


Article

A New Instrumental-Type Estimator for Quantile Regression Models

Li Tao ¹, Lingnan Tai ², Manling Qian ^{3,*} and Maozai Tian ^{4,5} ¹ School of Information, Beijing Wuzi University, Beijing 101149, China; taoli@bwu.edu.cn² School of Economics and Management, The Open University of China, Beijing 100039, China; tailn@ouchn.edu.cn³ School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia⁴ Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China; mztian@ruc.edu.cn⁵ School of Statistics and Information, Xinjiang University of Finance and Economics, Urumqi 830026, China

* Correspondence: manling@student.unimelb.edu.au

Abstract: This paper proposes a new instrumental-type estimator of quantile regression models for panel data with fixed effects. The estimator is built upon the minimum distance, which is defined as the weighted average of the conventional individual instrumental variable quantile regression slope estimators. The weights assigned to each estimator are determined by the inverses of their corresponding individual variance–covariance matrices. The implementation of the estimation has many advantages in terms of computational efforts and simplifies the asymptotic distribution. Furthermore, the paper shows consistency and asymptotic normality for sequential and simultaneous asymptotics. Additionally, it presents an empirical application that investigates the income elasticity of health expenditures.

Keywords: quantile regression; instrumental variables; minimum distance estimator; panel data

MSC: 62-04



Citation: Tao, L.; Tai, L.; Qian, M.; Tian, M. A New Instrumental-Type Estimator for Quantile Regression Models. *Mathematics* **2023**, *11*, 3412. <https://doi.org/10.3390/math11153412>

Academic Editors: Laleh Tafakori and Marco Bee

Received: 22 June 2023

Revised: 2 August 2023

Accepted: 3 August 2023

Published: 4 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Panel data not only captures the individual heterogeneity inherent in cross-sectional data but also offers valuable insights into the dynamics of time series data. Quantile regression effectively captures the relationship between the independent variable X and the dependent variable Y , enabling a comprehensive understanding of how covariates systematically influence the location, scale, and shape of the conditional distribution of the response. Moreover, distinguishing it from the ordinary least squares method, quantile regression does not necessitate any assumptions about the overall distribution.

The quantile regression model applied to panel data has the capability to comprehensively describe the conditional distribution of the response variable and effectively manage variability. When endogeneity is disregarded, three estimation approaches can be employed for quantile regression in panel data with fixed effects: penalty estimation, two-step estimation, and minimum distance estimation.

The penalty estimation primarily involved the addition of a penalty term to the objective function. Koenker [1] introduced a comprehensive methodology for estimating quantile regression models in longitudinal data through the utilization of l_1 regularization techniques. It was highlighted that shrinkage could offer advantages in managing the variability arising from the estimation of fixed effects parameters. The selection of the l_1 penalty maintained the linear programming structure and preserved the sparsity of the resulting design matrix. Nevertheless, the mentioned article did not provide a method for selecting the penalty parameter. Lamarche [2] theoretically demonstrated that there

was an optimal penalty parameter for the penalized quantile regression in panel data with fixed effects. As mentioned earlier, the utilization of l_1 shrinkage provides several statistical and computational advantages. However, when the number of panel subjects, represented by N , is large, the calculation of this kind of penalty estimation becomes rather complicated. Building upon Koenker [1], Gu and Volgushev [3] proposed a linear quantile regression method which allowed the researcher to learn the particular group structure of the latent effects together with common parameters of interest in the model. Galvao et al. [4] formalized the properties of bootstrap inference methods for quantile regression panel data models with fixed effects. They investigated a random-weighted bootstrap and demonstrated that it could be used to construct confidence intervals and perform inference for the parameters of interest.

The primary concept behind two-step estimation is to remove the fixed effects in the initial step, followed by employing simple quantile regression on the transformed data in the second step. Canay [5] introduced a two-step estimator for quantile regression model with fixed effects. In the first step, compute $\hat{\eta}_i = \frac{1}{N} \sum_{t=1}^T (y_{it} - \mathbf{x}'_{it} \hat{\beta})$, where N is the number of panel subjects and T is the number of observation times, y_{it} represents the response variable for subject i at time t , \mathbf{x}_{it} denotes a vector of observable variables, β represents the parameter of interest, and $\hat{\beta}$ stands for a \sqrt{NT} -consistent estimator of β . Additionally, η_i denotes the individual effect. Estimate β by a quantile regression of $\hat{y}_{it} = y_{it} - \hat{\eta}_i$ on \mathbf{x}_{it} in the second step. Asymptotic properties of the two-step estimator were presented in the paper. It is worth noting that the two-step estimation method eliminates the fixed effects in the initial step. This approach leads to a significant reduction in the number of estimated parameters in quantile regression and avoids the need to select penalty parameters. However, compared to y_{it} , the new variable \hat{y}_{it} has changed the meaning of the dependent variables in the original regression model, making the discussion of the large sample properties of the estimators more complex. Chen and Huo [6] revisited the estimator proposed by Canay [5] and pointed out that the bias of the estimator due to the estimation of fixed effects was erroneously omitted in his asymptotic analysis. They proposed a two-step estimation method based on smoothed quantile regression that was easy to implement. At the same time, this paper rigorously derived the asymptotic bias of the estimation method and proposed a correct statistical inference method based on bias correction.

Galvao and Wang [7] developed a novel minimum distance quantile regression (MD-QR) estimator for panel data quantile regression model. The MD-QR estimator is defined as the weighted average of the individual quantile regression slope estimators, where the weights are determined by the inverses of their corresponding individual variance-covariance matrices. The proposed estimator exhibits efficiency within the class of minimum distance estimators. Moreover, it exhibits fast computational performance, particularly for large cross-sections. However, it should be noted that based on the definition of MD-QR, the model cannot adequately accommodate time-invariant independent variables. Furthermore, the MD-QR estimator does not account for endogenous issues. Galvao et al. [8] proved that unbiased asymptotic normality of both the fixed effects quantile regression (FE-QR) and MD-QR estimators under conditions on N , T that were very close. This result represented a significant advancement in the existing theory and showed that quantile regression was applicable to the same type of panel data (in terms of N , T) as other nonlinear models.

Chernozhukov and Hansen [9–11] extended the results of Koenker and Bassett [12] for the quantile regression (QR) model with all exogenous variables to a model with endogeneity when instruments were available. They proposed an instrumental variable quantile regression (IVQR) estimator that appropriately modified the conventional quantile regression. In this paper, considering the endogenous problems we extend the results of Galvao and Wang [7] and Chernozhukov and Hansen [9–11], combining the minimum distance estimator and instrumental variable quantile regression.

With considering the endogeneity, there has been growing work at the three types of estimation. Harding and Lamarche [13] gave the instrumental variable quantile regression with fixed effects (IVFEQR) estimator to overcome the endogenous problem, where they allowed the endogenous variables to be correlated with unobserved factors affecting the response variables. Galvao and Montes-Rojas [14] also developed an instrumental variables estimator for quantile regression in panel data with fixed effects aiming to mitigate the bias caused by the presence of measurement error (endogeneity) in the model. They showed IVFEQR estimator and proposed its asymptotic properties. Galvao and Montes-Rojas [15] proposed penalized quantile regression for dynamic panel data, utilizing instrumental variables to address the endogeneity issue. Galvao [16] adapted instrumental variable approach to investigate a dynamic panel model. Harding and Lamarche [17] further explored the static fixed effect panel data model using Huasman–Taylor instrumental variables. Zhang et al. [18] studied a penalized instrumental variables quantile regression for spatial panel model with fixed effects. They shrank the individual fixed effects to a common value with a tuning parameter, aiming to control additional variability and enhance the accuracy of parameter estimation. Besstremyannaya and Golovan [19] reviewed various methods for estimating longitudinal models for quantile regression, they pointed out that method of smoothed quantile regression could be considered as a way to mitigate the asymptotic bias of the estimator in short panels. Furthermore, there has been a growing body of research focusing on quantile regression for panel data, see, e.g., [20–24].

In this paper, we propose a minimum distance instrumental variable quantile regression (MD-IVQR) estimator for panel data with fixed effects to resolve biased parameter estimation caused by endogenous variables and to simplify cumbersome computation caused by large N and T . The proposed estimator expands upon the minimum distance quantile regression estimator and is defined as a weighted average of the individual instrumental variable quantile regression slope estimators. The weights are determined by taking the inverses of the corresponding individual variance–covariance matrices. As a result, MD-IVQR can be constructed from a series of classical instrumental variable quantile regressions. MD-IVQR estimation splits the data into smaller parts and estimates each of them individually instead of solving a single larger optimization problem, making the approach computationally simple to implement. We also study the asymptotic properties of the MD-IVQR estimator. The consistency results for the MD-IVQR estimator are presented in Theorems 1 and 2, while the limiting distributions of MD-IVQR estimator under different assumptions are given in Theorems 3 and 4. The validity of the proposed approach is demonstrated through Monte Carlo simulations conducted under various parameter sets. We compare the bias and root mean squared error (RMSE) of the proposed estimator with those of the IVFEQR estimator developed by Harding and Lamarche [13] as well as the MD-QR estimator proposed by Galvao and Wang [7]. We also compare the computing time for the IVFEQR estimator and MD-IVQR estimator.

