslow and Clayton (1993) and Harding and Hausman (2007). The Laplace approximation is a two-term Taylor expansion on the log density function. If $\tilde{v}$ denotes the maxima of a density function $f(v)$, then it is also the maxima of the log probability density function $p(v) = \log f(v)$ and because $\dot{p}(\tilde{v}) = 0$, we can write:

$$\begin{aligned} p(v) &\simeq p(\tilde{v}) + (v-\tilde{v})\dot{p}(\tilde{v}) + \frac{1}{2}(v-\tilde{v})^2\ddot{p}(\tilde{v}) = p(\tilde{v}) + 0 + \frac{1}{2}(v-\tilde{v})^2\ddot{p}(\tilde{v}) \\ &= c + \frac{1}{2}(v-\tilde{v})^2\ddot{p}(\tilde{v}) = c - \frac{(v-a)^2}{2b^2}, \end{aligned}$$

where $a = \tilde{v}$, $b = \{-\ddot{p}(\tilde{v})\}^{-1/2}$ with $\ddot{p}(\tilde{v}) < 0$ because $\tilde{v}$ is a maxima, and $c = p(\tilde{v})$. By assuming the mean of the disturbance v_i is zero; see Assumption (A2) later, thus $f(v)$ can be approximated by:

$$\tilde{f}(v) = c \cdot e^{-\frac{v^2}{2\sigma^2}},$$

where σ^2 is the variance of v_i. Since $\tilde{f}(v)$ is actually a probability density function, c is a function of σ with $c = 1/\sqrt{2\pi}\sigma$. Therefore, the approximation yields

$$l(y_{it}, q_{-v}(y_{it}|\boldsymbol{x}_{it})) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\ l(y_{it}, q_\tau(y_{it}|\boldsymbol{x}_{it}, v_{i,\tau}))\ e^{-\frac{v_{i,\tau}^2}{2\sigma^2}}\ dv_{i,\tau}. \tag{6}$$

According to Komunjer (2005), different choices of $l(y, q)$ would affect the asymptotic theory of QMLE for quantile models. In this paper, for simplicity, we define $l(y, q)$ as

$$l(y, q) = e^{-\rho_\tau(y-q)}, \tag{7}$$

which is a probability density function, which is a member of the so-called tick-exponential family; see Komunjer (2005) for details. Substituting (7) into the integrated quasi-likelihood function for the τth conditional quantile in (6), we have

$$l(y_{it}, q_{-v}(y_{it}|\boldsymbol{x}_{it})) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\ \exp\Big\{ -\rho_\tau\big(y_{it} - q_\tau(y_{it}|\boldsymbol{x}_{it}, v_{i,\tau})\big) - \frac{v_{i,\tau}^2}{2\sigma^2}\Big\}\ dv_{i,\tau}.$$

By a simple calculation, we get

$$l_\tau(u_{it}, \sigma) \equiv l(y_{it}, q_{-v}(y_{it}|\boldsymbol{x}_{it})) = \exp\{-\rho_\tau(u_{it})\}\ \nu_\tau(u_{it}, \sigma)\ \big(I_{u_{it}\geq 0} + \exp\{-u_{it}\} I_{u_{it}<0}\big), \tag{8}$$

[1]For any probability densify function that is smooth and unimodal, it can be approximated by a normal density function, which is the so-called Laplace approximation. Indeed, it is widely used in the literature; see, for instance, the papers by Breslow and Clayton (1993) and Harding and Hausman (2007).

where $u_{it} = y_{it} - q_\tau(y_{it}|\boldsymbol{x}_{it}, v_{i,\tau})$, $\nu_\tau(u_{it}, \sigma) = \exp\{\tau^2\sigma^2/2\}\Phi(u_{it}/\sigma - \tau\sigma) + \exp\{(\tau-1)^2\sigma^2/2\}\Phi\big(-u_{it}/\sigma + (\tau-1)\sigma\big)\exp\{u_{it}\}$, and $\Phi(\cdot)$ is the standard normal distribution function. Therefore, the log quasi-likelihood function $L_\tau(\xi_\tau, \sigma)$ is given by

$$\begin{aligned} L_\tau(\xi_\tau, \sigma) &= \sum_{i=1}^{N}\sum_{t=1}^{T} \ln l_\tau(u_{it}, \sigma) \\ &= \sum_{i=1}^{N}\sum_{t=1}^{T} \big[-\rho_\tau(u_{it}) + \ln(\nu_\tau(u_{it}, \sigma)) + \ln(I_{u_{it}\geq 0} + \exp\{-u_{it}\}I_{u_{it}<0})\big] \\ &= \sum_{i=1}^{N}\sum_{t=1}^{T} \big[-\rho_\tau(u_{it}) + \ln(\nu_\tau(u_{it}, \sigma)) - u_{it}I_{u_{it}<0}\big] \\ &= \sum_{i=1}^{N}\sum_{t=1}^{T} \left[-\tau u_{it} + \ln(\nu_\tau(u_{it}, \sigma)\right]. \end{aligned}$$

Hence, for a given $\tilde{\sigma}$, the QMLE of $\boldsymbol{\xi}_\tau$ is obtained by

$$\hat{\boldsymbol{\xi}}_\tau = \arg\max_{\boldsymbol{\xi}_\tau} L_\tau(\boldsymbol{\xi}_\tau, \tilde{\sigma}), \tag{9}$$

and the QMLE of $\boldsymbol{\delta}_\tau$ is given by

$$\hat{\boldsymbol{\delta}}_\tau = \boldsymbol{e}_1'\hat{\boldsymbol{\xi}}_\tau, \tag{10}$$

where $\boldsymbol{e}_1$ is a column vector with the first K elements are 1's and other elements are 0's. Clearly, we can estimate both $\boldsymbol{\xi}_\tau$ and σ by iterating $\tilde{\boldsymbol{\xi}}_\tau = \arg\max_{\boldsymbol{\xi}_\tau} L_\tau(\boldsymbol{\xi}_\tau, \tilde{\sigma})$ and $\tilde{\sigma} = \arg\max_\sigma L_\tau(\tilde{\boldsymbol{\xi}}_\tau, \sigma)$ until convergence.

2.3 Large Sample Theory

This section provides asymptotic results of $\hat{\boldsymbol{\xi}}_\tau$ defined in Section 2.2.2. All condtions and proofs are relegated to the Appendix. First, let $\hat{\boldsymbol{\delta}}_\tau$ be the QMLE of $\boldsymbol{\delta}_\tau$ obtained by solving (9) and (10). Next, some additional notations are defined here for the following asymptotic theorem. To this end, let $\boldsymbol{g}_{it} = (\boldsymbol{x}_{it}, \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT})'$, $u = y - q_\tau(y|\boldsymbol{x})$, $\psi_\tau(u, \sigma) = -\partial \ln l_\tau(u, \sigma)/\partial u \equiv \tau - h_\tau(u, \sigma)$, and $\dot{h}_\tau(u, \sigma) = \partial h_\tau(u, \sigma)/\partial u$, where $h_\tau(u, \sigma) = \partial \mathrm{ln}(\nu_\tau(u, \sigma))/\partial u$. Also, define $D_{g\dot{h}}(\sigma) = \Sigma_0 E[-\dot{h}_\tau(y_{it} - q_\tau(y_{it}|\boldsymbol{x}_{it}), \sigma)]$, $D_g(\sigma) = \Sigma_0 E[\psi_\tau(y_{it} - q_\tau(y_{it}|\boldsymbol{x}_{it}), \sigma)]^2$ and $D_{g_{1t}}(\sigma) = \Sigma_{t-1} E[\psi_\tau(y_{i1} -$

$q_\tau(y_{i1}|\boldsymbol{x}_{i1}),\sigma)\psi_\tau(y_{it}-q_\tau(y_{it}|\boldsymbol{x}_{it}),\sigma)]$, where

$$\Sigma_0 = \begin{pmatrix} 1 & E(\boldsymbol{x}'_{it}) & E(\boldsymbol{x}'_{i1}) & \cdots & E(\boldsymbol{x}'_{iT}) \\ E(\boldsymbol{x}_{it}) & E(\boldsymbol{x}_{it}\boldsymbol{x}'_{it}) & E(\boldsymbol{x}_{it}\boldsymbol{x}'_{i1}) & \cdots & E(\boldsymbol{x}_{it}\boldsymbol{x}'_{iT}) \\ E(\boldsymbol{x}_{i1}) & E(\boldsymbol{x}_{i1}\boldsymbol{x}'_{it}) & E(\boldsymbol{x}_{i1}\boldsymbol{x}'_{i1}) & \cdots & E(\boldsymbol{x}_{i1}\boldsymbol{x}'_{iT}) \\ \vdots & \vdots & \cdots & \vdots & \\ E(\boldsymbol{x}_{iT}) & E(\boldsymbol{x}_{iT}\boldsymbol{x}'_{it}) & E(\boldsymbol{x}_{iT}\boldsymbol{x}'_{i1}) & \cdots & E(\boldsymbol{x}_{iT}\boldsymbol{x}'_{iT}) \end{pmatrix}$$

and

$$\Sigma_{t-1} = \begin{pmatrix} 1 & E(\boldsymbol{x}'_{it}) & E(\boldsymbol{x}'_{i1}) & \cdots & E(\boldsymbol{x}'_{iT}) \\ E(\boldsymbol{x}_{i1}) & E(\boldsymbol{x}_{i1}\boldsymbol{x}'_{it}) & E(\boldsymbol{x}_{i1}\boldsymbol{x}'_{i1}) & \cdots & E(\boldsymbol{x}_{i1}\boldsymbol{x}'_{iT}) \\ E(\boldsymbol{x}_{i1}) & E(\boldsymbol{x}_{i1}\boldsymbol{x}'_{it}) & E(\boldsymbol{x}_{i1}\boldsymbol{x}'_{i1}) & \cdots & E(\boldsymbol{x}_{i1}\boldsymbol{x}'_{iT}) \\ \vdots & \vdots & \cdots & \vdots & \\ E(\boldsymbol{x}_{iT}) & E(\boldsymbol{x}_{iT}\boldsymbol{x}'_{it}) & E(\boldsymbol{x}_{iT}\boldsymbol{x}'_{i1}) & \cdots & E(\boldsymbol{x}_{iT}\boldsymbol{x}'_{iT}) \end{pmatrix}.$$

Theorem 1 (Asymptotic Normality)**.** *Under Assumptions A1-A5 in the Appendix, we have*

$$\sqrt{NT}(\hat{\boldsymbol{\delta}}_\tau - \boldsymbol{\delta}_\tau) \;\to\; \mathcal{N}(0,\Sigma(\sigma,T)),$$

where $\Sigma(\sigma,T) = \boldsymbol{e}'_1 D^{-1}_{gh}(\sigma)\big\{D_g(\sigma) + \sum_{t=2}^T 2(1-t/T+1/T)D_{g_{1t}}(\sigma)\big\}D^{-1}_{gh}(\sigma)\boldsymbol{e}_1$.

Theorem 1 shows that the convergence rate of the estimator $\hat{\boldsymbol{\delta}}_\tau$ is $\sqrt{NT}$. We can observe that the covariance-variance matrix $\Sigma(\sigma,T)$ is a function of both T and σ. When $\sigma=0$, the model degenerates to the case that $v_{i,\tau}=0$. To be specific, when $v_{i,\tau}=0$, i.e., $\sigma=0$,

$$h_\tau(u,0) = \lim_{\sigma\to 0^+}\frac{\partial \ln(\nu_\tau(u,\sigma))}{\partial u} = \lim_{\sigma\to 0^+}\frac{e^{u+\frac{1-2\tau}{2}\sigma^2}\Phi(-\frac{u}{\sigma}+(\tau-1)\sigma)}{e^{u+\frac{1-2\tau}{2}\sigma^2}\Phi(-\frac{u}{\sigma}+(\tau-1)\sigma)+\Phi(\frac{u}{\sigma}-\tau\sigma)} = I_{u<0},$$

and

$$D_g(0) = \Sigma_0 E[-\dot{h}_\tau(y_{it}-q_\tau(y_{it}|\boldsymbol{x}_{it}),0)] = \Sigma_0 f_y(F^{-1}(\tau)),$$

where $f_y(\cdot)$ is the density function of dependent variable y_{it} and $F(\cdot)$ is its CDF. The second equality follows from (A.7)-(A.9) in the Appendix if we use the Knight identity in Knight (1998) for cross sectional data; see the detailed proof of Theorem 1 in the Appendix. Then, we can rewrite the asymptotic covariance in Theorem 1 for this special case as follows

$$\begin{aligned}\Sigma(0,T) =& \boldsymbol{e}'_1 f_y^{-2}(F^{-1}(\tau))\Sigma_0^{-1}\Big\{\tau(1-\tau)\Sigma_0 \\ &+\sum_{t=2}^T 2(1-t/T+1/T)\Sigma_{t-1}E[\psi_\tau(y_{i1}-q_\tau(y_{i1}|\boldsymbol{x}_{i1}),\sigma)\psi_\tau(y_{it}-q_\tau(y_{it}|\boldsymbol{x}_{it}),\sigma)]\Big\}\Sigma_0^{-1}\boldsymbol{e}_1,\end{aligned}$$

Clearly, $\Sigma(0,T)$ is the asymptotic covariance with no disturbance $\{v_i\}$. The existence of the second term $\sum_{t=2}^{T} 2(1-t/T+1/T)\Sigma_{t-1}E[\psi_\tau(u_{i1},\sigma)\psi_\tau(u_{it},\sigma)]$ is because we allow $\{\boldsymbol{x}_{it}\}$ to be correlated over t; see Assumption A1 in the Appendix. When $T=1$,

$$\Sigma(0,1) = \boldsymbol{e}_1'\tau(1-\tau)f_y^{-2}(F^{-1}(\tau))\Sigma_0^{-1}\Sigma_0\Sigma_0^{-1}\boldsymbol{e}_1 = \boldsymbol{e}_1'\tau(1-\tau)f_y^{-2}(F^{-1}(\tau))\Sigma_0^{-1}\boldsymbol{e}_1,$$

which is the asymptotic covariance matrix for classical linear quantile model for i.i.d. cross sectional data. Therefore, model (3) and its corresponding estimator can be viewed as an extension of quantile regression from cross sectional data to panel data.