The remainder of the paper is structured as follows. Section 2 introduces the MD-IVQR estimator for panel data models with fixed effects. Section 3 delves into the asymptotic behavior of the MD-IVQR estimator. Section 4 provides a detailed description of the Monte Carlo experiment. In Section 5, we apply the proposed method to investigate the income elasticity of health expenditures using panel data from 166 countries between 1995 and 2016. The results of this application are presented to illustrate the effectiveness of the proposed approach. Finally, Section 6 summarizes the findings and conclusions of this study.

2. Model and Methods

2.1. Basic Model

This paper considers the following model,

$$y_{it} = \mathbf{d}'_{it}\boldsymbol{\alpha} + \mathbf{x}'_{it}\boldsymbol{\beta} + \eta_i + u_{it}, \quad i = 1, \dots, N; t = 1, \dots, T \quad (1)$$

where N is the number of panel subjects and T is the number of observation times, y_{it} is the response variable for subject i at time t , \mathbf{d}_{it} is a vector of endogenous variables

that is correlated with unobserved factors affecting the response variable, \mathbf{x}_{it} is a vector of exogenous variables, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ represent the parameters of interest, and η_i denotes the individual effect and u_{it} is the error term.

It is convenient to write model (1) in matrix form as,

$$\mathbf{y} = \mathbf{d}\boldsymbol{\alpha} + \mathbf{x}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\eta} + \mathbf{u}, \tag{2}$$

where $\mathbf{y} = (y_{it})$ is a $NT \times 1$ matrix, $\mathbf{d} = (\mathbf{d}_{it})$ is a $NT \times \dim(\boldsymbol{\alpha})$ matrix, $\mathbf{x} = (\mathbf{x}_{it})$ is a $NT \times \dim(\boldsymbol{\beta})$ matrix, $\mathbf{Z} = I_N \otimes l_T$, l_T is a $T \times 1$ vector of ones, I_N is the identity matrix of order N , $\boldsymbol{\eta} = (\eta_1, \dots, \eta_N)'$ is the $N \times 1$ vector of individual specific effects or intercepts, and $\mathbf{u} = (u_{it})$ is a $NT \times 1$ matrix. Note that \mathbf{Z} represents an incidence matrix that identifies the N distinct individuals in the sample.

We assume that the τ -th quantile of the error u_{it} is equal to zero. We consider the following model for the τ -th conditional quantile functions of y_{it} ,

$$Q_{y_{it}}(\tau | \mathbf{d}_{it}, \mathbf{x}_{it}, \mathbf{Z}_i) = \mathbf{d}'_{it}\boldsymbol{\alpha}(\tau) + \mathbf{x}'_{it}\boldsymbol{\beta}(\tau) + \mathbf{Z}'_i\boldsymbol{\eta}(\tau), \tag{3}$$

where \mathbf{Z}_i is the i -th column element of matrix \mathbf{Z} , used to identify the individual effect η_i .

2.2. IVFEQR Estimator

Note that the endogenous variable \mathbf{d}_{it} is correlated with the random error item u_{it} , this situation will cause biased estimation. Therefore, researchers consider using instrumental variables to reduce bias, see Harding and Lamarche [13] and Galvao and Montes-Rojas [14].

The IVFEQR estimator proposed by Harding and Lamarche [13] (also referred to in Galvao and Montes-Rojas [14]) is defined as follows:

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha} \in \mathcal{A}} \{ \hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha})' \mathbf{A} \hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}) \},$$

for a given positive definite matrix \mathbf{A} and $(\hat{\boldsymbol{\eta}}(\boldsymbol{\alpha}), \hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}), \hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}))$ solves

$$\min_{\boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\gamma}} \sum_{t=1}^T \sum_{i=1}^N \rho_{\tau}(y_{it} - \mathbf{d}'_{it}\boldsymbol{\alpha} - \mathbf{x}'_{it}\boldsymbol{\beta} - \mathbf{z}'_{it}\boldsymbol{\eta} - \boldsymbol{\omega}'_{it}\boldsymbol{\gamma}), \tag{4}$$

where \mathcal{A} denotes compact parameter set, $\boldsymbol{\omega}_{it}$ is a vector of the available instrumental variables, $\boldsymbol{\gamma}$ is the coefficient of the instrumental variables, and $\rho_{\tau}(u) = u(\tau - I(u < 0))$ is the check function and $I(\cdot)$ is the indicator function (see, e.g., [12]). The instrument $\boldsymbol{\omega}_{it}$ needs to satisfy the following two conditions: (i) instruments $\boldsymbol{\omega}_{it}$ can impact \mathbf{d}_{it} , and $\dim(\boldsymbol{\omega}_{it}) \geq \dim(\mathbf{d}_{it})$; (ii) $\boldsymbol{\omega}_{it}$ is independent of the random error.

2.3. MD-IVQR Estimator

Because of the endogeneity, we cannot eliminate the fixed effects by transforming \mathbf{x}_{it} , \mathbf{d}_{it} and y_{it} to deviations from individual means to reduce the estimated parameters. That is, for instrumental variable quantile regression, we are required to deal directly with the full problem, which means when N , T and p (the dimension of $\dim(\boldsymbol{\omega}) + \dim(\boldsymbol{\alpha})$) are large, there are a large number of parameters that need to be estimated. Moreover, given a suitable set of values $\{\boldsymbol{\alpha}_j, j = 1, \dots, J\}$, the estimators obtained through Formula (4) essentially correspond to the FE-QR estimators ([1]). In other words, in order to obtain IVFEQR estimators, we need to calculate a series of FE-QR estimators. However, FE-QR estimator involves optimization with large number of parameters to be estimated, making the problem computationally cumbersome, and often intractable. Motivated by the practical challenges of implementing FE-QR, Galvao and Wang [7] presented an efficient minimum distance quantile regression estimator for panels with fixed effects which was simple to implement.

Inspired by Galvao and Wang [7], we propose a minimum distance instrumental variable quantile regression estimator for model (1) to reduce the bias caused by the

presence of endogeneity and to simplify cumbersome computation. The proposed estimator is a weighted combination of the IVQR estimators that can be constructed from a series of conventional instrumental variable quantile regressions. Thus, the estimation approach is computationally convenient and simple to implement in many typical applications.

Let $\theta = (\alpha', \beta')'$, we consider a minimum distance instrumental variable quantile regression (MD-IVQR) estimator, denoted as $\hat{\theta}_{MD-IVQR}$, which is defined as follows:

$$\hat{\theta}_{MD-IVQR} = \left(\sum_{i=1}^N V_i^{-1} \right)^{-1} \sum_{i=1}^N V_i^{-1} \hat{\theta}_i, \tag{5}$$

where $\hat{\theta}_i = (\hat{\alpha}'_i, \hat{\beta}'_i)'$ is the IVQR estimator of slope coefficient for each individual using the time-series data, and V_i denotes the associated variance–covariance matrix of $\hat{\theta}_i$ for each individual. As we can see, the MD-IVQR estimator is defined as the weighted average of the conventional IVQR slope estimators, with weights given by the inverses of the corresponding individual variance-covariance matrices.

Specifically, the IVQR estimator $\hat{\theta}_i$ is defined as follows. Define

$$\hat{\alpha}_i = \arg \min_{\alpha \in \mathcal{A}} \{ \hat{\gamma}_i(\alpha)' \mathbf{A} \hat{\gamma}_i(\alpha) \},$$

where \mathbf{A} is any uniformly positive definite matrix, $\hat{\beta}_i(\alpha)$, $\hat{\gamma}_i(\alpha)$ and $\hat{\eta}_i(\alpha)$ represent the estimated coefficients of the covariates \mathbf{d}_{it} and \mathbf{x}_{it} , as well as the estimated values of individual fixed effect for each individual given α , with

$$(\hat{\beta}_i(\alpha), \hat{\eta}_i(\alpha), \hat{\gamma}_i(\alpha)) := \arg \min_{\beta, \eta_i, \gamma} Q_{iT}(\tau, \alpha, \beta, \eta_i, \gamma),$$

$$Q_{iT}(\tau, \alpha, \beta, \eta_i, \gamma) = \frac{1}{T} \sum_{t=1}^T \rho_{\tau}(y_{it} - \mathbf{d}'_{it}\alpha - \mathbf{x}'_{it}\beta - \eta_i - \omega'_{it}\gamma),$$

ω_{it} is a vector of the available instrument instrumental variables, which is related to \mathbf{d}_{it} but dependent on u_{it} .

The IVQR estimator is given by

$$\hat{\theta}_i := (\hat{\alpha}'_i, \hat{\beta}'_i)' := (\hat{\alpha}'_i, \hat{\beta}_i(\hat{\alpha}_i))'. \tag{6}$$

Moreover, obtaining the IVQR estimator involves the following three steps.

Step 1: For a given quantile τ , it is necessary to define a suitable set of values $\{\alpha_j, j = 1, \dots, J\}$. This involves performing the τ -quantile regression of $y_{it} - \mathbf{d}'_{it}\alpha_j$ on $(\mathbf{x}_{it}, \omega_{it})$ to obtain the ordinary QR estimators of β, η_i, γ :

$$(\hat{\beta}_i(\alpha), \hat{\eta}_i(\alpha), \hat{\gamma}_i(\alpha)) := \arg \min_{\beta, \eta_i, \gamma} Q_{iT}(\tau, \alpha, \beta, \eta_i, \gamma). \tag{7}$$

Step 2: Choose $\hat{\alpha}$ from the set of values $\{\alpha_j, j = 1, \dots, J\}$ that minimizes a weighted distance function defined on γ , aiming to bring the value closest to zero:

$$\hat{\alpha}_i = \arg \min_{\alpha \in \mathcal{A}} \{ \hat{\gamma}_i(\alpha)' \hat{A}(\alpha) \hat{\gamma}_i(\alpha) \},$$

where $\hat{A}(\alpha)$ is defined in Chernozhukov and Hansen [11].