2.4 Covariance Estimation

To make a statistical test, one needs a consistent estimate of the asymptotic covariance matrix $\Sigma(\sigma,T)$. The explicit expression of $\Sigma(\sigma,T)$ in Theorem 1 provides a natural consistent estimator. We will show $\hat{\Sigma}(\sigma,T)$ is a consistent estimator of $\Sigma(\sigma,T)$ for any given σ. First, define

$$\hat{D}_g(\sigma) = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{X}_{0,it}\cdot(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\psi_\tau^2(y_{it}-\hat{y}_{it,\tau},\sigma),$$

$$\hat{D}_{g_{1t}}(\sigma) = (N(T-t))^{-1}\sum_{i=1}^{N}\sum_{s=1}^{T-t}\boldsymbol{X}_{t,is}\cdot(N(T-t))^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T-t}\psi_\tau(y_{is}-\hat{y}_{is,\tau},\sigma)\psi_\tau(y_{i,(s+t)}-\hat{y}_{i,(s+t),\tau},\sigma),$$

and

$$\hat{D}_{g\dot{h}}(\sigma) = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{X}_{0,it}\cdot(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{h}_\tau(y_{it}-\hat{y}_{it,\tau},\sigma),$$

where $\hat{y}_{it,\tau} = (1,\boldsymbol{x}_{it}',\boldsymbol{x}_{i1}',\ldots,\boldsymbol{x}_{iT}')\hat{\boldsymbol{\xi}}_\tau$,

$$\boldsymbol{X}_{0,it} = \begin{pmatrix} 1 & \boldsymbol{x}_{it}' & \boldsymbol{x}_{i1}' & \cdots & \boldsymbol{x}_{iT}' \\ \boldsymbol{x}_{it} & \boldsymbol{x}_{it}\boldsymbol{x}_{it}' & \boldsymbol{x}_{it}\boldsymbol{x}_{i1}' & \cdots & \boldsymbol{x}_{it}\boldsymbol{x}_{iT}' \\ \boldsymbol{x}_{i1} & \boldsymbol{x}_{i1}\boldsymbol{x}_{it}' & \boldsymbol{x}_{i1}\boldsymbol{x}_{i1}' & \cdots & \boldsymbol{x}_{i1}\boldsymbol{x}_{iT}' \\ \vdots & \vdots & \cdots & \vdots & \\ \boldsymbol{x}_{iT} & \boldsymbol{x}_{iT}\boldsymbol{x}_{it}' & \boldsymbol{x}_{iT}\boldsymbol{x}_{i1}' & \cdots & \boldsymbol{x}_{iT}\boldsymbol{x}_{iT}' \end{pmatrix},$$

and

$$\boldsymbol{X}_{t,is} = \begin{pmatrix} 1 & \boldsymbol{x}'_{i,(s+t)} & \boldsymbol{x}'_{i1} & \cdots & \boldsymbol{x}'_{iT} \\ \boldsymbol{x}_{is} & \boldsymbol{x}_{is}\boldsymbol{x}'_{i,(s+t)} & \boldsymbol{x}_{is}\boldsymbol{x}'_{i1} & \cdots & \boldsymbol{x}_{is}\boldsymbol{x}'_{iT} \\ \boldsymbol{x}_{i1} & \boldsymbol{x}_{i1}\boldsymbol{x}'_{i,(s+t)} & \boldsymbol{x}_{i1}\boldsymbol{x}'_{i1} & \cdots & \boldsymbol{x}_{i1}\boldsymbol{x}'_{iT} \\ \vdots & \vdots & \cdots & \vdots & \\ \boldsymbol{x}_{iT} & \boldsymbol{x}_{iT}\boldsymbol{x}'_{i,(s+t)} & \boldsymbol{x}_{iT}\boldsymbol{x}'_{i1} & \cdots & \boldsymbol{x}_{iT}\boldsymbol{x}'_{iT} \end{pmatrix}.$$

It can be easily shown that $\hat{D}_g(\sigma) = D_g(\sigma) + o_p(1)$, $\hat{D}_{g_{1t}}(\sigma) = D_{g_{1t}}(\sigma) + o_p(1)$, and $\hat{D}_{g\dot{h}}(\sigma) = D_{g\dot{h}}(\sigma) + o_p(1)$ by the law of large numbers. Then, the consistent estimator of $\Sigma(\sigma, T)$ is given by

$$\hat{\Sigma}(\sigma, T) = \boldsymbol{e}'_1 \hat{D}^{-1}_{g\dot{h}}(\sigma) \Big\{ \hat{D}_g(\sigma) + \sum_{t=2}^{T} 2(1 - t/T + 1/T)\hat{D}_{g_{1t}}(\sigma) \Big\} \hat{D}^{-1}_{g\dot{h}}(\sigma)\boldsymbol{e}_1.$$

However, we still do not know the value of nuisance parameter σ. We will replace it by $\tilde{\sigma}$ which is obtained by iterating $\tilde{\boldsymbol{\xi}}_\tau = \arg\max_{\boldsymbol{\xi}_\tau} L_\tau(\boldsymbol{\xi}_\tau, \tilde{\sigma})$ and $\tilde{\sigma} = \arg\max_\sigma L_\tau(\tilde{\boldsymbol{\xi}}_\tau, \sigma)$ until convergence for statistical inference.

To test $H_0 : R\boldsymbol{\delta}_\tau = r_\tau$ (including the simple test $H_0 : \delta_{j,\tau} = 0$ for each j), where R is a $J \times K$ known matrix with the rank J and r_τ is a known constant, a Wald type test statistic can be constructed as follows:

$$W_N = N(R\hat{\boldsymbol{\delta}}_\tau - r_\tau)' \Big[R\hat{\Sigma}(\tilde{\sigma}, T)R' \Big]^{-1} (R\hat{\boldsymbol{\delta}}_\tau - r_\tau).$$

The limiting distribution of W_N is stated in the following theorem which can be established easily by Theorem 1 and Slutsky's theorem and omitted.

Theorem 2 (Asymptotic χ^2 Test)**.** *Under Assumptions A1-A5 in the Appendix and the null hypothesis $H_0 : R\boldsymbol{\delta}_\tau = r_\tau$, we have*

$$W_N \xrightarrow{d} \chi^2_J,$$

where χ^2_J is the χ^2-distribution with J degrees of freedom.

3 Simulation Study

In this section, we examine the finite sample performance of the proposed estimator of our model. We consider the following data generating process:

$$y_{it} = \delta_0 + \delta_1 x_{it,1} + \delta_2 x_{it,2} + \alpha_i + (0.1 + 0.3x_{it,1} + 0.2x_{it,2})e_{it}, \tag{11}$$

where $\alpha_i = \sum_{t=1}^{T} \eta_{t,1} x_{it,1} + \eta_{t,2} x_{it,2} + v_i$ and T is set by 2. Here, the sequence $\{e_{it}\}$ is generated from $\mathcal{N}(0,1)$ and the covariates $x_{it,1}$ and $x_{it,2}$ are generated from i.i.d. $U(0,2)$ and $U(0,3)$, respectively. Furthermore, the measurement error term $\{v_i\}$ is a sequence of i.i.d. random variables generated from $\mathcal{N}(0, 0.1^2)$, $\mathcal{N}(0, 0.2^2)$, $\mathcal{N}(0, 0.3^2)$, $Laplace(0, 0.2^2)$, and $t(5)$ distribution. Given that variance of $t(5)$ is much bigger than that of the other distributions, we set $2.5 v_i \sim t(5)$. The coefficients are set by $\delta_0 = 3$, $\delta_1 = -2.5$, $\delta_2 = 2$, $\eta_{1,1} = 1.5$, $\eta_{1,2} = -2$, $\eta_{2,1} = -3$, and $\eta_{2,2} = 2.5$, respectively. Therefore, the quantile model of the DGP in (11) is

$$q_\tau(y_{it}) = \delta_{0,\tau} + \delta_{1,\tau} x_{it,1} + \delta_{2,\tau} x_{it,2} + \alpha_i,$$

where $\delta_{0,\tau} = 3 + 0.1\Phi^{-1}(\tau)$, $\delta_{1,\tau} = -2.5 + 0.3\Phi^{-1}(\tau)$, $\delta_{3,\tau} = 2 + 0.2\Phi^{-1}(\tau)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and $\alpha_i = \sum_{t=1}^{T} \eta_{t,1} x_{it,1} + \eta_{t,2} x_{it,2} + v_i$. The individual effect $\alpha_{i,\tau}$ in model (1) is actually $\delta_{0,\tau} + \alpha_i$ here.

To evaluate the performance of $\hat{\delta}_{j,\tau}$, for $0 \le j \le 2$, we use the absolute deviation error (ADE) of estimator, which is defined as

$$\text{ADE}(\hat{\delta}_{j,\tau}) = |\delta_{j,\tau} - \hat{\delta}_{j,\tau}|,$$

for $0 \le j \le 2$.

We consider three sample sizes $N = 200$, 500, and 1000. For each setting, we conduct the simulation 500 times and calculate 500 absolute deviation error values. For comparison of the performance of estimators in three different sample sizes, we also calculate the median and standard deviation (SD) of 500 absolute deviation error values for each sample size.

Table 1 reports the median and standard deviation of 500 absolute deviation error values for three sample sizes. As seen from Table 1, the median of 500 absolute deviation error values decreases when the sample size N increases. For example, in the case of $v_i \sim \mathcal{N}(0, 0.1^2)$ and $\tau = 0.2$, when the sample size increases from 200 to 1000, the median of 500 ADE values for $\hat{\delta}_{0,\tau}$, $\hat{\delta}_{1,\tau}$, and $\hat{\delta}_{2,\tau}$ decreases quickly from 0.0862 to 0.0421, from 0.075 to 0.0333, and from 0.0414 to 0.0218, respectively. In the case of $v_i \sim Laplace(0, 0.2^2)$ and $\tau = 0.2$, the median of 500 absolute error values for $\hat{\delta}_{0,\tau}$, $\hat{\delta}_{1,\tau}$ and $\hat{\delta}_{2,\tau}$ decreases from 0.1389 to 0.1272, from 0.0734 to 0.0399, and from 0.0528 to 0.0278, respectively, when the sample size increases. The standard deviation of 500 ADE values also decreases when the sample size increases. For example, the standard deviation of 500 ADE values for $\hat{\delta}_{0,\tau}$, $\hat{\delta}_{1,\tau}$, and $\hat{\delta}_{2,\tau}$ decreases from 0.088 to 0.0389, from 0.0664 to 0.0295, and from

0.0396 to 0.0218, respectively, when $v_i \sim \mathcal{N}(0, 0.1^2)$ and $\tau = 0.2$. Similar results can been obtained when $\tau = 0.5$ or 0.8 and the error term v_i is generated from *Laplace* and t distribution. We also notice that when the distribution of v_i becomes heavy-tailed, the median and SD of 500 ADE values also increased. In sum, the decrease of medians and standard deviations of ADE values with larger sample size indicates that our simulation result is in consistent with the asymptotic results above.

Table 1: The median and SD of 500 ADE values for $\hat{\delta}_{0,\tau}$,$\hat{\delta}_{1,\tau}$,$\hat{\delta}_{2,\tau}$.

	$\tau = 0.2$			$\tau = 0.5$			$\tau = 0.8$		
	$\hat{\delta}_{0,\tau}$	$\hat{\delta}_{1,\tau}$	$\hat{\delta}_{2,\tau}$	$\hat{\delta}_{0,\tau}$	$\hat{\delta}_{1,\tau}$	$\hat{\delta}_{2,\tau}$	$\hat{\delta}_{0,\tau}$	$\hat{\delta}_{1,\tau}$	$\hat{\delta}_{2,\tau}$
$v_i \sim \mathcal{N}(0, 0.1^2)$									
200	0.0862	0.075	0.0414	0.0725	0.06	0.0398	0.0893	0.0714	0.0455
	(0.088)	(0.0664)	(0.0396)	(0.0671)	(0.0584)	(0.0346)	(0.0801)	(0.0671)	(0.0424)
500	0.0569	0.0468	0.0338	0.0447	0.0416	0.0243	0.0585	0.0455	0.0282
	(0.0565)	(0.0416)	(0.0259)	(0.0453)	(0.0349)	(0.0222)	(0.0551)	(0.0387)	(0.0271)
1000	0.0421	0.0333	0.0218	0.0329	0.0274	0.0188	0.046	0.0282	0.0206
	(0.0389)	(0.0295)	(0.0191)	(0.0325)	(0.0253)	(0.0158)	(0.0418)	(0.0299)	(0.0187)
$v_i \sim \mathcal{N}(0, 0.2^2)$									
200	0.1244	0.0771	0.0415	0.0795	0.0636	0.0398	0.1155	0.0775	0.0489
	(0.1057)	(0.068)	(0.0417)	(0.0737)	(0.0584)	(0.0346)	(0.0991)	(0.0688)	(0.0441)
500	0.1038	0.0498	0.0345	0.0507	0.04	0.0242	0.1031	0.0513	0.0302
	(0.075)	(0.0441)	(0.0281)	(0.0473)	(0.0353)	(0.0223)	(0.072)	(0.0411)	(0.0284)
1000	0.0997	0.0403	0.0274	0.0358	0.0268	0.0189	0.1016	0.0383	0.0237
	(0.055)	(0.0331)	(0.0218)	(0.0362)	(0.0254)	(0.0156)	(0.0605)	(0.0329)	(0.0206)
$v_i \sim \mathcal{N}(0, 0.3^2)$									
200	0.2021	0.0808	0.0481	0.0891	0.0612	0.0384	0.1937	0.0865	0.0519
	(0.1338)	(0.0709)	(0.0445)	(0.0829)	(0.0584)	(0.0347)	(0.1279)	(0.072)	(0.0463)
500	0.1928	0.0583	0.0401	0.0618	0.041	0.0238	0.1904	0.0604	0.0342
	(0.0949)	(0.0496)	(0.0324)	(0.0519)	(0.0359)	(0.0227)	(0.0909)	(0.0466)	(0.0316)
1000	0.1863	0.053	0.0373	0.0402	0.0273	0.0181	0.1895	0.0543	0.0338
	(0.0664)	(0.0386)	(0.026)	(0.0407)	(0.0254)	(0.0156)	(0.0718)	(0.0376)	(0.024)
$v_i \sim Laplace(0, 0.2^2)$									
200	0.1389	0.0734	0.0528	0.0879	0.0627	0.0422	0.1431	0.0743	0.048
	(0.1125)	(0.0628)	(0.0477)	(0.0788)	(0.0558)	(0.038)	(0.1163)	(0.0697)	(0.0508)
500	0.1282	0.0518	0.0329	0.0556	0.0391	0.0282	0.1295	0.0517	0.0356
	(0.0859)	(0.0464)	(0.0285)	(0.0493)	(0.0373)	(0.0234)	(0.0838)	(0.0468)	(0.0305)

1000	0.1272	0.0399	0.0278	0.0442	0.0279	0.018	0.1266	0.0404	0.0272
	(0.0681)	(0.0351)	(0.0214)	(0.0366)	(0.0255)	(0.0166)	(0.0649)	(0.0372)	(0.0237)
$2.5v_i \sim t(5)$									
200	0.3107	0.0848	0.0622	0.1056	0.067	0.0439	0.2967	0.0925	0.0575
	(0.1866)	(0.0762)	(0.049)	(0.1037)	(0.0587)	(0.0394)	(0.1733)	(0.0781)	(0.0543)
500	0.2882	0.0736	0.05	0.0716	0.0414	0.028	0.2904	0.066	0.0466
	(0.1227)	(0.0566)	(0.0366)	(0.068)	(0.0365)	(0.0247)	(0.1248)	(0.0538)	(0.037)
1000	0.2819	0.0696	0.0416	0.0503	0.0255	0.0209	0.2823	0.0641	0.0438
	(0.0843)	(0.0435)	(0.0294)	(0.0464)	(0.0263)	(0.0181)	(0.0878)	(0.0419)	(0.0292)

4 Empirical Study

4.1 Empirical Models

To illustrate our model empirically, we test the pecking order theory in the field of corporate finance, established by Myers (1984) and Myers and Majluf (1984). There are two important reasons for using the panel quantile regression to examine the pecking order theory. On the one hand, prior literature employing mean models does not reach a consensus on whether corporate financing behaviors conform to the pecking order hypothesis (Brown et al., 2019). Some of prior studies confirm the pecking order theory analytically and empirically, indicating that firms prefer internal to external funds (i.e., the first rung of financial hierarchy), and prefer debt to equity issues (i.e., the second rung of financial hierarchy) when internal funds are exhausted and external funds are sought (Myers, 1984; Myers and Majluf, 1984; Shyam-Sunder and Myers, 1999; Fama and French, 2002; Lemmon and Zender, 2010). Nevertheless, the existing studies also document departures of financing hierarchy from the pecking order theory in different settings (Brennan and Kraus, 1987; Jung et al., 1996; Frank and Goyal, 2003; Fama and French, 2005; Brown et al., 2009; Bharath et al., 2009). Their either-or conclusions can be caused by the biased estimation of mean models. In contrast to the quantile regressions, their use of mean regressions does not take into consideration the fact that the distribution of the dependent variable is highly skewed. For instance, the distribution of corporate investments is highly skewed, which can be observed evidently from Figure 1 (the left figure for the original data and the right figure for the log-transformed data) for the investment amount in year 2016. In addition, the employment of mean regressions fails to capture the phenomenon that the contribution of a given source of funds to financing investments could change across different levels of investments (Chay et al., 2015). On the other hand, Chay et al. (2015) used quantile regressions to test the pecking order hypothesis but they did not control firm and year fixed effects due to difficulties in theoretical modeling and computational challenges. Since it is difficult to identify valid instrumental variables for financing variables, the control of unobserved firm-specific heterogeneity, taken as a principal improvement of panel quantile regressions, is therefore critical for the examination of the pecking order theory.

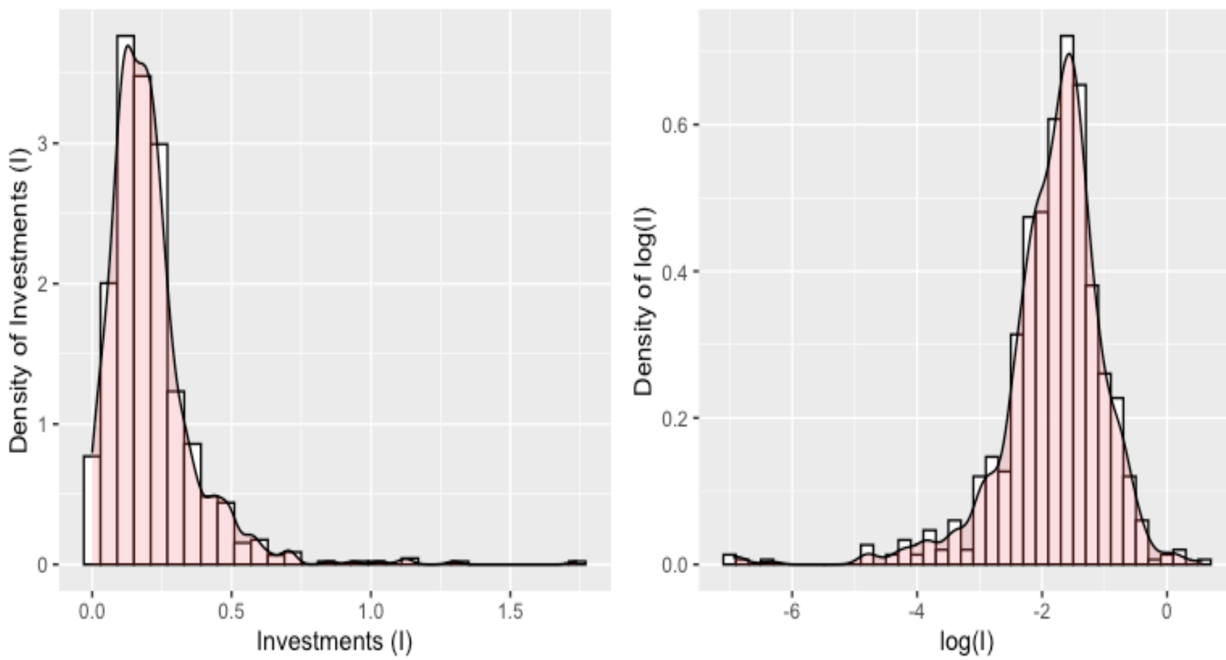

Figure 1: The density of investment amount for year 2016: Left figure for the original data and right figure for the log-transformed data.

Following Brown et al. (2009) and Chay et al. (2015), we use a modified investment model which captures the responsiveness of investments to internal and external funds:

$$q_\tau(I_{it}) = \alpha_{i,\tau} + \delta_{1,\tau} CF_{it} + \delta_{2,\tau} EXTF_{it} + \sum_{k=1}^{K} \gamma_{k,\tau} Control_{it,k}, \tag{12}$$

where I denotes the corporate investments in operating activities, measured as capital expenditures of firm i in fiscal year t. Concerning on the key independent variables, internal funds (CF) is defined as the amount of cash flow and external funds ($EXTF$) is defined as the sum of debt issues ($DEBT$) and equity issues ($EQUITY$). All of the variables mentioned above are scaled by net fixed assets. Market-to-book value (MB), cash holdings ($Cash$), firm size ($Size$), leverage (Lev) and year dummy variables $\{d_t\}$ are also added as control variables to the regression above, and to all of the other regressions. Year dummy variables $\{d_t\}$ is set for t = 2001, 2006, 2011, and 2016 since we set 1996 as the base year. We calculate the mean values of all responsive and explanatory variables in each year; see Figure 2. Indeed, we can observe there are significant differences in sample data across different years, which is consistent with the regression results of year dummies presented in Table 2. α_i is the individual effects used to control the unobserved firm-specific heterogeneity. In this paper, we approximate α_i, as discussed earlier, by a projection on the explanatory variables of firm i. Hence, the relative importance of internal and external fund can be observed from the magnitude of $\delta_{1,\tau}$ and $\delta_{2,\tau}$.

To take a further step, external funds ($EXTF$) is divided into two separate components, i.e., debt issues ($DEBT$) and equity issues ($EQUITY$). This derivative equation allows us to test the second rung of the pecking order theory:

$$q_\tau(I_{it}) = \alpha_{i,\tau} + \delta_{1,\tau} CF_{it} + \delta_{2,\tau} DEBT_{it} + \delta_{3,\tau} EQUITY_{it} + \sum_{k=1}^{K} \gamma_{k,\tau} Control_{it,k}, \tag{13}$$

where the coefficients of interest are $\delta_{1,\tau}$, $\delta_{2,\tau}$ and $\delta_{3,\tau}$, representing the responsiveness of investments to internal funds, debt issues and equity issues, respectively.

4.2 Data and Sample Construction

To generate our sample, we begin with all listed companies in the U.S. that have annual data regarding financial information in *Compustat*. We construct a balanced panel of non-financial and non-utility companies by drawing the sample every five-years, i.e., 1996, 2001, 2006, 2011, and 2016. We exclude observations with missing values from the sample and trim the outliers in all key variables following the rules as those in Cleary (1999). Therefore, our final sample consists of 3785 firm-year observations.