Step 3: Subsequently, the estimation of β , denoted as $\hat{\beta}_i(\hat{\alpha}_i)$, can be obtained.

However, in applications, the estimator $\hat{\theta}_{MD-IVQR}$, defined in (5), is infeasible unless V_i is known for every individual. Thus, it is suggested to use the corresponding consistent estimator \hat{V}_i to replace the unknown V_i . Then, the MD-IVQR estimator is given by

$$\hat{\theta}_{MD-IVQR} = \left(\sum_{i=1}^N \hat{V}_i^{-1} \right)^{-1} \sum_{i=1}^N \hat{V}_i^{-1} \hat{\theta}_i. \tag{8}$$

Actually, the proposed MD-IVQR estimator is a product of a simple optimization problem. Under the restrictions,

$$\zeta - l_N \otimes \theta = 0,$$

where $\zeta = (\theta'_1, \dots, \theta'_N)$, θ_i is a $(\dim(\alpha) + \dim(\beta)) \times 1$ vector containing the slope coefficients for each individual i , l_N denotes an N -vector of ones, and θ is $(\dim(\alpha) + \dim(\beta)) \times 1$ vector of parameters of interest form (3). The restriction states that all the slope coefficients from different individuals are the same. The estimated value of ζ is denoted as $\hat{\zeta}$, where $\hat{\zeta} = (\hat{\theta}'_1, \dots, \hat{\theta}'_N)'$ and the definition of $\hat{\theta}_i$ can be found in (6). The restriction $\hat{\zeta} - l_N \otimes \theta = 0$ cannot be exactly satisfied, so we need to obtain its least-square solution. Thus, we consider minimizing the following quadratic optimization problem

$$\min_{\theta} (\hat{\zeta} - l_N \otimes \theta)^T U (\hat{\zeta} - l_N \otimes \theta),$$

where U is a positive definite matrix. Under independence across individuals, we can simplify U by diagonalizing it, i.e., $U = \text{diag}(U_1, \dots, U_N)$. Thus, the above optimization can be written as

$$\min_{\theta} \sum_{i=1}^N (\hat{\theta}_i - \theta)^T U_i (\hat{\theta}_i - \theta). \tag{9}$$

The Equation (9) can be solved using the following expression:

$$\hat{\theta} = \left(\sum_{i=1}^N U_i \right)^{-1} \sum_{i=1}^N U_i \hat{\theta}_i, \tag{10}$$

This solution is analogous to the MD-IVQR estimator (8), where the MD-IVQR estimator replaces U_i with the inverse of the estimated variance-covariance matrix of the individual regression parameter, denoted as \hat{V}_i^{-1} .

In addition, when U_i takes different values, one can obtain the class of minimum distance estimators, denoted by \mathcal{M} . As mentioned in Galvao and Wang [7], the optimal weights among all the estimators in \mathcal{M} are the inverse of the asymptotic variance-covariance matrices of the slope regression quantiles. This means that the estimator defined in (8) is the most efficient estimator in \mathcal{M} . In other words, $\hat{\theta}_{MD-IVQR}$ has the smallest asymptotic variance-covariance matrix among all the estimators in \mathcal{M} . For the proof of the inverse of the covariance matrices being optimal weights, see Rao [25], Serfling [26], and Hsiao [27].

3. Asymptotic Theory

Now we briefly discuss the asymptotic properties of the MD-IVQR estimator. The existence of the parameter η_i will raise some new issues for the asymptotic analysis of estimator as N tends to infinity. We investigate the asymptotic properties of the MD-IVQR estimator when both T and N tend towards infinity, either sequentially or simultaneously. The sequential asymptotics, denoted by $(T, N)_{seq} \rightarrow \infty$, is defined by taking T to infinity first, followed by $N \rightarrow \infty$. The simultaneous asymptotics, denoted by $(T, N) \rightarrow \infty$, means that T and N tend to infinity at the same time. In order to establish the asymptotic properties of the MD-IVQR estimator, we impose the following regularity conditions.

(A1) $(y_{it}, \mathbf{d}_{it}, \mathbf{x}_{it})$ is independent across i , and i.i.d. within each i .

(A2) For all τ , $(\alpha(\tau), \beta(\tau)) \in \text{int} \mathcal{A} \times \mathcal{B}$, and $\mathcal{A} \times \mathcal{B}$ is compact and convex. Let $\theta(\tau) = (\alpha'(\tau), \beta'(\tau))'$, and $k = \dim(\theta) = \dim(\alpha) + \dim(\beta)$.

(A3) There is a constant M such that $\max_{1 \leq i \leq N, 1 \leq t \leq T} \|\mathbf{d}_{it}\| < M$, $\max_{1 \leq i \leq N, 1 \leq t \leq T} \|\mathbf{x}_{it}\| < M$, $\max_{1 \leq i \leq N, 1 \leq t \leq T} \|\omega_{it}\| < M$.

(A4) For each i , let $Y_i = (y_{i1}, \dots, y_{iT})$ be a T -vector, $D_i = (\mathbf{d}_{i1}, \dots, \mathbf{d}_{iT})$ be a $\dim(\alpha) \times T$ matrix, $X_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})$ be a $\dim(\beta) \times T$ matrix, $W_i = (\omega_{i1}, \dots, \omega_{iT})$ be a $\dim(\gamma) \times T$ matrix and \mathbf{e}_T be a $T \times 1$ vector of ones. Denote $\tilde{\mathbf{x}}_i = [X_i', \mathbf{e}_T']'$, $\tilde{\beta}_i = (\beta', \eta_i)'$, $\tilde{\mathbf{x}}_i = [W_i', \tilde{\mathbf{x}}_i']'$. For each i , $\partial E[I(Y_i < D_i' \alpha + \tilde{\mathbf{x}}_i' \tilde{\beta}_i + W_i' \gamma) \tilde{\mathbf{x}}_i(\tau)] / \partial(\tilde{\beta}_i', \gamma')$ has full rank at each α in \mathcal{A} .

(A5) $\partial E[I(Y_i < D_i' \alpha + \tilde{\mathbf{x}}_i' \tilde{\beta}_i) \tilde{\mathbf{x}}_i'(\tau)] / \partial(\alpha', \tilde{\beta}_i')$ has full rank at $(\alpha'(\tau), \beta'(\tau), \eta_i(\tau))'$.

(A6) For each i , the function $(\alpha, \beta, \eta_i) \mapsto E[\tau - I(Y_i < \alpha D_i + X_i \beta + \eta_i) \tilde{\mathbf{x}}_i'(\tau)]$ is one-to-one over $\mathcal{A} \times \mathcal{B} \times \mathcal{N}$.

(A7) Let $F_i(u|D, X, W)$ be the conditional distribution function of $u_{it} = y_{it} - \mathbf{d}_{it}' \alpha - \mathbf{x}_{it}' \beta - \eta_i$ given $\mathbf{d}_{it} = D$, $\mathbf{x}_{it} = X$, $\omega_{it} = W$ and have conditional density $f_i(u|D, X, W)$. The conditional density $f_i(u|\mathbf{d}_{it}, \mathbf{x}_{it}, \omega_{it})$ is continuously differentiable. There exist $0 < C_L < C_U < \infty$ such that $f_i(u|\mathbf{d}_{it}, \mathbf{x}_{it}, \omega_{it}) \leq C_U$ and $f_i(0|\mathbf{d}_{it}, \mathbf{x}_{it}, \omega_{it}) \geq C_L$ for $i \geq 1$; there exists $C_f > 0$ such that $|f_i^{(1)}(u|\mathbf{d}_{it}, \mathbf{x}_{it}, \omega_{it})| \leq C_f$.

(A8) Given $\alpha \in \mathcal{A}$, $\vartheta_i(\alpha) = (\eta_i(\alpha), \beta'(\alpha), \gamma'(\alpha))' := \text{arginf} E Q_{iT}(\tau, \alpha, \beta, \eta_i, \gamma)$. For each $\delta > 0$,

$$\epsilon_\delta = \inf_{i \geq 1} \inf_{\|\vartheta_i(\alpha)\| = \delta} E \left[\int_0^{\tilde{\mathbf{x}}_i' \vartheta_i(\alpha)} \{F_i(s|x_{it}, \omega_{it}) - \tau\} ds \right] > 0.$$

(A9) Define $q_i = (\alpha', \beta', \eta_i, \gamma)'$, $\tilde{\mathbf{x}}_{it}^* = (\mathbf{d}_{it}', \mathbf{x}_{it}', 1, \omega_{it}')'$, there exists $\delta_\omega > 0$ such that $\min_{1 \leq i \leq N} \text{mineig}(\tilde{\omega}_i) \geq \delta_\omega$, where $\tilde{\omega}_i = E(\tilde{\mathbf{x}}_{it}^* \tilde{\mathbf{x}}_{it}^{*'})$

Conditions (A1) is commonly seen in the quantile regression literature. Condition (A2) restricts the compactness on the parameter space of θ . Such a condition is needed since the objective function is not convex in θ . Condition (A3) requires the restriction boundary conditions of the variables. Condition (A3) also guarantees that $\tilde{\omega}_i$ exists and has a uniform bound. Condition (A4) is sufficient for the estimates $(\beta(\alpha, \tau), \eta_i(\alpha, \tau), \gamma(\alpha, \tau))$ to be asymptotically normal. Condition (A5) requires that each matrix has full rank. Condition (A6) imposes that global identifiability must hold. Condition (A5) and Condition (A6) are required to carry out the direct inference but are not required in the dual approach, as discussed in Chernozhukov and Hansen [10]. Condition (A7) imposes the smoothness and boundedness of the conditional density and its derivatives. Condition (A8) is important for the parameters' identification. Condition (A9) assures that $\tilde{\omega}_i^{-1}$ is bounded uniformly for each i .