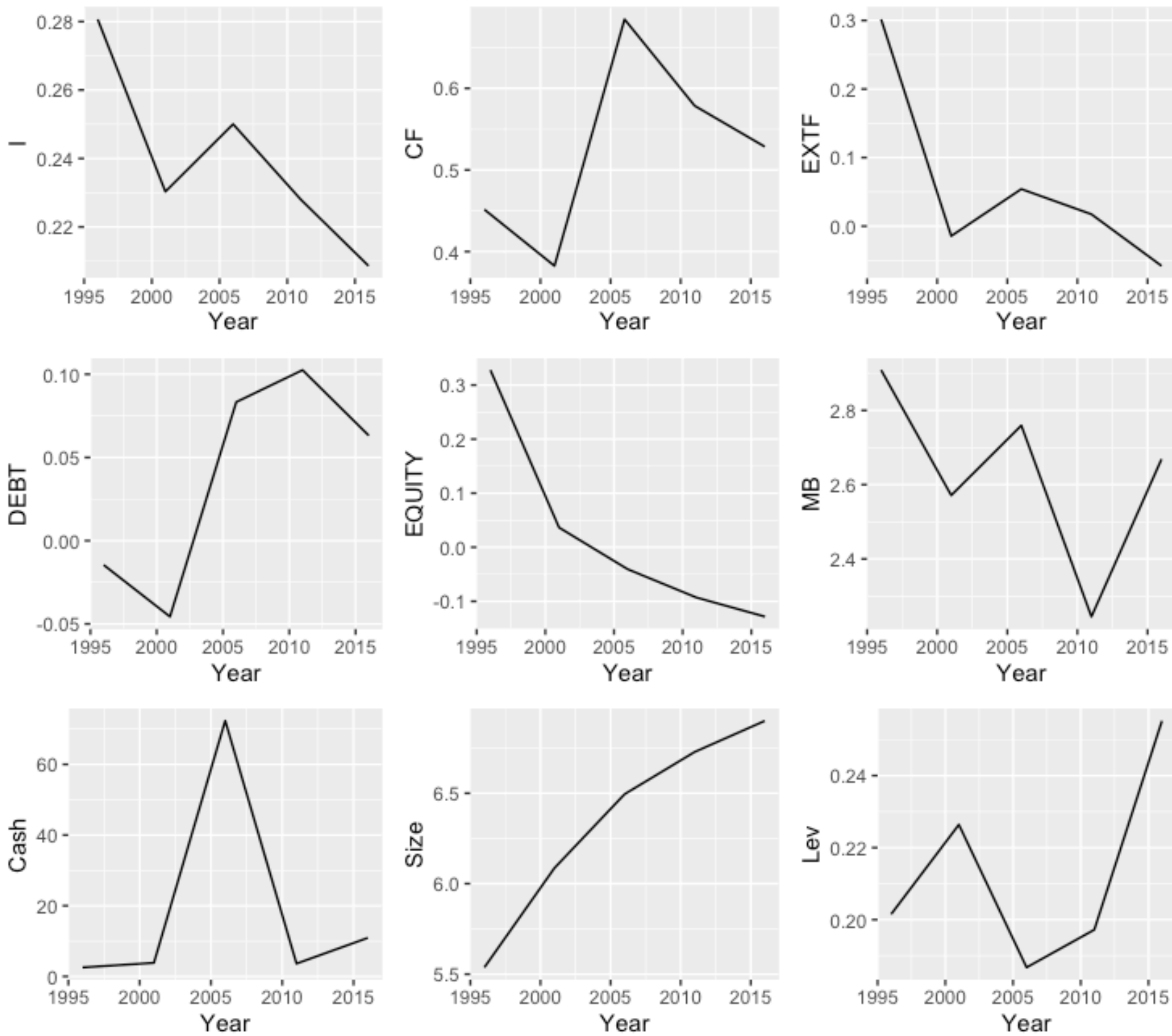

Figure 2: Time series of mean values for responsive and explanatory variables in each year.

4.3 Empirical Results

4.3.1 Testing on the First Rung of Pecking Order Theory

Panel A of Table 2 provides the panel quantile regression results as well as the OLS regression results estimated for the first rung of the pecking order theory (i.e., internal to external funds). The OLS results illustrate a simple situation on average, while the results of the quantile regression model depict a full picture of the financing hierarchy across different scales of investments.

The results of OLS regression indicate that both internal (CF) and external funds ($EXTF$) have statistically significant effect on investment activities (I). Consistent with the assumption in the first rung of pecking order theory, the responsiveness of investments to internal funds (0.148) is significantly more profound than external funds (0.068).

Concerning on the results of the panel quantile regression, the financing hierarchy of investment activities gets more complicated across different scales of investments. The estimates of $\delta_{1,\tau}$ and $\delta_{2,\tau}$ are positive and statistically significant with p-values < 0.1 at all levels of investments. To illustrate the trends of each estimate and compare the relative relationship between them, we then plot the estimates of internal and external funds as shown in Figure 3. The horizontal axis measures different scales of investment activities (I) and the vertical axis measures the estimated coefficients of different sources of funds. The solid points indicate the point estimates from the panel quantile regression and the shaded area represents the corresponding 90% confidence interval. The two dashed lines indicate the estimates of internal and external funds from the OLS regression, respectively. It is observed that as the investment quantile ranges from 10% to 90%, the magnitude of coefficients of external funds ($EXTF$) generally experiences an upward trend from 0.042 to 0.117. In the mean time, the magnitude of coefficients of internal funds (CF) rises to the second highest point at the 50% quantile and declines at the 60% quantile. Then the magnitude of coefficients of internal funds remain stable from 60% to 80% quantiles while it rises to the peak at the highest quantile. Generally, the estimated coefficients of internal funds are significantly larger than that of external funds except for the lowest investments quantile. Our results, thus, suggest that the corporate financing hierarchy is in accordance with the first rung of the pecking order theory across all investments (I) quantiles from 10% to 90%. It is reasonable that our results supports the view that internal funds are preferred to external funds, as prior literature indicate that internal funds help firms avoid transaction costs and information asymmetry costs on debt or equity issuances (Faulkender and Wang, 2006) and liquidation cost on debt issuance (Al-Najjar and Belghitar, 2011).

Table 2: Baseline Tests of Pecking Order

		Selected quantiles of investments (I)								
	OLS	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Panel A: First rung of pecking order										
CF	0.148	0.086	0.135	0.156	0.205	0.224	0.186	0.19	0.191	0.271
	(0.018)***	(0.022)***	(0.021)***	(0.022)***	(0.023)***	(0.025)***	(0.027)***	(0.03)***	(0.036)***	(0.048)***
$EXTF$	0.068	0.042	0.026	0.052	0.041	0.062	0.071	0.077	0.069	0.117
	(0.016)***	(0.016)***	(0.015)*	(0.016)***	(0.017)**	(0.018)***	(0.02)***	(0.021)***	(0.026)***	(0.035)***
MB	0.104	0.054	0.072	0.09	0.077	0.064	0.055	0.059	0.055	0.1
	(0.016)***	(0.017)***	(0.016)***	(0.017)***	(0.018)***	(0.02)***	(0.021)**	(0.023)***	(0.028)**	(0.038)***
$Cash$	0.129	-0.043	0.027	0.063	0.094	0.116	0.141	0.176	0.221	0.192
	(0.02)***	(0.025)*	(0.025)	(0.026)**	(0.027)***	(0.029)***	(0.032)***	(0.034)***	(0.042)***	(0.056)***
$Size$	-0.176	0.136	0.074	-0.005	-0.068	-0.125	-0.203	-0.288	-0.385	-0.566
	(0.053)***	(0.183)	(0.179)	(0.189)	(0.196)	(0.211)	(0.232)	(0.251)	(0.301)	(0.403)
Lev	-0.117	-0.1	-0.124	-0.083	-0.1	-0.081	-0.096	-0.125	-0.134	-0.075
	(0.021)***	(0.028)***	(0.027)***	(0.029)***	(0.03)***	(0.032)**	(0.035)***	(0.038)***	(0.046)***	(0.061)
d_{2001}	-0.184	-0.111	-0.155	-0.16	-0.162	-0.169	-0.19	-0.181	-0.205	-0.199
	(0.042)***	(0.05)**	(0.049)***	(0.051)***	(0.054)***	(0.058)***	(0.063)***	(0.068)***	(0.082)**	(0.111)*
d_{2006}	-0.131	-0.098	-0.123	-0.129	-0.136	-0.111	-0.113	-0.133	-0.141	-0.057
	(0.046)***	(0.059)*	(0.057)**	(0.061)**	(0.064)**	(0.069)	(0.075)	(0.081)	(0.098)	(0.132)
d_{2011}	-0.198	-0.18	-0.189	-0.19	-0.2	-0.183	-0.185	-0.172	-0.203	-0.106
	(0.049)***	(0.067)***	(0.065)***	(0.069)***	(0.072)***	(0.078)**	(0.084)**	(0.091)*	(0.11)*	(0.149)
d_{2016}	-0.27	-0.224	-0.233	-0.254	-0.239	-0.212	-0.235	-0.239	-0.293	-0.23
	(0.051)***	(0.073)***	(0.071)***	(0.075)***	(0.078)***	(0.085)**	(0.092)**	(0.099)**	(0.12)**	(0.162)
N	3785	3785	3785	3785	3785	3785	3785	3785	3785	3785
Panel B: Second rung of pecking order										
CF	0.147	0.102	0.145	0.163	0.205	0.224	0.194	0.186	0.219	0.267
	(0.019)***	(0.022)***	(0.021)***	(0.022)***	(0.023)***	(0.025)***	(0.028)***	(0.03)***	(0.036)***	(0.05)***
$DEBT$	0.049	0.022	0.03	0.029	0.03	0.054	0.061	0.053	0.077	0.086
	(0.015)***	(0.013)*	(0.013)**	(0.013)**	(0.014)**	(0.016)***	(0.017)***	(0.018)***	(0.022)***	(0.031)***
$EQUITY$	0.044	0.014	0.007	0.035	0.029	0.056	0.061	0.075	0.047	0.063
	(0.017)**	(0.019)	(0.018)	(0.019)*	(0.02)	(0.022)**	(0.024)**	(0.026)***	(0.031)	(0.044)
MB	0.104	0.044	0.086	0.09	0.076	0.061	0.059	0.061	0.064	0.111
	(0.016)***	(0.017)***	(0.016)***	(0.017)***	(0.018)***	(0.02)***	(0.021)***	(0.023)***	(0.027)**	(0.039)***
$Cash$	0.13	-0.015	0.027	0.085	0.104	0.126	0.13	0.19	0.23	0.233
	(0.02)***	(0.025)	(0.024)	(0.026)***	(0.027)***	(0.029)***	(0.032)***	(0.034)***	(0.041)***	(0.058)***
$Size$	-0.173	0.146	0.071	-0.007	-0.07	-0.148	-0.215	-0.262	-0.423	-0.482
	(0.054)***	(0.185)	(0.179)	(0.189)	(0.198)	(0.215)	(0.236)	(0.252)	(0.301)	(0.422)
Lev	-0.117	-0.089	-0.128	-0.076	-0.096	-0.089	-0.098	-0.131	-0.142	-0.076
	(0.021)***	(0.028)***	(0.027)***	(0.028)***	(0.03)***	(0.033)***	(0.036)***	(0.038)***	(0.046)***	(0.064)
d_{2001}	-0.186	-0.126	-0.155	-0.159	-0.172	-0.176	-0.2	-0.192	-0.154	-0.165
	(0.042)***	(0.049)**	(0.048)***	(0.05)***	(0.053)***	(0.058)***	(0.063)***	(0.067)***	(0.081)*	(0.114)
d_{2006}	-0.133	-0.116	-0.127	-0.126	-0.131	-0.121	-0.107	-0.134	-0.11	-0.086
	(0.046)***	(0.059)**	(0.057)**	(0.06)**	(0.063)**	(0.069)*	(0.075)	(0.079)*	(0.096)	(0.136)
d_{2011}	-0.202	-0.207	-0.194	-0.197	-0.205	-0.179	-0.18	-0.187	-0.146	-0.143
	(0.049)***	(0.066)***	(0.064)***	(0.067)***	(0.072)***	(0.078)**	(0.085)**	(0.09)**	(0.108)	(0.153)
d_{2016}	-0.273	-0.247	-0.242	-0.268	-0.247	-0.2	-0.227	-0.259	-0.227	-0.277
	(0.051)***	(0.072)***	(0.07)***	(0.073)***	(0.078)***	(0.084)**	(0.092)**	(0.097)***	(0.118)*	(0.166)*
N	3785	3785	3785	3785	3785	3785	3785	3785	3785	3785

Table 2 reports the results of tests on pecking order theory employing the OLS regression and the panel quantile regression. Panel A concerning on the first rung of the pecking order reports the estimates of the internal and external funds on investment activities. Panel B concerning on the second order of pecking order reports the estimates of the internal funds, debt issues and equity issues, respectively. The dependent variable (I) denotes the corporate investments in operating activities, measured as capital expenditures scaled by net fixed assets. The independent variable CF is defined as cash flow, scaled by net fixed assets. Another variable of interest is $EXTF$, measured as the sum of debt issues ($DEBT$) and equity issues ($EQUITY$), scaled by net fixed assets, where $DEBT$ is calculated as funds raised by debt issues and $EQUITY$ is calculated as funds raised by equity issues. We control for market-to-book value (MB), Cash holdings ($Cash$), firm size ($Size$) and leverage (Lev). Panel B reports the results of the internal funds, debt issues and equity issues, respectively. All control variables are winsorised at the 1% and 99% percentile levels, and all continuous variables are standardized afterwards. Firm-specific effects and year-specific effects are included. Standard errors are presented in parentheses. Finally, *, ** and *** represent significance levels of 10%, 5% and 1%, respectively.