In applications, it is necessary to estimate the variance–covariance matrices since they are typically unknown. Since the estimation of V_i depends on the conditional densities and the conditional densities, which are also unknown, we study the kernel estimation of the V_i . Let $\{h_T\}$ and $\{h_N\}$ denote sequences of positive numbers (bandwidths). When T and N tend to infinity sequentially, we propose the following condition.

(A10) $\hat{V}_i = V_i + o_p(1)$ for some $h_T \downarrow 0$ uniformly across i and $Th_T^2 \rightarrow \infty$ as $T \rightarrow \infty$, h_T is a bandwidth. Assume that $(N^{-1} \sum_{i=1}^N V_i^{-1})^{-1} \rightarrow V$, where $V := \lim_{N \rightarrow \infty} (N^{-1} \sum_{i=1}^N V_i^{-1})^{-1}$ exists and is nonsingular.

(A10') $\hat{V}_i = V_i + O_p(T^{-1/2} h_N^{-1/2})$ for some $h_N \downarrow 0$ uniformly across i and $\lim_{N \rightarrow \infty} \frac{N}{Th_N} = 0$ as $N \rightarrow \infty$, h_N is a bandwidth. Assume that $(N^{-1} \sum_{i=1}^N V_i^{-1})^{-1} \rightarrow V$, where $V := \lim_{N \rightarrow \infty} (N^{-1} \sum_{i=1}^N V_i^{-1})^{-1}$ exists and is nonsingular.

Later in the paper, we will discuss examples that satisfy both conditions (A10) and (A10').

First, we show the consistency results for the MD-IVQR estimators in Theorems 1 and 2.

Theorem 1. Under conditions (A1)–(A8) and (A10),

$$\widehat{\boldsymbol{\theta}}_{MD-IVQR} \xrightarrow{p} \boldsymbol{\theta},$$

as $(T, N)_{seq} \rightarrow \infty$.

Theorem 2. Under conditions (A1)–(A8) and (A10'),

$$\widehat{\boldsymbol{\theta}}_{MD-IVQR} \xrightarrow{p} \boldsymbol{\theta},$$

as $(T, N) \rightarrow \infty$ and $\frac{\log N}{T} \rightarrow 0$.

The proofs of Theorems 1 and 2 can be found in Appendices A.1 and A.2, respectively.

Next, we give the limiting distributions of MD-IVQR estimator under sequential and simultaneous limits in Theorems 3 and 4. For the sequential limits, denoted by $(T, N)_{seq} \rightarrow \infty$, we let T diverge to infinity first, and then $N \rightarrow \infty$. For the simultaneous limits, we denote $(T, N)_{sim} \rightarrow \infty$ when $(T, N) \rightarrow \infty$ and $\frac{N^2(\log N)}{T} |\log \frac{(\log N)^{0.5}}{T^{0.5}}|^2 \rightarrow 0$.

Theorem 3. Under conditions (A1)–(A10), as $(T, N)_{seq} \rightarrow \infty$,

$$\sqrt{NT}(\widehat{\boldsymbol{\theta}}_{MD-IVQR} - \boldsymbol{\theta}) \xrightarrow{d} N(0, V),$$

where $V = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N V_i^{-1}$, $V_i = \Xi_1 \tilde{V}_i \Xi_1^\top$, $\Xi_1 = [I_{k \times k} | \mathbf{0}_{k \times 1}]$, $\tilde{V}_i = (K'_i, L'_i)' S_i (K'_i, L'_i)$, $\epsilon_i = Y_i - D'_i \boldsymbol{\alpha}(\tau) - \tilde{\mathbf{x}}'_i \tilde{\boldsymbol{\beta}}_i(\tau)$, $S_i = \tau(1 - \tau) E[\tilde{\mathbf{x}}_i \tilde{\mathbf{x}}'_i]$, $K_i = (J'_{i,\boldsymbol{\alpha}} H_i J_{i,\boldsymbol{\alpha}})^{-1} J'_{i,\boldsymbol{\alpha}} H_i$, $H_i = \bar{J}'_{i,\boldsymbol{\gamma}} A[\boldsymbol{\alpha}(\tau)] \bar{J}_{i,\boldsymbol{\gamma}}$, $L_i = \bar{J}'_{i,\tilde{\boldsymbol{\beta}}_i} M_i$, $M_i = I - J_{i,\boldsymbol{\alpha}} K_i$, $J_{i,\boldsymbol{\alpha}} = E[f_i(0 | \tilde{\mathbf{x}}_i, W_i, D_i) \tilde{\mathbf{x}}_i D'_i]$, and $[\bar{J}'_{i,\tilde{\boldsymbol{\beta}}_i}, \bar{J}'_{i,\boldsymbol{\gamma}}]$ is a partition of $J_{i,\psi} := (E[f_i(0 | \tilde{\mathbf{x}}_i, W_i) \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}'_i])^{-1}$ such that $\bar{J}'_{i,\tilde{\boldsymbol{\beta}}_i}$ is a $\dim(\tilde{\boldsymbol{\beta}}_i) \times \dim(\tilde{\boldsymbol{\beta}}_i, \boldsymbol{\gamma})$ matrix and $\bar{J}'_{i,\boldsymbol{\gamma}}$ is a $\dim(\boldsymbol{\gamma}) \times \dim(\tilde{\boldsymbol{\beta}}_i, \boldsymbol{\gamma})$ matrix.

Theorem 4. Under conditions (A1)–(A9) and (A10'), as $(T, N)_{sim} \rightarrow \infty$,

$$\sqrt{NT}(\widehat{\boldsymbol{\theta}}_{MD-IVQR} - \boldsymbol{\theta}) \xrightarrow{d} N(0, V),$$

where V is defined in Theorem 3.

The proofs of Theorems 3 and 4 can be found in Appendices A.3 and A.4, respectively.

Here are some remarks of Theorems 3 and 4. First, when $\dim(\boldsymbol{\gamma}) = \dim(\boldsymbol{\alpha})$, the choice of \mathbf{A} does not affect asymptotic variance, and V_i has the simple vision

$$V_i = \Xi_1 J_{i,\boldsymbol{\theta}}^{-1} S_i (J'_{i,\boldsymbol{\theta}})^{-1} \Xi_1^\top,$$

where S_i is defined above and $J'_{i,\boldsymbol{\theta}} = E[f_i(0 | \tilde{\mathbf{x}}_i, W_i, D_i) \tilde{\mathbf{x}}_i [D'_i, \tilde{\mathbf{x}}'_i]]$. This is particularly convenient since the variance formula V_i becomes simple once the instrument W_i is collapsed to the same dimension as D_i .

Second, if $\widehat{\tilde{V}}_i$ satisfies $\widehat{\tilde{V}}_i = \tilde{V}_i + o_p(1)$ for each i , then we can obtain $\widehat{V}_i = V_i + o_p(1)$ as $V_i = \Xi_1 \tilde{V}_i \Xi_1^\top$ and $\widehat{V}_i = \Xi_1 \widehat{\tilde{V}}_i \Xi_1^\top$. Following Koenker's [28] analysis for ordinary QR, Chernozhukov and Hansen [11] gave the estimating variance $\widehat{\tilde{V}}_i$ that satisfied $\widehat{\tilde{V}}_i = \tilde{V}_i +$

$o_p(1)$. The components of \widehat{V}_i that need to be estimated include $J_{i,\alpha}$, $J_{i,\psi}$ and S_i . They can be estimated as follows:

$$\widehat{S}_i = \frac{1}{T} \sum_{t=1}^T \widehat{s}_{it} \widehat{s}'_{it}, \widehat{J}_{i,\alpha} = \frac{1}{T} \sum_{t=1}^T [K(\widehat{\epsilon}_{it}/h_T)/h_T] \check{x}_{it} [\mathbf{d}'_{it}, \check{x}'_{it}]', \widehat{J}_{i,\psi} = \frac{1}{T} \sum_{t=1}^T [K(\widehat{\epsilon}_{it}/h_T)/h_T] \check{x}_{it} \check{x}'_{it},$$

where $\widehat{\epsilon}_{it} := Y_{it} - \mathbf{d}'_{it} \widehat{\alpha}(\tau) - \mathbf{x}'_{it} \widehat{\beta}(\tau) - \widehat{\eta}_i$, $\widehat{s}_{it} = [\tau - 1(\widehat{\epsilon}_{it} < 0)] \check{x}_{it}$, $\check{x}_{it} = [\omega'_{it}, \mathbf{x}'_{it}]'$, h_T is a bandwidth chosen such that $h_T \rightarrow 0$ and $Th_T^2 \rightarrow \infty$ as $T \rightarrow \infty$, $K(\cdot)$ is a kernel function.