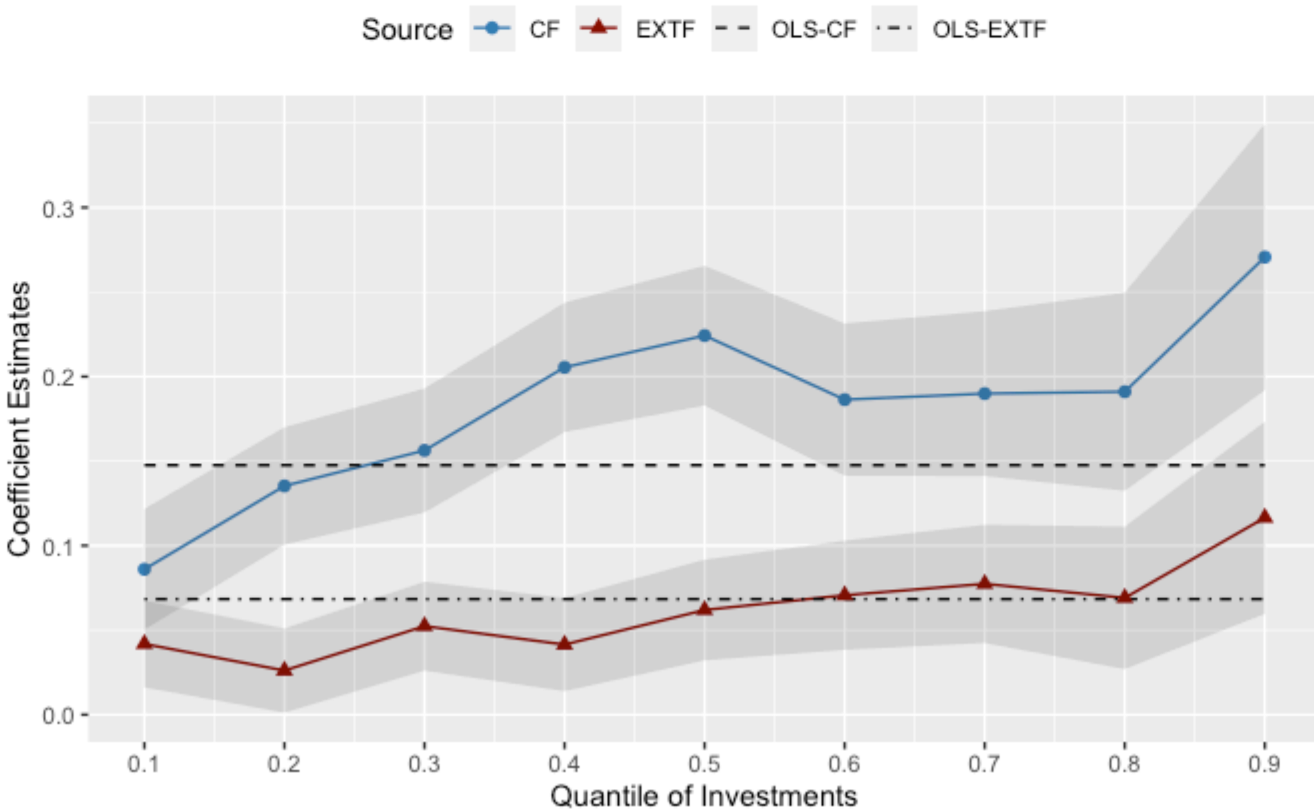

Figure 3: First Rung of Pecking Order

4.3.2 Testing on the Second Rung of Pecking Order Theory

Provided that managers in pursuit of maximizing the value for original shareholders have information advantages over outside creditors and investors, the second rung of the pecking order hypothesis theoretically assumes that managers will use debt and then equity issues in sequence to invest projects with a positive net present value. Accordingly, we take a further empirical test for the second rung of the pecking order theory.

Panel B of Table 2 represents the regression results estimated where external funds are divided into funds raised by debt issues and equity issues. The results, however, are inconsistent with the pecking order hypothesis in the second rung for both OLS regression and the panel quantile regression, despite that the first run still holds. For all quantiles of investments, the estimates of $\delta_{2,\tau}$ and $\delta_{3,\tau}$ are positive, and the estimates of $\delta_{2,\tau}$ are statistically significant with p-values < 0.1. However, the estimates of $\delta_{3,\tau}$ are statistically significant only at 30%, 50%, 60% and 70%

quantiles with p-values < 0.1.

To compare the magnitudes of estimated coefficients on cash flow (CF), debt issues ($DEBT$) and equity issues ($EQUITY$) and determine their relative importance for investment activities at each investment level, we plot the estimates of internal and two types of external funds with the similar patterns as Figure 3. As is shown in Figure 4, firms prefer internal funds to both debt and equity issues across all investment levels, in support of the first rung of the pecking order theory. As for the second rung, the magnitude of $\delta_{2,\tau}$ keeps an upward tendency while that of $\delta_{3,\tau}$ generally experience growth but with slight fluctuations at some of the quantiles. However, the shaded areas for the estimates of debt issues ($DEBT$) and equity issues ($EQUITY$) overlap for all quantiles of investments (I). It implies that the relative magnitudes of the two sources of external funds are not significantly distinctive from each other, resulting in an evidence against the second rung of the pecking order theory.

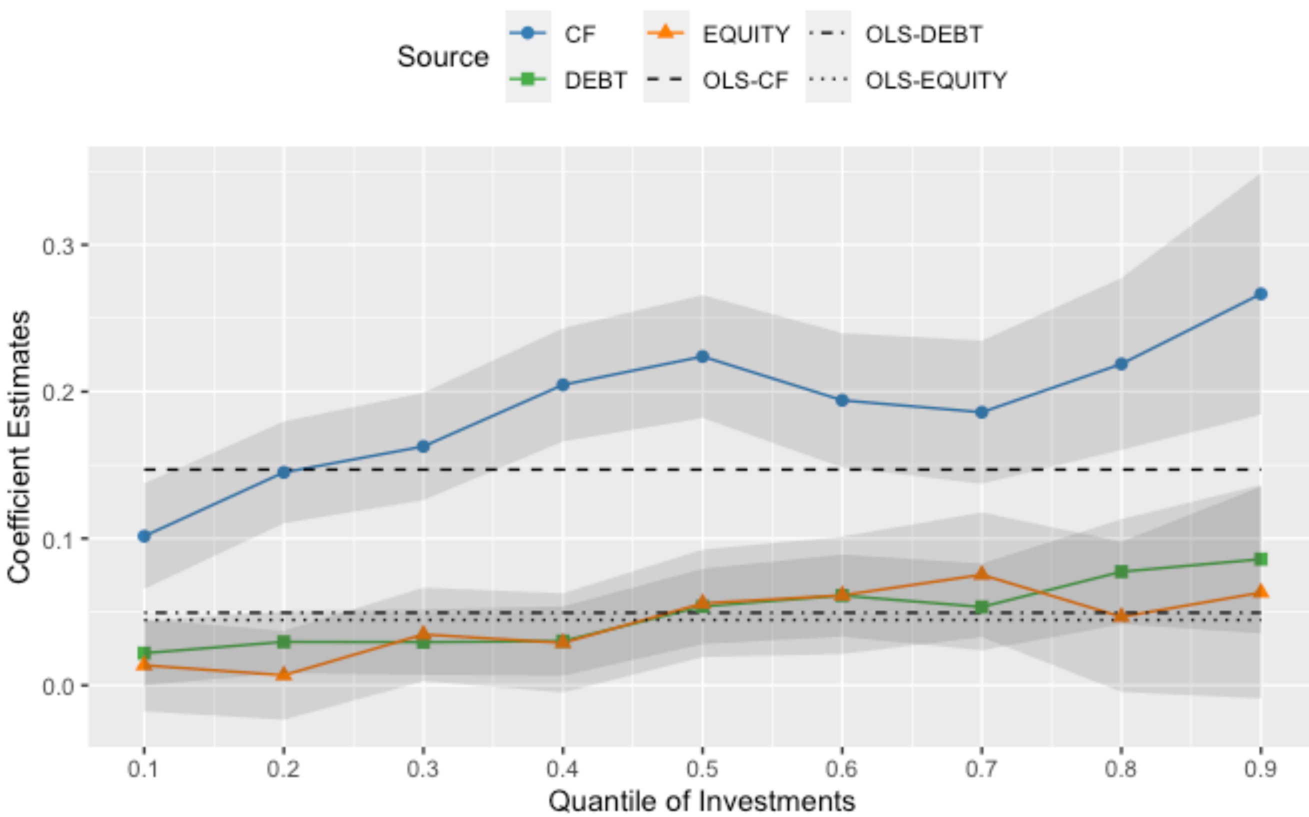

Figure 4: Second Rung of Pecking Order

Although our results reveal some interesting patterns seemingly contradictory to the traditional pecking order hypothesis, we acknowledge that the results above are not against the view that information asymmetry is the driving force of adverse selection and then corporate financing behavior in terms of pecking order theory. Still, uninformed investors rationally demand different premiums on debt issues and equity issues given that firms are confronted with severe adverse selection problems (Myers, 1984; Myers and Majluf, 1984). Our results could be attributable to the improved information environments in the U.S. thanks to the gradually updated FASB standards (Khan et al., 2018) and the growing number of small listed companies since 1990s (Frank and Goyal, 2003; Chay et al., 2015). Frank and Goyal (2003) documented that the net equity issues track the financing deficit much more closely than net debt issues do, and the bulk of external financing takes the form of equity. A particularly striking phenomenon is that a lot more small and unprofitable firms with less severe adverse selection problems became publicly traded during the 1990s so that they do not behave according to pecking order theory.

The overall results indicate that the financing hierarchy departs from the pecking order theory at any of the investments quantiles, as presented clearly in Figure 4. Meanwhile, our results are distinctive from the results documented by previous studies and we presume the difference mainly arises from our deployment of the panel quantile regression controlling firm-specific and year fixed effects.

5 Conclusion

This paper studies the estimation and inference of a quantile regression model for panel data with correlated random effects. Most previous literature concerning panel quantile models with fixed effects assume both the sample size N and the time period T go to infinity, but we only require N to go to infinity and allow T to be a fixed finite number under the framework of Abrevaya and Dahl (2008). The Monte Carlo simulation result is consistent with the asymptotic theory. We then illustrate the method with an application to testing the pecking order theory using non-financial and non-utility U.S. firms in *Compustat* database.

Finally, we note several possible extensions of the present study. For example, it is worth pointing out that y_{it} might not be independent for many applications. To characterize the dependence among individuals for panel data, it might need to assume large T. In such a way, two methods of using factors are commonly used in the literature: the common correlated effect

(CCE) by Pesaran (2006) and the interactive fixed effect (IFE) approach proposed in Bai (2009), so that our quantile model for panel data can be generalized to quantile model for panel data with cross-sectional dependence and large T. We leave such extensions as possible future research topics.

References

Abrevaya, J. and Dahl, C. M. (2008). The Effects of Birth Inputs on Birthweight: Evidence from Quantile Estimation on Panel Data. *Journal of Business and Economic Statistics*, 26(4):379–397.

Al-Najjar, B. and Belghitar, Y. (2011). Corporate Cash Holdings and Dividend Payments: Evidence from Simultaneous Analysis. *Managerial and Decision Economics*, 32(4):231–241.

Bai, J. (2009). Panel Data Models With Interactive Fixed Effects. *Econometrica*, 77(4):1229–1279.

Bharath, S. T., Pasquariello, P., and Wu, G. (2009). Does Asymmetric Information Drive Capital Structure Decisions? *Review of Financial Studies*, 22(8):3211–3243.

Brennan, M. and Kraus, A. (1987). Efficient financing under asymmetric information. *The Journal of Finance*, 42(5):1225–1243.

Breslow, N. E. and Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 88(421):9–25.

Brown, J. R., Fazzari, S. M., and Peterson, B. C. (2009). Financing Innovation and Growth: Cash Flow, External Equity, and the 1990s R&D Boom. *The Journal of Finance*, 64(1):151–185.

Brown, S., Dutordoir, M., Veld, C., and Veld-Merkoulova, Y. (2019). What is the role of institutional investors in corporate capital structure decisions? A survey analysis. *Journal of Corporate Finance*, 58:270–286.

Cai, Z., Chen, L., and Fang, Y. (2018). A Semiparametric Quantile Panel Data Model with an Application to Estimating the Growth Effect of FDI. *Journal of Econometrics*, 206(2):531–553.

Cai, Z. and Xiao, Z. (2012). Semiparametric Quantile Regression Estimation in Dynamic Models with Partially Varying Coefficients. *Journal of Econometrics*, 167(2):413–425.

Cai, Z. and Xu, X. (2008). Nonparametric Quantile Estimations for Dynamic Smooth Coefficient Models. *Journal of the American Statistical Association*, 103(484):1595–1608.

Chamberlain, G. (1982). Multivariate Regression Models for Panel Data. *Journal of Econometrics*, 18(1):5–46.

Chamberlain, G. (1984). *Panel Data*, volume 2. Handbook of Econometrics, Amsterdam: North-Holland.

Chay, J., Park, S. H., Kim, S., and Suh, J. (2015). Financing Hierarchy : Evidence from Quantile Regression. *Journal of Corporate Finance*, 33:147–163.

Cleary, S. (1999). The Relationship between Firm Investment and Financial Status. *The Journal of Finance*, 54(2):673–692.

Fama, E. F. and French, K. R. (2002). Testing trade-off and pecking order predictions about dividends and debt. *The Review of Financial Studies*, 15(1):1–33.

Fama, E. F. and French, K. R. (2005). Financing Decisions: Who Issues Stock? *Journal of Financial Economics*, 76(3):549–582.

Faulkender, M. and Wang, R. (2006). Corporate Financial Policy and the Value of Cash. *The Journal of Finance*, 61(4):1957–1990.

Frank, M. Z. and Goyal, V. K. (2003). Testing the Pecking Order Theory of Capital Structure. *Journal of Financial Economics*, 67(2):217–248.

Galvao, A. F., Lamarche, C., and 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.

Gu, J. and Volgushev, S. (2019). Panel data quantile regression with grouped fixed effects. *Journal of Econometrics*, 213(1):68–91.

Harding, M. and Lamarche, C. (2017). Penalized Quantile Regression with Semiparametric Correlated Effects: An Application with Heterogeneous Preferences. *Journal of Applied Econometrics*, 32(2):342–358.

Harding, M. C. and Hausman, J. (2007). Using a Laplace Approximation to Estimate the Random Coefficients Logit Model by Nonlinear Least Squares. *International Economic Review*, 48(4):1311–1328.

Hausman, J., Luo, Y. E., and Palmer, C. (2019). Errors in the Dependent Variable of Quantile Regression Models. *NBER Working Paper No. 25819*.

Jung, K., Kim, Y.-C., and Stulz, R. (1996). Timing, Investment Opportunities, Managerial Discretion, and the Security Issue Decision. *Journal of Financial Economics*, 42(2):159–185.

Khan, U., Li, B., Rajgopal, S., and Venkatachalam, M. (2018). Do the FASB's Standards Add Shareholder Value? *The Accounting Review*, 93(2):209–247.

Knight, K. (1998). Limiting Distributions for L1 Regression Estimators under General Conditions. *The Annals of Statistics*, 26(2):755–770.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89.

Koenker, R. and Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46(1):33–50.

Koenker, R. and Zhao, Q. (1996). Conditional Quantile Estimation and Inference for ARCH Models. *Econometric Theory*, 12(5):793–813.

Komunjer, I. (2005). Quasi-maximum Likelihood Estimation for Conditional Quantiles. *Journal of Econometrics*, 128(1):137–164.

Lamarche, C. (2010). Robust Penalized Quantile Regression Estimation for Panel Data. *Journal of Econometrics*, 157(2):396–408.

Lemmon, M. L. and Zender, J. F. (2010). Debt Capacity and Tests of Capital Structure Theories. *Journal of Financial and Quantitative Analysis*, 45(5):1161–1187.

Machado, J. A. and Santos Silva, J. M. (2019). Quantiles via Moments. *Journal of Econometrics*, 213(1):145–173.

Myers, S. (1984). Capital Structure Puzzle. *The Journal of Finance*, 39(3):575–592.

Myers, S. C. and Majluf, N. S. (1984). Corporate Financing and Investment Decisions When Firms Have Information that Investors Do Not Have. *Journal of Financial Economics*, 13(2):187–221.

Pesaran, M. H. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica*, 74(4):967–1012.

Powell, D. (2016). Quantile Regression with Nonadditive Fixed Effects. *RAND Labor and Population Working Paper*, pages 1–22.

Shyam-Sunder, L. and Myers, S. C. (1999). Testing static tradeoff against pecking order models of capital structure. *Journal of Financial Economics*, 51(2):219–244.

APPENDIX: Conditions and Proofs

First, we make the following assumptions, which are necessary to establish the asymptotic normality of our estimator by following the analysis of Cai and Xu (2008) and Cai and Xiao (2012). All notations can be found in Section 2.3.

Assumptions:

A1. The series $\{y_{it}, \boldsymbol{x}_{it}\}$ is i.i.d. across individual i but can be correlated over t for fixed i;

A2. The series $\{v_{i,\tau}\}$ is i.i.d. and independent of $\{\boldsymbol{x}_{it}\}$, and has a smooth and unimodal density function with zero mean;

A3. $\Sigma_x = E(\boldsymbol{x}_{it}\boldsymbol{x}'_{it})$ is invertible;

A4. $E(\|\boldsymbol{x}_{it}\|^{\theta}) < \infty$ with $\theta > 4$;

A5. $T \geq 2$.

Assumption A1 assumes that the observations $\{\boldsymbol{x}_{it}, y_{it}\}$ are i.i.d. across individual i, but can be correlated between different time t for fixed i. Assumption A2 is for model identification. For those mean of $v_{i,\tau}$ is not equal to zero, one can always subtract the non-zero mean and add it to the constant coefficient. Assumptions A3-A4 about the observations $\{\boldsymbol{x}_{it}\}$ guarantee the matrix $\boldsymbol{G}'\boldsymbol{G}$ in model (3) is invertible, and $E(\|\boldsymbol{g}_{it}\|)^{\theta} < \infty$ for some $\theta > 4$. Finally, Assumption A5 excludes the cross sectional data.

Next, we give the proof of Theorem 1. To this end, two lemmas and one proposition are needed as follows.

Lemma 1. *Let $V_{NT}(\Delta)$ be a vector function that satisfies*

(i) $-\Delta' V_{NT}(\lambda\Delta) \geq -\Delta' V_{NT}(\Delta), \lambda \geq 1,$

(ii) $\sup_{\|\Delta\| \leq M} \|V_{NT}(\Delta) - A_{NT} + D\Delta\| = o_p(1),$

where $\|A_{NT}\| = O_p(1)$, $0 < M < \infty$, and D is a positive-definite matrix. Suppose $\hat{\Delta}$ is a vector such that $V_{NT}(\hat{\Delta}) = o_p(1)$. Then, we have $\|\hat{\Delta}\| = O_p(1)$ and

$$\hat{\Delta} = D^{-1} A_{NT} + o_p(1).$$

Proof. See the proof of Lemma A.4 in Koenker and Zhao (1996). □

Proposition 1. *Some properties about $\psi_\tau(u, \sigma) = \tau - h_\tau(u, \sigma)$ are listed as follows.*

(1) $\psi_\tau(u,\sigma)$ is a non-decreasing function about u when σ is fixed.

(2) $\psi_\tau(u,\sigma)$ is a bounded function, and both the first and second derivatives are also bounded.

(3) $\psi_\tau(u,\sigma)$ has a continuous second derivative function about u.

Proof. (1): Since $\psi_\tau(u,\sigma) = \tau - h_\tau(u,\sigma)$, we only need to show that $h_\tau(u,\sigma)$ is a non-increasing function about parameter u when σ is fixed. Denote $m_\tau(u) \equiv h_\tau(u,\sigma)$, then

$$\begin{aligned} m_\tau(u) &= \frac{\partial \ln(\nu_\tau(u,\sigma))}{\partial u} \\ &= \frac{\exp\{u + \frac{1-2\tau}{2}\sigma^2\}\Phi(-\frac{u}{\sigma} + (\tau-1)\sigma)}{\exp\{u + \frac{1-2\tau}{2}\sigma^2\}\Phi(-\frac{u}{\sigma} + (\tau-1)\sigma) + \Phi(\frac{u}{\sigma} - \tau\sigma)} \equiv \frac{p(u)}{q(u)}, \end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Since $m'_\tau(u) = \frac{p'(u)q(u) - p(u)q'(u)}{q^2(u)}$, we only need to show that

$$\begin{aligned} m_1(u) \equiv& p'(u)q(u) - p(u)q'(u) \\ =& \exp\left\{u + \frac{1-2\tau}{2}\sigma^2\right\}\left\{\Phi\left(\frac{u}{\sigma} - \tau\sigma\right)\left[\Phi\left(-\frac{u}{\sigma} + (\tau-1)\sigma\right) - \frac{1}{\sigma}\phi\left(-\frac{u}{\sigma} + (\tau-1)\sigma\right)\right]\right. \\ & \left. - \frac{1}{\sigma}\Phi\left(-\frac{u}{\sigma} + (\tau-1)\sigma\right)\phi\left(\frac{u}{\sigma} - \tau\sigma\right)\right\} \le 0, \quad u \in \mathbb{R}, \end{aligned} \tag{A.1}$$

where $\phi(\cdot)$ is the density probability function of the standard normal distribution. Let x denote $-\frac{u}{\sigma} + (\tau-1)\sigma$, then

$$\begin{aligned} m_2(x) &\equiv m_1(u)/\exp\left\{u + \frac{1-2\tau}{2}\sigma^2\right\} \\ &= \Phi(-x-\sigma)\Phi(x) - \frac{1}{\sigma}\Phi(-x-\sigma)\phi(x) - \frac{1}{\sigma}\Phi(x)\phi(-x-\sigma). \end{aligned}$$

Notice that $m_2(x)$ is a symmetric function about $x = -\frac{\sigma}{2}$ and $\lim_{x\to\pm\infty} m_2(x) = 0$, so (A.1) follows from

$$m'_2(x) \ge 0, \quad x > -\frac{\sigma}{2}. \tag{A.2}$$

The next step is to show (A.2) holds. We only prove (A.2) when $\sigma = 1$. For $0 < \sigma < 1$ and $\sigma > 1$, one can show it similarly.

For $\sigma = 1$, $m_2(x) = \Phi(-x-1)\Phi(x) - \Phi(-x-1)\phi(x) - \Phi(x)\phi(-x-1)$, and the first derivative of $m_2(x)$ is

$$\begin{aligned}
m_2'(x) = &-\phi(-x-1)\Phi(x) + \Phi(-x-1)\phi(x) \\
&- [-\phi(-x-1)\phi(x) + \Phi(-x-1)\phi(x)(-x)] \\
&- [\phi(x)\phi(-x-1) + \Phi(x)\phi(-x-1)(-x-1)] \\
= &-\phi(-x-1)\Phi(x) + \Phi(x)\phi(-x-1)(x+1) \\
&+ \Phi(-x-1)\phi(x) + \Phi(-x-1)\phi(x)x \\
=&\phi(-x-1)\Phi(x)x + \Phi(-x-1)\phi(x)(1+x) \\
=&\phi(x+1)\Phi(x)x + \Phi(-x-1)\phi(x)(1+x),
\end{aligned}$$

where the last equality is by symmetry of $\phi(x)$. Obviously, $m_2'(-\frac{1}{2}) = 0$, $m_2'(0) > 0$, and

$$m_2'(x) \geq 0, \quad x \geq 0.$$

We still need to prove that

$$m_2'(x) \geq 0, \quad -\frac{1}{2} < x < 0. \tag{A.3}$$

For $-\frac{1}{2} < x < 0$, $-x^2 - x + 1 \geq 1$, and the second derivative of $m_2(x)$ is

$$\begin{aligned}
m_2''(x) = &-\phi(x+1)(x+1)\Phi(x)x + \phi(x+1)\left[\phi(x)x + \Phi(x)\right] \\
&+ \phi(-x-1)(-1)\phi(x)(1+x) + \Phi(-x-1)\left[\phi(x)(-x)(1+x) + \phi(x)\right] \\
=&\phi(x+1)\Phi(x)\left[-(x+1)x+1\right] + \phi(x+1)\phi(x)[x-1-x] \\
&+ \Phi(-x-1)\phi(x)[-x(1+x)+1] \\
=&(-x^2-x+1)\left[\phi(x+1)\Phi(x) + \Phi(-x-1)\phi(x)\right] - \phi(x+1)\phi(x) \\
\geq&\phi(x+1)\Phi(x) + \Phi(-x-1)\phi(x) - \phi(-x-1)\phi(x) \equiv m_3(x).
\end{aligned}$$