When $\dim(\gamma) = \dim(\alpha)$, following Galvao and Wang's [7] analysis for ordinary MD-QR, the estimating variance satisfying condition (A10') is defined as follows:

$$\tau(1 - \tau) \Xi_1 \left(\frac{1}{T} \sum_{t=1}^T K_{h_N}(\widehat{\epsilon}_{it}) \check{x}_{it} [\mathbf{d}'_{it}, \check{x}'_{it}] \right)^{-1} \frac{1}{T} \sum_{t=1}^T \check{x}_{it} \check{x}'_{it} \left(\frac{1}{T} \sum_{t=1}^T K_{h_N}(\widehat{\epsilon}_{it}) [\mathbf{d}'_{it}, \check{x}'_{it}]' \check{x}'_{it} \right)^{-1} \Xi_1^T,$$

where $\widehat{\epsilon}_{it} := Y_{it} - \mathbf{d}'_{it} \widehat{\alpha}(\tau) - \mathbf{x}'_{it} \widehat{\beta}(\tau) - \widehat{\eta}_i$, h_N is a bandwidth chosen such that $h_N \rightarrow 0$ and $\frac{N}{Th_N} \rightarrow 0$ as $N \rightarrow \infty$, $K_{h_N}(\cdot)$ is defined in Kato [29].

In Theorems 3 and 4, we provide the limiting distributions for both sequential and simultaneous asymptotics. It is noteworthy that Theorems 3 and 4 demonstrate that the MD-IVQR estimator possesses the same limiting distribution in both sequential and simultaneous asymptotic scenarios. The reasons underlying this observation are explained in Galvao and Wang [7]. However, the two asymptotic properties primarily differ in terms of the divergence rates of T and N that are required. For the sequential limits asymptotics, we begin with letting T tend to infinity, followed by N . The proofs of the sequential limits entail great mathematical simplifications by drawing on the conclusions of predecessors, such as Chernozhukov and Hansen [11]. In contrast, the simultaneous asymptotics entail stringent requirements on the growth rate of T , which should tend to infinity faster than $N^2 \log N$. Consequently, the proof for the simultaneous asymptotics is relatively complex.

4. Monte Carlo Simulation

The samples are generated from the following model,

$$\begin{aligned} y_{it} &= \eta_i + d_{it}\alpha + x_{it}\beta + u_{it}, \\ v_{it} &= a * v_{it-1} + \epsilon_{it} + b * \epsilon_{it-1}, \\ x_{it} &= \varphi \kappa_i + v_{it}, \\ \eta_i &= \phi \bar{x}_i + \varepsilon_i, \end{aligned}$$

where $\bar{x}_i = T^{-1} \sum_{t=1}^T x_{it}$, $d_{it} = y_{it-1}$, $\kappa_i \sim U[0, 1]$ and $\varepsilon_i \sim N(0, 1)$. In the simulations, we fix $\varphi = 1, \phi = 0.5, a = 0.6, b = 0.2, \alpha = 0.5, \beta = 1$. We generate ϵ_{it} and u_{it} from the same distributions, which are Gaussian distribution $N(0, 1)$, t distribution with 3 degrees of freedom $t(3)$ and Chi-squared distribution with 3 degrees of freedom $\chi^2(3)$. In the generation process, we set $v_{i,-50} = 0$ and $y_{i,-50} = 0$, and discard their respective first 50 observations, using the observations $t = 0$ through T for the estimation. The design of the experiment carried here is based on Galvao [30] and Galvao [16]. As mentioned in [31], both random effects and fixed effects models exhibit endogeneity issues in dynamic panel data models because the lagged dependent variable y_{it-1} is correlated with the disturbance, even if it is assumed that u_{it} is not itself autocorrelated. Thus, we use $\omega_{it} = x_{it-1}$ as an instrument variable. The choice of the values of x_{it} lagged one periods as an instrument variable has been used in previous studies, such as Galvao [30] and Galvao [16]. We set the number of replications to 1000.

For the sake of comparing the performance and efficiency between different methods, we compare the bias, RMSE of the following estimators: the instrumental variables estimator for quantile regression with fixed effects (IVFEQR) as in Harding and Lamarche [13]; the minimum distance quantile regression estimator (MD-QR) as in Galvao and Wang [7]

and the proposed MD-IVQR estimator. In the simulations, we report results considering $\{N\} = \{50, 75, 100\}$, $\{T\} = \{50, 100, 200\}$, and quantiles $\tau = \{0.25, 0.50, 0.75\}$. As the true value of α is 0.5, we generate a series of α_j values from 0.3 to 0.7 in steps of 0.01 in the estimation process. In practice, we suggest first calculating the QR or MD-QR estimator and then setting a series of α_j values near this estimate in the initial step.

Moreover, in quantile regression, the error term u_{it} follows a general distribution F , and the derivative of the quantile function $d\frac{F^{-1}(\tau)}{d\tau}$ is referred to as the sparsity function (see [32]). Therefore, to examine whether the finite sample performance of the estimator is affected by sparsity function estimation in the weights, we calculate both the true and estimated sparsity functions within the corresponding variance–covariance matrix for the MD-IVQR and MD-QR estimators. We denote the MD-IVQR using the true sparsity as MDIVT and the MD-IVQR using the estimated sparsity as MDIVE. For MDIVE, we estimate the variance-covariance matrix using kernel estimation with the Gaussian kernel function. Following Galvao and Wang [7], we set the bandwidth as $h = 1.3h_{HS}$, where h_{HS} represents the Hall–Sheather bandwidth. Similarly, the MD-QR using the true sparsity is denoted as MDT, while the MD-QR using the estimated sparsity is abbreviated as MDE. Given the table layout and size, we abbreviate IVFEQR as IVFE.

4.1. Bias and RMSE

Table 1 presents the bias and RMSE of the estimators for α and β when $u_{it} \sim N(0, 1)$. In terms of the bias of α , it is evident that the IVFE, MDIVE, and MDIVT estimators outperform the MDE and MDT estimators. This implies that the methods using IV yield superior estimation results. The RMSE results are consistent with the bias findings. The RMSE of the MDE and MDT estimators is generally larger than that of the other three estimators.

Regarding the coefficient of the exogenous variable β , the overall results obtained by the five estimators do not significantly differ from each other. However, when N and T are small, the MDE and MDT estimators perform noticeably worse than the other three estimators. Furthermore, the RMSE of the MDIVT estimator is smaller than that of MDIVE, and the RMSE of the MDT estimator is also smaller than that of MDE. Nevertheless, the impact of sparsity estimation diminishes as the sample size increases. In other words, the estimation results of MDIVE and MDIVT estimators do not vary significantly with the increase in sample size. Whether the sparsity function is estimated or known, it does not significantly affect the estimation performance as the sample size grows.

Furthermore, as noted by Galvao and Wang [7], it is observed that the bias decreases as T increases for all estimators. However, this reduction in bias does not occur as N increases due to the incidental parameter problem. For a fixed N , the bias and RMSE of the five estimators decrease as T increases. Similarly, for a fixed T , the bias and RMSE of the MDIVE and MDIVT estimators tend to decrease with an increase in N . Meanwhile, there is little disparity in the estimation effects on the parameters at different quantiles.

Table 2 displays the bias and RMSE of estimators for α and β when $u_{it} \sim t(3)$. The results are comparable to those obtained from the $N(0, 1)$ distribution. The MDIVE and MDIVT estimators outperform the MDE and MDT estimators, and exhibit similar performance to the IVFE estimators.

Table 3 provides the the bias and RMSE of estimators for α and β when $u_{it} \sim \chi^2(3)$. The results demonstrate similarities to those obtained from the $N(0, 1)$ and $t(3)$ distributions at the 0.25 and 0.5 quantiles. However, the estimation results are relatively poorer at higher quantile points.

Table 1. Bias and RMSE of estimators for α and β when $u_{it} \sim N(0,1)$.

τ	N	T	α					β						
			IVFE	MDE	MDT	MDIVE	MDIVT	IVFE	MDE	MDT	MDIVE	MDIVT		
0.25	Bias	50	-0.001	-0.012	-0.017	0.006	0.002	-0.003	0.022	0.010	0.002	-0.003		
		50	100	-0.003	-0.006	-0.008	0.003	-0.001	0.004	0.012	0.009	0.001	0.001	
			200	0.000	-0.003	-0.004	0.002	0.001	-0.001	0.003	0.002	-0.001	-0.001	
			50	0.000	-0.009	-0.015	0.005	0.003	0.000	0.013	0.010	0.001	-0.003	
		75	100	0.001	-0.004	-0.006	0.004	0.003	-0.001	0.010	0.006	-0.001	-0.002	
			200	-0.001	-0.003	-0.004	0.001	0.001	0.000	0.004	0.003	0.001	-0.001	
	50		-0.002	-0.011	-0.015	0.004	0.002	0.002	0.021	0.013	0.005	-0.002		
	100	100	0.000	-0.004	-0.007	0.004	0.002	-0.001	0.009	0.007	-0.001	-0.001		
		200	0.001	-0.003	-0.004	0.002	0.001	-0.001	0.007	0.004	0.000	0.000		
		50	0.013	0.027	0.025	0.019	0.014	0.021	0.044	0.027	0.032	0.025		
	RMSE	50	100	0.010	0.014	0.015	0.012	0.012	0.014	0.027	0.020	0.022	0.017	
			200	0.008	0.010	0.009	0.009	0.008	0.011	0.017	0.013	0.015	0.013	
			50	0.010	0.017	0.020	0.014	0.012	0.012	0.026	0.019	0.020	0.018	
		75	100	0.007	0.012	0.011	0.011	0.009	0.011	0.021	0.016	0.019	0.015	
			200	0.005	0.007	0.007	0.007	0.006	0.008	0.013	0.010	0.011	0.009	
			50	0.009	0.018	0.020	0.013	0.011	0.015	0.032	0.023	0.023	0.019	
		100	100	0.008	0.011	0.011	0.012	0.009	0.012	0.019	0.016	0.018	0.015	
			200	0.006	0.008	0.007	0.008	0.007	0.006	0.013	0.009	0.012	0.010	
50			-0.002	-0.014	-0.015	0.002	0.002	0.001	0.014	0.012	0.002	-0.003		
0.5		Bias	50	100	-0.001	-0.008	-0.006	0.002	0.002	0.002	0.007	0.006	-0.002	0.000
				200	0.000	-0.004	-0.004	0.000	0.000	0.000	0.004	0.003	-0.001	0.000
				50	0.001	-0.016	-0.017	0.004	0.003	0.000	0.017	0.015	-0.003	-0.004
	75		100	0.000	-0.006	-0.007	0.002	0.002	-0.001	0.005	0.006	-0.003	-0.003	
			200	0.000	-0.005	-0.004	0.001	0.000	0.001	0.005	0.005	-0.001	-0.001	
			50	-0.001	-0.017	-0.017	0.001	0.002	0.001	0.009	0.011	-0.005	-0.004	
	100	100	-0.001	-0.008	-0.009	0.001	0.001	0.001	0.009	0.008	0.000	-0.001		
		200	0.000	-0.004	-0.004	0.001	0.000	0.001	0.004	0.005	-0.001	-0.001		
		50	0.014	0.022	0.023	0.017	0.015	0.021	0.031	0.029	0.027	0.024		
	RMSE	50	100	0.011	0.014	0.013	0.013	0.012	0.014	0.019	0.016	0.019	0.015	
			200	0.007	0.009	0.008	0.008	0.007	0.011	0.014	0.013	0.013	0.012	
			50	0.010	0.021	0.020	0.013	0.012	0.014	0.026	0.022	0.019	0.017	
		75	100	0.008	0.010	0.011	0.008	0.009	0.012	0.018	0.015	0.015	0.014	
			200	0.006	0.008	0.008	0.006	0.006	0.008	0.011	0.009	0.009	0.008	
			50	0.010	0.022	0.022	0.012	0.011	0.012	0.020	0.020	0.018	0.015	
		100	100	0.007	0.011	0.011	0.008	0.007	0.010	0.016	0.013	0.012	0.011	
			200	0.006	0.007	0.007	0.007	0.007	0.007	0.010	0.009	0.009	0.009	
			50	0.004	-0.017	-0.014	0.004	0.006	-0.003	0.004	0.005	-0.004	-0.002	
0.75		Bias	50	100	0.001	-0.011	-0.010	0.003	0.003	-0.001	0.001	0.006	-0.004	-0.002
				200	-0.001	-0.004	-0.003	0.001	0.001	0.002	0.001	0.003	-0.003	0.000
				50	0.001	-0.020	-0.014	0.003	0.003	0.001	0.011	0.015	-0.003	-0.002
	75		100	0.001	-0.010	-0.007	0.001	0.002	0.003	0.009	0.008	-0.001	0.002	
			200	0.000	-0.005	-0.003	0.001	0.001	-0.002	0.006	0.007	-0.001	-0.001	
			50	0.000	-0.016	-0.012	0.001	0.002	0.002	0.015	0.014	0.001	0.000	
	100	100	0.000	-0.012	-0.010	0.000	0.001	0.000	0.009	0.008	0.000	0.000		
		200	0.001	-0.005	-0.004	0.000	0.000	0.000	0.003	0.002	0.000	0.000		
		50	0.013	0.027	0.022	0.015	0.015	0.020	0.035	0.028	0.030	0.025		
	RMSE	50	100	0.010	0.018	0.016	0.014	0.012	0.014	0.025	0.020	0.028	0.018	
			200	0.007	0.010	0.009	0.009	0.008	0.009	0.014	0.012	0.015	0.012	
			50	0.009	0.026	0.019	0.013	0.012	0.013	0.031	0.026	0.023	0.019	
		75	100	0.008	0.015	0.012	0.010	0.010	0.011	0.019	0.016	0.016	0.013	
			200	0.006	0.009	0.007	0.007	0.007	0.009	0.011	0.010	0.012	0.012	
			50	0.010	0.023	0.019	0.013	0.012	0.014	0.024	0.020	0.025	0.019	
		100	100	0.007	0.015	0.013	0.009	0.009	0.008	0.017	0.016	0.014	0.012	
			200	0.005	0.008	0.008	0.007	0.006	0.008	0.011	0.011	0.012	0.010	

Table 2. Bias and RMSE of estimators for α and β when $u_{it} \sim t(3)$.

τ	N	T	α					β					
			IVFE	MDE	MDT	MDIVE	MDIVT	IVFE	MDE	MDT	MDIVE	MDIVT	
0.25	Bias	50	0.001	-0.008	-0.007	0.004	0.004	0.001	0.009	0.005	0.001	-0.004	
		50	100	-0.001	-0.004	-0.003	0.003	0.001	0.003	0.004	0.004	0.000	0.000
			200	0.000	-0.002	-0.003	0.001	0.000	0.000	0.003	0.002	0.000	0.000
			50	-0.001	-0.004	-0.010	0.005	0.003	0.001	0.009	0.007	0.002	-0.004
		75	100	-0.001	0.001	-0.004	0.003	0.000	0.001	0.000	0.003	-0.002	-0.001
			200	0.000	-0.003	-0.003	0.002	0.001	0.000	0.003	0.003	0.001	-0.001
	50		0.001	-0.007	-0.007	0.006	0.004	-0.003	0.002	0.003	-0.003	-0.004	
	100	100	0.000	-0.006	-0.004	0.003	0.001	0.000	0.003	0.004	-0.001	-0.002	
		200	0.000	-0.002	-0.002	0.001	0.001	0.000	0.003	0.002	0.001	-0.001	
		50	0.013	0.017	0.015	0.015	0.013	0.018	0.026	0.018	0.025	0.019	
	RMSE	50	100	0.010	0.011	0.008	0.012	0.011	0.016	0.023	0.014	0.023	0.015
			200	0.007	0.007	0.005	0.008	0.007	0.009	0.014	0.008	0.017	0.009
			50	0.013	0.013	0.013	0.017	0.014	0.018	0.023	0.016	0.024	0.016
		75	100	0.008	0.008	0.007	0.012	0.008	0.012	0.016	0.010	0.020	0.012
			200	0.006	0.007	0.005	0.006	0.005	0.008	0.015	0.007	0.016	0.008
			50	0.009	0.010	0.010	0.015	0.010	0.013	0.015	0.013	0.022	0.016
	100	100	0.007	0.006	0.007	0.010	0.007	0.009	0.012	0.010	0.014	0.008	
		200	0.005	0.005	0.004	0.006	0.005	0.008	0.010	0.006	0.011	0.007	
50		-0.003	-0.007	-0.008	-0.002	0.002	0.002	0.008	0.007	0.003	0.001		
0.5	Bias	50	100	0.001	-0.003	-0.003	0.001	0.001	0.000	0.005	0.002	0.001	-0.001
			200	0.000	-0.002	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	-0.001
			50	-0.001	-0.005	-0.006	0.001	0.002	0.002	0.003	0.006	-0.001	-0.002
		75	100	-0.001	-0.003	-0.003	0.001	0.001	0.002	0.003	0.004	-0.001	0.000
			200	0.000	-0.001	-0.001	0.000	0.000	-0.001	0.000	0.000	0.000	0.000
			50	-0.002	-0.007	-0.008	0.000	0.001	0.004	0.007	0.008	0.001	-0.001
	100	100	-0.001	-0.003	-0.004	0.000	0.001	0.001	0.003	0.003	0.001	0.000	
		200	0.000	-0.001	-0.002	0.000	0.000	0.001	0.000	0.001	0.000	0.000	
		50	0.013	0.013	0.011	0.014	0.013	0.019	0.022	0.018	0.023	0.019	
	RMSE	50	100	0.008	0.008	0.007	0.011	0.009	0.012	0.014	0.011	0.014	0.012
			200	0.007	0.005	0.003	0.008	0.007	0.009	0.010	0.006	0.012	0.008
			50	0.010	0.011	0.011	0.012	0.011	0.014	0.018	0.015	0.017	0.015
		75	100	0.009	0.007	0.006	0.010	0.009	0.013	0.012	0.009	0.014	0.011
			200	0.005	0.004	0.003	0.006	0.005	0.008	0.009	0.006	0.009	0.008
			50	0.009	0.010	0.010	0.010	0.009	0.013	0.014	0.014	0.014	0.012
	100	100	0.006	0.008	0.007	0.007	0.006	0.008	0.010	0.008	0.011	0.008	
		200	0.004	0.005	0.004	0.005	0.004	0.006	0.006	0.005	0.008	0.006	
		50	0.002	-0.012	-0.007	0.002	0.004	0.000	0.005	0.006	-0.005	-0.002	
0.75	Bias	50	100	-0.001	-0.007	-0.005	-0.002	0.001	0.001	0.002	0.004	-0.002	-0.001
			200	0.000	-0.003	-0.002	-0.001	0.001	0.000	0.000	0.001	-0.001	-0.001
			50	0.002	-0.012	-0.009	0.002	0.003	-0.001	0.004	0.008	-0.006	-0.005
		75	100	0.002	-0.004	-0.002	0.002	0.003	-0.002	0.001	0.002	-0.004	-0.004
			200	0.001	-0.004	-0.002	-0.001	0.000	-0.002	0.003	0.002	0.000	-0.002
			50	-0.001	-0.015	-0.009	-0.003	0.001	-0.001	0.005	0.006	-0.005	-0.003
	100	100	-0.002	-0.007	-0.005	-0.001	0.000	0.001	0.002	0.003	-0.003	-0.002	
		200	0.001	-0.003	-0.002	0.001	0.000	-0.001	0.002	0.003	-0.002	-0.001	
		50	0.014	0.018	0.015	0.017	0.014	0.019	0.028	0.020	0.028	0.021	
	RMSE	50	100	0.009	0.011	0.009	0.013	0.010	0.013	0.019	0.012	0.022	0.013
			200	0.008	0.008	0.006	0.012	0.008	0.010	0.015	0.008	0.017	0.011
			50	0.011	0.017	0.013	0.015	0.012	0.014	0.020	0.015	0.023	0.014
		75	100	0.008	0.010	0.007	0.012	0.009	0.010	0.018	0.010	0.020	0.013
			200	0.006	0.008	0.005	0.009	0.006	0.008	0.014	0.008	0.014	0.008
			50	0.009	0.020	0.012	0.015	0.010	0.012	0.021	0.014	0.023	0.015
	100	100	0.006	0.010	0.007	0.009	0.006	0.009	0.017	0.009	0.017	0.010	
		200	0.006	0.006	0.004	0.007	0.006	0.007	0.013	0.007	0.014	0.007	