By $m_2'(-\frac{1}{2}) = 0$ and $m_2'(0) > 0$, (A.3) is implied by

$$m_3(x) \geq 0, -\frac{1}{2} < x < 0. \tag{A.4}$$

We now consider $m_3(x)$ for $-\frac{1}{2} < x < 0$. It is easy to show that

$$\begin{aligned}
m_3(x) =&\frac{1}{\sqrt{2\pi}}e^{-\frac{(x+1)^2}{2}}\Phi(x) + \Phi(-x-1)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} - \phi(x+1)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \\
=&\left\{e^{-\frac{2x+1}{2}}\Phi(x) + \Phi(-x-1) - \phi(x+1)\right\}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.
\end{aligned}$$

The first derivative of $\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right)^{-1} m_3(x)$ is

$$\begin{aligned}\left[\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right)^{-1} m_3(x)\right]' &= \left\{-e^{-\frac{2x+1}{2}}\Phi(x) + e^{-\frac{2x+1}{2}}\phi(x) - \phi(x+1) - \phi(x+1)(-x-1)\right\} c_1 \\ &= \left\{-e^{\frac{2x+1}{2}}\Phi(x) + \phi(x+1)(x+1)\right\} c_1 \\ &= \left\{-e^{\frac{2x+1}{2}}\Phi(x) + \frac{1}{\sqrt{2\pi}}e^{-\frac{(x+1)^2}{2}}(x+1)\right\} c_1 \\ &= \{-\Phi(x) + \phi(x)(x+1)\}\, c_2,\end{aligned}$$

where c_1 and c_2 are positive constant numbers. The second derivative of $c_2^{-1}m_3(x)$ is

$$\begin{aligned}[(c_2)^{-1}m_3(x)]'' &= -\phi(x) + \phi(x)(-x)(x+1) + \phi(x) \\ &= -x(x+1)\phi(x),\end{aligned}$$

Notice that $m_3''(x) \geq 0$ for $-\frac{1}{2} < x < 0$, and $m_3''(0) = 0$, $m_3''(-\frac{1}{2}) > 0$. Therefore, $m_3'(x)$ is an increasing function for $-\frac{1}{2} < x < 0$. Combining the fact that $m_3'(0) < 0$ and $m_3'(-\frac{1}{2}) < 0$, we know that $m_3(x)$ is a decreasing function for $-\frac{1}{2} < x < 0$. By $m_3(0) > 0$ and $m_3(-\frac{1}{2}) > 0$, we can obtain (A.4) holds.

The proof of (2) and (3) follows from the proof of (1). □

Lemma 2. *Under the assumptions of Theorem 1,*

$$\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N}\sum_{t=1}^{T} \psi_\tau(u_{it}, \sigma)\boldsymbol{g}_{it} \right\| = O_p(1). \tag{A.5}$$

Proof. We rewrite (A.5) as follows

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N}\sum_{t=1}^{T} \psi_\tau(u_{it}, \sigma) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi_\tau(u_{it}, \sigma) \right) \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Psi_i.$$

By Proposition 1, Ψ_i has a mean $\mu_\Psi = 0$ and a variance $V_\Psi < \infty$. And by Lindeberg-Lévy central limit theory,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N}\sum_{t=1}^{T} \psi_\tau(u_{it}, \sigma) \xrightarrow{d} \mathcal{N}(0, V_\Psi).$$

Then, by Assumption (A4) and Slutsky's theorem,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N}\sum_{t=1}^{T} \psi_\tau(u_{it}, \sigma)\boldsymbol{g}_{it} \xrightarrow{d} \mathcal{N}\left(0, V_\Psi E(\boldsymbol{g}_{it})E(\boldsymbol{g}_{it}')\right)$$

Therefore,

$$\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \psi_\tau(u_{it}, \sigma) \boldsymbol{g}_{it} \right\| = O_p(1).$$

□

Proof of Theorem 1.

First, we derive the Bahadur representation of the panel quantile regression. We need to verify the assumptions listed in Lemma 1. Define a objective function as follows.

$$Z_{NT}(\Delta) = \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \ln l_\tau(u_{it} - \Delta' b_{NT}^{-1} \boldsymbol{g}_{it}) - \ln l_\tau(u_{it}) \right\},$$

where $b_{NT} = (NT)^{1/2}$, and $l_\tau(u)$ is actually $l_\tau(u, \sigma)$ that is defined in (8). In this proof, however, σ will be omitted for the estimation of $\boldsymbol{\xi}_\tau$ is performed with some fixed σ. Indeed, $Z_{NT}(\Delta)$ is the objective function of the following model:

$$q_\tau(u_{it}) = (\Delta b_{NT}^{-1})'_\tau \boldsymbol{g}_{it},$$

and the QMLE of $(\Delta b_{NT}^{-1})_\tau$ is $(\hat{\boldsymbol{\xi}}_\tau - \boldsymbol{\xi}_\tau)$ by (9). Therefore, we can obtain

$$\left. \frac{\partial Z_{NT}(\Delta)}{\partial \Delta} \right|_{\Delta = \sqrt{NT}(\hat{\boldsymbol{\xi}}_\tau - \boldsymbol{\xi}_\tau)} = 0. \tag{A.6}$$

Using the identity which has a similar form in Knight (1998), we have

$$-\ln l_\tau(u - v) + \ln l_\tau(u) = -v\psi_\tau(u) + \int_0^v \left[h(u - s) - h(u) \right] ds,$$

where $\psi_\tau(u) = -\frac{\partial \ln l_\tau(u)}{\partial u}$. Then,

$$Z_{NT}(\Delta) = \sum_{i=1}^{N} \sum_{t=1}^{T} \Delta' b_{NT}^{-1} \boldsymbol{g}_{it} \psi_\tau(u_{it}) - \sum_{i=1}^{N} \sum_{t=1}^{T} \int_0^{\Delta' b_{NT}^{-1} \boldsymbol{g}_{it}} \left[h(u_{it} - s) - h(u_{it}) \right] ds \tag{A.7}$$

By (A.7), $Z_{NT}(\Delta)$ is derivable. Then we define a new function as follows.

$$\begin{aligned} V_{NT}(\Delta) &\equiv \frac{\partial Z_{NT}(\Delta)}{\partial \Delta} \\ &= \sum_{i=1}^{N} \sum_{t=1}^{T} b_{NT}^{-1} \boldsymbol{g}_{it} \psi_\tau(u_{it}) - \sum_{i=1}^{N} \sum_{t=1}^{T} b_{NT}^{-1} \boldsymbol{g}_{it} \left[h(u_{it} - \Delta' b_{NT}^{-1} \boldsymbol{g}_{it}) - h(u_{it}) \right]. \end{aligned}$$

The next step is to show that $V_{NT}(\Delta)$ satisfies Assumption (i) in Lemma 1. By Proposition 1 that $\psi_\tau(u)$ is a non-dcreasing function, so

$$-\Delta' V_{NT}(\lambda\Delta) = \sum_{i=1}^{N}\sum_{t=1}^{T} \psi_\tau(u_{it} - \lambda\Delta' b_{NT}^{-1}\boldsymbol{g}_{it})(-\Delta' b_{NT}^{-1}\boldsymbol{g}_{it})$$

is a non-decreasing function of λ if $\Delta' b_{NT}^{-1}\boldsymbol{g}_{it} > 0$. Similarly,

$$-\Delta' V_{NT}(\lambda\Delta) = \sum_{i=1}^{N}\sum_{t=1}^{T} \psi_\tau(u_{it} - \lambda\Delta' b_{NT}^{-1}\boldsymbol{g}_{it})(-\Delta' b_{NT}^{-1}\boldsymbol{g}_{it})$$

is a non-decreasing function of λ if $\Delta' b_{NT}^{-1}\boldsymbol{g}_{it} < 0$. Thus,

$$-\Delta' V_{NT}(\lambda\Delta) = \sum_{i=1}^{N}\sum_{t=1}^{T} \psi_\tau(u_{it} - \lambda\Delta' b_{NT}^{-1}\boldsymbol{g}_{it})(-\Delta' b_{NT}^{-1}\boldsymbol{g}_{it})$$

is a non-dereasing function of λ. Therefore, we have for $\lambda > 1$

$$-\Delta' V_{NT}(\lambda\Delta) \geq \Delta' V_{NT}(\Delta),$$

so that Assumption (i) of Lemma 1 is verified.

We still need to prove that $V_{NT}(\Delta)$ satisfies Assumption (ii) in Lemma 1.

$$\begin{aligned} V_{NT}(\Delta) &= \sum_{i=1}^{N}\sum_{t=1}^{T} b_{NT}^{-1}\boldsymbol{g}_{it}\psi_\tau(u_{it}) - \sum_{i=1}^{N}\sum_{t=1}^{T} b_{NT}^{-1}\boldsymbol{g}_{it}\left[h(u_{it} - \Delta' b_{it}^{-1}\boldsymbol{g}_{it}) - h(u_{it})\right] \\ &= \sum_{i=1}^{N}\sum_{t=1}^{T} b_{NT}^{-1}\boldsymbol{g}_{it}\psi_\tau(u_{it}) - \sum_{i=1}^{N}\sum_{t=1}^{T} E(\lambda_{it}|\boldsymbol{g}_{it}) - \sum_{i=1}^{N}\sum_{t=1}^{T}\left[\lambda_{it} - E(\lambda_{it}|\boldsymbol{g}_{it})\right], \\ &= A_{NT} - \sum_{i=1}^{N}\sum_{t=1}^{T} E(\lambda_{it}|\boldsymbol{g}_{it}) - \sum_{i=1}^{N}\sum_{t=1}^{T}\left[\lambda_{it} - E(\lambda_{it}|\boldsymbol{g}_{it})\right], \end{aligned} \tag{A.8}$$

where $A_{NT} = \sum_{i=1}^{N}\sum_{t=1}^{T} b_{NT}^{-1}\boldsymbol{g}_{it}\psi_\tau(u_{it})$ and $\lambda_{it} = b_{NT}^{-1}\boldsymbol{g}_{it}\left[h(u_{it} - \Delta' b_{it}^{-1}\boldsymbol{g}_{it}) - h(u_{it})\right]$. By Taylor expansion,

$$
\begin{aligned}
\sum_{i=1}^{N}\sum_{t=1}^{T} E(\lambda_{it}|\boldsymbol{g}_{it}) &\equiv \sum_{i=1}^{N}\sum_{t=1}^{T} E_{it}(\lambda_{it}) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T} b_{NT}^{-1}\boldsymbol{g}_{it}E_{it}\left(h(u_{it}-\Delta' b_{NT}^{-1}\boldsymbol{g}_{it}) - h(u_{it})\right) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T} b_{NT}^{-1}\boldsymbol{g}_{it}E_{it}\left(-\dot{h}(u_{it})\boldsymbol{g}_{it}' b_{NT}^{-1}\Delta + \frac{1}{2}\ddot{h}(u_{it}^{*})\Delta' b_{NT}^{-1}\boldsymbol{g}_{it}\boldsymbol{g}_{it}' b_{NT}^{-1}\Delta\right) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T} E\left(-\dot{h}(u_{it})|\boldsymbol{g}_{it}\right) b_{NT}^{-1}\boldsymbol{g}_{it}\boldsymbol{g}_{it}' b_{NT}^{-1}\Delta + \sum_{i=1}^{N}\sum_{t=1}^{T} b_{NT}^{-1}\boldsymbol{g}_{it}\frac{1}{2}E\left(\ddot{h}(u_{it}^{*})|\boldsymbol{g}_{it}\right)\Delta' b_{NT}^{-1}\boldsymbol{g}_{it}\boldsymbol{g}_{it}' b_{NT}^{-1}\Delta \\
&\equiv B_1 + B_2,
\end{aligned}
\tag{A.9}
$$

where $E_{it}(\cdot)$ denotes the conditional expectation dependent on $\boldsymbol{g}_{it}$ and u_{it}^{*} is the interior point between u_{it} and $u_{it} + \Delta' b_{NT}^{-1}\Delta\boldsymbol{g}_{it}$. Note that $\dot{h} < 0$ by Proposition 1 (1). We decompose B_1 into two parts as follows.

$$
B_1 = \sum_{i=1}^{N}\sum_{t=1}^{T} E\left(-\dot{h}(u_{it})\right) b_{NT}^{-1}\boldsymbol{g}_{it}\boldsymbol{g}_{it}' b_{NT}^{-1}\Delta + \sum_{i=1}^{N}\sum_{t=1}^{T}\left\{E\left(-\dot{h}(u_{it})|\boldsymbol{g}_{it}\right) - E\left[E\left(-\dot{h}(u_{it})|\boldsymbol{g}_{it}\right)\right]\right\} b_{NT}^{-1}\boldsymbol{g}_{it}\boldsymbol{g}_{it}' b_{NT}^{-1}\Delta.
$$