Table 3. Bias and RMSE of estimators for α and β when $u_{it} \sim \chi^2(3)$.

τ	N	T	α					β						
			IVFE	MDE	MDT	MDIVE	MDIVT	IVFE	MDE	MDT	MDIVE	MDIVT		
0.25	Bias	50	−0.002	−0.008	−0.007	0.003	0.000	0.001	0.009	0.010	−0.010	0.001		
		50	100	0.000	−0.005	−0.004	0.002	0.001	0.000	0.004	0.006	−0.007	0.001	
			200	0.001	−0.004	−0.001	0.003	0.001	0.000	0.000	0.003	−0.009	0.000	
			50	0.000	−0.009	−0.008	0.004	0.000	0.001	0.012	0.014	−0.010	0.006	
		75	100	0.001	−0.006	−0.004	0.003	0.001	0.000	0.007	0.008	−0.008	0.001	
			200	0.000	−0.003	−0.002	0.002	0.000	0.000	−0.001	0.003	−0.007	0.001	
	50		0.001	−0.009	−0.007	0.005	0.002	0.001	0.014	0.012	−0.012	0.002		
	100	100	0.000	−0.005	−0.003	0.004	0.000	0.000	0.005	0.006	−0.010	0.001		
		200	0.001	−0.002	−0.002	0.002	0.001	−0.001	0.001	0.003	−0.008	0.000		
		50	0.015	0.015	0.011	0.018	0.017	0.019	0.032	0.023	0.029	0.017		
	RMSE	50	100	0.009	0.009	0.008	0.011	0.010	0.012	0.025	0.014	0.019	0.010	
			200	0.007	0.006	0.004	0.006	0.005	0.010	0.015	0.010	0.017	0.011	
50			0.011	0.014	0.011	0.012	0.012	0.016	0.024	0.022	0.024	0.015		
75		100	0.009	0.009	0.007	0.010	0.010	0.011	0.023	0.013	0.014	0.011		
		200	0.006	0.005	0.004	0.005	0.004	0.008	0.012	0.007	0.012	0.007		
		50	0.009	0.013	0.010	0.011	0.010	0.012	0.023	0.020	0.018	0.013		
100		100	0.006	0.009	0.005	0.006	0.006	0.010	0.014	0.009	0.016	0.008		
		200	0.005	0.004	0.003	0.005	0.005	0.008	0.009	0.008	0.011	0.007		
		50	−0.001	−0.018	−0.014	0.011	0.000	0.003	0.010	0.013	0.016	0.001		
0.5		Bias	50	100	−0.002	−0.008	−0.006	0.008	−0.001	0.000	0.000	0.006	−0.011	0.000
				200	0.000	−0.006	−0.004	0.005	0.000	−0.001	−0.001	0.004	0.000	0.000
				50	−0.002	−0.019	−0.016	0.012	−0.001	0.004	0.013	0.019	0.025	0.003
	75		100	0.000	−0.010	−0.008	0.008	0.000	0.001	0.005	0.011	−0.007	0.002	
			200	0.000	−0.004	−0.003	0.003	0.000	0.000	0.000	0.003	0.000	0.001	
			50	0.000	−0.016	−0.014	0.011	0.001	−0.003	−0.003	0.009	−0.011	−0.003	
	100	100	0.000	−0.008	−0.007	0.008	0.000	−0.001	−0.001	0.006	0.007	0.000		
		200	−0.001	−0.003	−0.003	0.002	0.000	0.002	0.000	0.005	0.000	0.001		
		50	0.013	0.026	0.019	0.019	0.014	0.019	0.050	0.030	0.055	0.023		
	RMSE	50	100	0.010	0.015	0.012	0.013	0.011	0.014	0.030	0.018	0.035	0.017	
			200	0.007	0.009	0.008	0.009	0.007	0.010	0.023	0.014	0.023	0.012	
			50	0.011	0.024	0.020	0.017	0.012	0.017	0.040	0.030	0.051	0.022	
75		100	0.009	0.013	0.010	0.011	0.009	0.013	0.022	0.018	0.033	0.015		
		200	0.006	0.009	0.006	0.008	0.005	0.008	0.017	0.013	0.021	0.009		
		50	0.011	0.021	0.017	0.015	0.010	0.015	0.029	0.030	0.043	0.019		
100		100	0.007	0.012	0.009	0.010	0.007	0.011	0.020	0.017	0.031	0.011		
		200	0.005	0.006	0.005	0.009	0.007	0.008	0.015	0.011	0.022	0.010		
		50	−0.001	−0.033	−0.024	0.023	0.000	−0.003	0.016	0.015	0.029	−0.003		
0.75		Bias	50	100	0.000	−0.016	−0.012	0.012	0.000	−0.001	−0.011	0.010	0.012	−0.002
				200	0.002	−0.010	−0.008	0.004	0.001	0.000	0.000	0.004	0.006	−0.002
				50	0.000	−0.040	−0.025	0.026	0.003	−0.005	0.025	0.024	0.018	−0.002
	75		100	−0.002	−0.014	−0.014	0.013	−0.001	0.001	−0.007	0.009	0.012	−0.002	
			200	0.000	−0.005	−0.004	0.002	0.000	0.000	−0.003	0.001	0.004	0.000	
			50	0.001	−0.029	−0.022	0.019	0.004	−0.001	0.007	0.016	0.017	−0.006	
	100	100	0.000	−0.015	−0.011	0.010	0.002	0.000	0.006	0.009	0.012	−0.005		
		200	0.000	−0.006	−0.005	0.005	0.001	−0.001	−0.005	0.005	0.006	−0.002		
		50	0.017	0.043	0.035	0.031	0.015	0.018	0.068	0.053	0.076	0.034		
	RMSE	50	100	0.010	0.026	0.018	0.022	0.011	0.014	0.057	0.032	0.063	0.022	
			200	0.008	0.016	0.010	0.011	0.008	0.011	0.036	0.021	0.041	0.021	
			50	0.012	0.036	0.028	0.025	0.016	0.021	0.072	0.044	0.074	0.030	
75		100	0.009	0.021	0.017	0.018	0.008	0.013	0.050	0.032	0.065	0.019		
		200	0.003	0.006	0.005	0.013	0.007	0.009	0.032	0.016	0.032	0.014		
		50	0.011	0.037	0.026	0.023	0.011	0.015	0.058	0.042	0.071	0.026		
100		100	0.007	0.023	0.015	0.016	0.007	0.009	0.043	0.020	0.057	0.016		
		200	0.005	0.013	0.009	0.011	0.005	0.009	0.030	0.019	0.034	0.014		

It is worth mentioning that, contrary to the findings of Galvao and Wang [7], we observe that MDIVT’s estimation results are no longer always better than MDIVE’s. The reason is that when solving the model, we first give a suitable set of values α , and then we analyze the new y instead of the original model. From the above analysis, on the one hand, we can find that MDIVT is better than MDIVE in most cases, while on the other hand, the effects of sparsity estimation diminish as the sample size increases.

Finally, we conducted an analysis to assess the bias in the MDIVE estimators for α as α varies. Considering $N = T = 100$, $\tau = 0.5$ and assuming $u_{it} \sim N(0, 1)$, we obtained estimates of the bias in the MDIVE estimators for different α values ranging from 0.2 to 0.8 with a step size of 0.05. The results of the bias estimation for the MDIVE estimators of α are presented in Figure 1. Notably, the figure demonstrates that the bias falls within the range of -0.0025 to 0.0035. Additionally, it is worth mentioning that incorporating x_{it-1} as an instrumental variable exhibited minimal impact on the bias of the MDIVE estimators across different α values. These findings align with the results reported by Galvao [16].