We rewrite the second part as follows and want to show it is equal to $o_p(1)$.

$$
\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{E\left(-\dot{h}(u_{it})|\boldsymbol{g}_{it}\right) - E\left[E\left(-\dot{h}(u_{it})|\boldsymbol{g}_{it}\right)\right]\right\} b_{NT}^{-1}\boldsymbol{g}_{it}\boldsymbol{g}_{it}' b_{NT}^{-1} \\
=&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{E\left(-\dot{h}(u_{it})|\boldsymbol{g}_{it}\right) - E\left[E\left(-\dot{h}(u_{it})|\boldsymbol{g}_{it}\right)\right]\right\}\boldsymbol{g}_{it}\boldsymbol{g}_{it}' \cdot \frac{1}{\sqrt{NT}}.
\end{aligned}
$$

Because $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{E\left(-\dot{h}(u_{it})|\boldsymbol{g}_{it}\right) - E\left[E\left(-\dot{h}(u_{it})|\boldsymbol{g}_{it}\right)\right]\right\} \xrightarrow{d} \mathcal{N}(0,\Sigma)$ where $\Sigma = Var\left[E\left(-\dot{h}(u_{it})|\boldsymbol{g}_{i}\right.\right.$ by Proposition 1 (2) that $h(u)$ is bounded and Assumption (A4), it follows that

$$
\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{E\left(-\dot{h}(u_{it})|\boldsymbol{g}_{it}\right) - E\left[E\left(-\dot{h}(u_{it})|\boldsymbol{g}_{it}\right)\right]\right\} b_{NT}^{-1}\boldsymbol{g}_{it}\boldsymbol{g}_{it}' b_{NT}^{-1} = O_p(1)\cdot o(1) = o_p(1).
$$

Therefore,

$$
B_1 = E\left[\boldsymbol{g}_{it}\boldsymbol{g}_{it}' E\left(-\dot{h}(u_{it},\sigma)\right)\right]\Delta + o_p(1) \equiv D_{g\dot{h}}(\sigma)\Delta + o_p(1), \tag{A.10}
$$

where $D_{g\dot{h}}(\sigma) = E\left[\boldsymbol{g}_{it}\boldsymbol{g}'_{it}E\left(-\dot{h}(u_{it},\sigma)\right)\right]$. Also,

$$\begin{aligned}\|B_2\| &= \left\|\sum_{i=1}^{N}\sum_{t=1}^{T} b_{NT}^{-1}\boldsymbol{g}_{it}\frac{1}{2}E\left(\ddot{h}(u^*)|\boldsymbol{g}_{it}\right)\Delta' b_{NT}^{-1}\boldsymbol{g}_{it}\boldsymbol{g}'_{it}b_{NT}^{-1}\Delta\right\| \\ &\leq b_{NT}^{-1}\frac{1}{2}\left|E\left(\ddot{h}(u^*)|\boldsymbol{g}_{it}\right)\right|\cdot\sum_{i=1}^{N}\sum_{t=1}^{T} b_{NT}^{-2}\|\boldsymbol{g}_{it}\|^3\|\Delta\|^2 \\ &\xrightarrow{p} o(1)\cdot\frac{1}{2}\left|E\left(\ddot{h}(u^*)|\boldsymbol{g}_{it}\right)\right| E\|\boldsymbol{g}_{it}\|^3\|\Delta\|^2 = 0,\end{aligned}$$

by Assumption (A4), Assumption (ii) in Lemma 1 that $\|\Delta\| \leq M$ and Proposition 1 (2). Thus,

$$B_2 = o_p(1). \tag{A.11}$$

By combining (A.10) and (A.11) , we have

$$\sum_{i=1}^{N}\sum_{t=1}^{T} E(\lambda_{it}|\boldsymbol{g}_{it}) = D_{g\dot{h}}(\sigma) + o_p(1). \tag{A.12}$$

Next, we will verify that

$$\sum_{i=1}^{N}\sum_{t=1}^{T}[\lambda_{it} - E(\lambda_{it}|\boldsymbol{g}_{it})] = o_p(1).$$

For those satisfying $\Delta' b_{NT}^{-1}\boldsymbol{g}_{it} > 0$, $h(u_{it}+\Delta' b_{NT}^{-1}\boldsymbol{g}_{it}) - h(u_{it}) > 0$, and $\left[h(u_{it}+\Delta' b_{NT}^{-1}\boldsymbol{g}_{it}) - h(u_{it})\right] \in (0,1)$ by $h(u) \in (0,1)$, we can show that

$$\begin{aligned}\sum_{i=1}^{N}\sum_{t=1}^{T} E(\lambda_{it}\lambda'_{it}) &= \sum_{i=1}^{N}\sum_{t=1}^{T} E[E(\lambda_{it}\lambda'_{it}|\boldsymbol{g}_{it})] \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} E\left\{E[\left(h(u_{it}+\Delta' b_{NT}^{-1}\boldsymbol{g}_{it}) - h(u_{it})\right)^2\boldsymbol{g}_{it}\boldsymbol{g}'_{it}|\boldsymbol{g}_{it}]\right\} \\ &\leq E\left\{E[\left(h(u_{it}+\Delta' b_{NT}^{-1}\boldsymbol{g}_{it}) - h(u_{it})\right)\boldsymbol{g}_{it}\boldsymbol{g}'_{it}|\boldsymbol{g}_{it}]\right\} \\ &= E\left\{\dot{h}(u^*_{it})\Delta' b_{NT}^{-1}\boldsymbol{g}_{it}\boldsymbol{g}_{it}\boldsymbol{g}'_{it}\right\} \\ &\leq \max_{u\in\mathbb{R}}|\dot{h}(u)|\cdot\frac{1}{\sqrt{NT}}\cdot E(\Delta'\boldsymbol{g}_{it}\boldsymbol{g}_{it}\boldsymbol{g}'_{it}),\end{aligned}$$

where $u^*_{it} \in (u_{it}, u_{it}+\Delta' b_{NT}^{-1}\boldsymbol{g}_{it})$. For those satisfying $\Delta' b_{NT}^{-1}\boldsymbol{g}_{it} < 0$, we can obtain the same result. Thus, for $0 < M < \infty$,

$$\left\|\sum_{i=1}^{N}\sum_{t=1}^{T} E(\lambda_{it}\lambda'_{it})\right\| \leq \max_{u\in\mathbb{R}}|\dot{h}(u)|\cdot\frac{1}{\sqrt{NT}}\cdot M\cdot E\|\boldsymbol{g}_{it}\|^3 = o(1),$$

by Proposition 1 (2) and Assumption (A4).

As a result, we prove that

$$\sum_{i=1}^{N}\sum_{t=1}^{T}E(\lambda_{it}\lambda_{it}')=o(1).$$

Therefore,

$$\begin{aligned}
&Var\left(\sum_{i=1}^{N}\sum_{t=1}^{T}[\lambda_{it}-E(\lambda_{it}|\boldsymbol{g}_{it})]\right)\\
\equiv&Var\left(\sum_{i=1}^{N}\sum_{t=1}^{T}[\lambda_{it}-E_{it}(\lambda_{it})]\right)\\
=&\sum_{i=1}^{N}\sum_{t=1}^{T}Var[\lambda_{it}-E_{it}(\lambda_{it})]+\sum_{i=1}^{N}\sum_{t\neq s}Cov\left[\lambda_{it}-E_{it}(\lambda_{it}),\lambda_{is}-E_{is}(\lambda_{is})\right]\\
\leq&\sum_{i=1}^{N}\sum_{t=1}^{T}Var[\lambda_{it}-E_{it}(\lambda_{it})]+(T-1)\sum_{i=1}^{N}\sum_{t=1}^{T}Var\left[\lambda_{it}-E_{it}(\lambda_{it})\right]\\
=&T\sum_{i=1}^{N}\sum_{t=1}^{T}Var[\lambda_{it}-E_{it}(\lambda_{it})]=T\sum_{i=1}^{N}\sum_{t=1}^{T}E[\lambda_{it}-E_{it}(\lambda_{it})][\lambda_{it}-E_{it}(\lambda_{it})]'\\
\leq&T\sum_{i=1}^{N}\sum_{t=1}^{T}E(\lambda_{it}\lambda_{it}')=o(1).
\end{aligned}$$

Thus by Chebyshev's inequality, we have verified

$$\sum_{i=1}^{N}\sum_{t=1}^{T}[\lambda_{it}-E(\lambda_{it}|\boldsymbol{g}_{it})]=o_p(1). \tag{A.13}$$

By (A.8), (A.12), and (A.13), for $\|\Delta\|\leq M$, $0<M<\infty$,

$$V_{NT}(\Delta)=A_{NT}-D_{g\dot{h}}(\sigma)+o_p(1)$$

Therefore, for $0<M<\infty$,

$$\sup_{\|\Delta\|\leq M}\left\|V_{NT}(\Delta)+D_{g\dot{h}}(\sigma)-A_{NT}\right\|=o_p(1). \tag{A.14}$$

By Lemma 2,

$$\|A_{NT}\|=\left\|\sum_{i=1}^{N}\sum_{t=1}^{T}b_{NT}^{-1}\boldsymbol{g}_{it}\psi_\tau(u_{it})\right\|=O_p(1), \tag{A.15}$$

and by (A.6),

$$\left|V_{NT}\left(\sqrt{NT}(\hat{\boldsymbol{\xi}}_\tau-\boldsymbol{\xi}_\tau)\right)\right|=\left|\frac{\partial Z_{NT}(\Delta)}{\partial\Delta}\bigg|_{\Delta=\sqrt{NT}(\hat{\boldsymbol{\xi}}_\tau-\boldsymbol{\xi}_\tau)}\right|=0=o_p(1). \tag{A.16}$$

Then by (A.14), (A.15) and (A.16), we conclude that Assumption (ii) of Lemma 1 is verified. Thus, we obtain the Bahadur representation

$$\sqrt{NT}(\hat{\boldsymbol{\xi}}_\tau - \boldsymbol{\xi}_\tau) \equiv \hat{\Delta} = D_{g\dot{h}}^{-1}(\sigma) A_{NT} + o_p(1). \tag{A.17}$$

Since $\boldsymbol{\delta}_\tau = \boldsymbol{e}_1' \boldsymbol{\xi}_\tau$, then the Bahadur representation of $\hat{\boldsymbol{\delta}}_\tau$ is

$$\sqrt{NT}(\hat{\boldsymbol{\delta}}_\tau - \boldsymbol{\delta}_\tau) \simeq \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{e}_1' D_{g\dot{h}}^{-1}(\sigma) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi_\tau(u_{it}, \sigma) \boldsymbol{g}_{it} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Gamma_i.$$

It is easy to see the mean of Γ_i is zero for $E[\psi_\tau(u, \sigma)] = 0$ and the variance of Γ_i is given by

$$\begin{aligned}
Var(\Gamma_i) &= \boldsymbol{e}_1' D_{g\dot{h}}^{-1}(\sigma) \frac{1}{T} Var\left(\sum_{t=1}^{T} \psi_\tau(u_{it}, \sigma) \boldsymbol{g}_{it} \right) D_{g\dot{h}}^{-1}(\sigma_1) \boldsymbol{e}_1 \\
&= \boldsymbol{e}_1' D_{g\dot{h}}^{-1}(\sigma) \left\{ Var[\psi_\tau(u_{it}, \sigma) \boldsymbol{g}_{it}] + \sum_{t=2}^{T} \frac{2(T-t+1)}{T} Cov(\psi_\tau(u_{i1}, \sigma) \boldsymbol{g}_{i1}, \psi_\tau(u_{it}, \sigma) \boldsymbol{g}_{it}) \right\} D_{g\dot{h}}^{-1}(\sigma) \boldsymbol{e}_1 \\
&= \boldsymbol{e}_1' D_{g\dot{h}}^{-1}(\sigma) \left\{ E[\psi_\tau(u_{it}, \sigma)]^2 E[\boldsymbol{g}_{it} \boldsymbol{g}_{it}'] + \sum_{t=2}^{T} \frac{2(T-t+1)}{T} E[\psi_\tau(u_{i1}, \sigma) \psi_\tau(u_{it}, \sigma)] E[\boldsymbol{g}_{i1} \boldsymbol{g}_{it}'] \right\} D_{g\dot{h}}^{-1}(\sigma) \boldsymbol{e}_1.
\end{aligned}$$

Finally, since Γ_i is i.i.d. across individual i, we can follow the Lindeberg-Lévy central limit theorem to get the asymptotic normality of $\hat{\boldsymbol{\delta}}_\tau$.