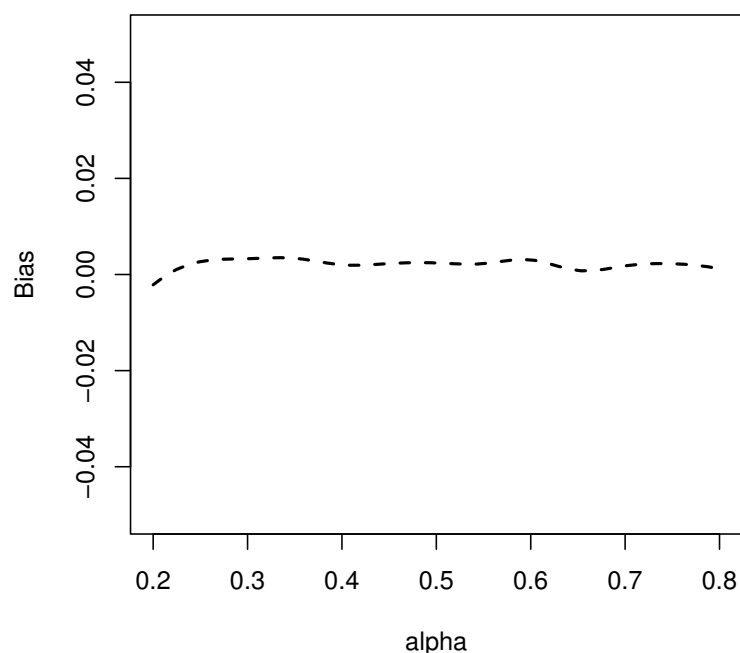


Figure 1. Bias of the MDIVE estimators for α when α varies.

4.2. Estimation Speed

As discussed by Galvao and Wang [7], the estimation procedure for FE-QR models can be quite cumbersome, as is the case with IVFEQR models, which we have previously discussed. Consequently, we propose using MD-IVQR as an alternative estimator because it is both straightforward to implement and computationally feasible.

In order to compare the computing time for IVFEQR and MD-IVQR estimators, the durations of IVFEQR and MD-IVQR estimations in a simulation for various levels of N and T where we estimate the above model for one particular quantile ($\tau = 0.5$) are presented. The code of IVFEQR is provided by Harding and Lamarche on their respective websites. We compare the time in terms of three quantities: user, system, and elapsed. User time gives the CPU time spent by the current process and system time gives the CPU time spent by the kernel (the operating system) on behalf of the current process. Elapsed time is in our best interests and gives the time charged to the CPU(s) for one replication of the simulation. In addition, the processor of computer we use is an Intel Core i5-5257U cpu@2.70 GHz RAM 8.00 GB and the program is implemented using R 3.1.1.

Table 4 illustrates the durations (in seconds) of the IVFEQR and MD-IVQR estimations in a simulation for sample sizes $N, T \in \{50, 100, 150, 250, 300, 450\}$. From Table 4, we can see that when N, T are more than 150, the duration of MD-IVQR estimation is less than IVFEQR estimation. With the increase of N and T , the durations of MD-IVQR estimation and IVFEQR estimation both increase, and the duration of IVFEQR estimation is significantly more than that of MD-IVQR estimation. That is, the duration of MD-IVQR estimation increases slowly, while the duration of IVFEQR increases more obviously. The significant disparity in the computational time required for estimating the parameters, as discussed in Galvao and Wang [7], arises from the distinct approach employed by IVFEQR estimation

and MD-IVQR estimation. IVFEQR estimation tackles a single, larger optimization problem, whereas MD-IVQR estimation subdivides the data into smaller segments and estimates them individually.

Table 4. The durations of IVFEQR and MD-IVQR estimations in a simulation for various levels of N and T .

	IVFEQR			MD-IVQR		
	User	System	Elapsed	User	System	Elapsed
$N = T = 50$	3.97	1.13	5.62	5.67	0.39	6.45
$N = T = 100$	9.00	1.55	10.89	10.99	0.63	11.77
$N = T = 150$	18.16	2.47	21.80	15.86	0.58	16.86
$N = T = 250$	84.44	19.25	108.00	39.06	0.63	40.31
$N = T = 300$	147.63	35.28	186.66	49.03	1.11	50.75
$N = T = 450$	421.64	96.61	525.32	59.05	1.29	60.93

Furthermore, by holding N and T fixed separately, we observe whether the duration of the estimation is more sensitive to T or N . Table 5 shows the durations (in seconds) of the IVFEQR and MD-IVQR estimations in a simulation for sample sizes $T \in \{50, 100, 250, 500, 750, 1000\}$ with fixed N . With the increase in time, T , the duration of both estimation procedures increase, but both are relatively flat. While from Table 6 with fixed T , we can see that as N increases, the duration of computing each estimator increases, although the increment is mild for MD-IVQR, it is drastic for IVFEQR. When N is greater than 250, IVFEQR consumes significantly more computing time than MD-IVQR. By comparison of Tables 5 and 6, it can be concluded that both estimators demonstrate greater sensitivity to variations in the size of N . In addition, MD-IVQR is much less sensitive to sample size than IVFEQR.

Table 5. The durations of IVFEQR and MD-IVQR estimations in a simulation for various levels of T when $N = 100$.

$N = 100$	IVFEQR			MD-IVQR		
	User	System	Elapsed	User	System	Elapsed
$T = 50$	6.27	1.55	8.08	10.33	0.64	11.28
$T = 100$	8.58	1.32	10.47	11.07	0.39	11.93
$T = 250$	7.95	2.50	20.71	13.23	0.67	14.08
$T = 500$	31.77	4.76	37.39	19.76	0.80	21.28
$T = 750$	48.42	13.9	68.13	21.22	0.90	23.23
$T = 1000$	54.92	12.94	68.96	35.95	0.58	37.67

Table 6. The durations of IVFEQR and MD-IVQR estimations in a simulation for various levels of N when $T = 100$.

$T = 100$	IVFEQR			MD-IVQR		
	User	System	Elapsed	User	System	Elapsed
$N = 50$	5.34	1.00	6.53	5.41	0.55	6.28
$N = 100$	8.64	1.33	10.78	10.65	0.45	11.77
$N = 250$	29.02	5.20	34.50	27.11	0.55	28.10
$N = 500$	78.34	17.72	96.09	38.89	0.23	39.16
$N = 750$	202.97	43.34	246.52	59.61	0.52	60.14
$N = 1000$	432.44	66.79	499.68	77.63	0.83	78.48

In summary, when the sample size N and time series length T are sufficiently large, IVFEQR requires substantially more computation time compared to MD-IVQR. Both esti-

mators are more sensitive to the magnitude of N rather than T . Therefore, when dealing with a large sample, particularly for large N , it is advisable to use the MD-IVQR estimator.

5. Application

In this section, we demonstrate the effectiveness of our new approach by applying it to investigate the income elasticity of health expenditure. The substantial increase in healthcare spending has presented a challenge for many countries, and the growing proportion of healthcare expenditures as a share of GDP is an important economic trend. We apply our approach to a panel of data from 166 countries spanning from 1995 to 2016: the healthcare expenditure dataset.

The literature has extensively examined the strong and positive relationship between GDP and health expenditures. Notable studies include Newhouse [33], Gerdtham and Jönsson [34], Moran and Simon [35], Murphy and Robert [36], Farag et al. [37], Acemoglu et al. [38], and Tian et al. [39]. Farag et al. [37] employed fixed effects and instrumental variables models to investigate the income elasticity of health care spending in developed and developing countries. They noted that endogeneity may arise due to unobservable variables driving both health expenditures and income. Acemoglu et al. [38] analyzed the income elasticity of health expenditure by using global oil prices as an instrument for local area income. Their study used time-series variation in oil prices between 1970 and 1990, interacted with cross-sectional variation in oil reserves across different areas of the Southern United States.

Like Acemoglu et al. [38], the panel data model can be established as

$$\ln ch_{it} = \alpha \ln cgdp_{it} + \eta_i + u_{it}, \tag{11}$$

where $i = 1, \dots, 166$ indicates different countries and regions, and $t = 1, \dots, 22$ indicates different years, η_i denotes regional heterogeneity. Variable descriptions are summarized in Table 7. One advantage of the above modeling is that it can control the unobserved country factors. The result of Hausman test shows that the fixed effects model is preferred.

Table 7. Variable descriptions.

Variable	Definition	Data Sources
$\ln ch$	log of health expenditure per capita	World Health Organization
$\ln cgdp$	log of GDP per capita	Worldbank Indicators
$\ln pop$	log of total population	Worldbank Indicators
$\ln oilprice$	log of oil price	West Texas Intermediate series
$indensity$	the total amount of oil reserves in country or region	U.S. Energy Information Administration

To address concerns of endogeneity, we propose using instrumental variables in estimating our model, while Farag et al. [37] used the share of agriculture in the economy and primary school net enrollment ratio as instruments for income, these may not be appropriate. The primary school net enrollment ratio reflects higher national human capital and income, but also higher education levels that may lead to increased medical expenditure. Instead, following Acemoglu et al. [38], we consider using global oil prices interacted with cross-sectional oil reserves as instruments. The use of oil shocks as an instrument for GDP began with Bruckner et al. [40]. In this study, the instrument we used is the least squares projection of $\ln cgdp$ on the instruments $\ln oilprice_{t-1} \times indensity_i$ and $\ln pop$, where the total amount of oil reserves, $indensity_i$ is calculated as estimated remaining reserves plus total cumulative oil production as of 1995.

The estimation results are illustrated in Figure 2. The findings indicate that the estimated income elasticity of medical expenditure at each quantile is greater than 1, which is consistent with the results of previous studies by Newhouse [33], Gerdtham and

Jönsson [34], and Murphy and Robert [36]. Furthermore, the income elasticity of medical expenditures at medium quantiles is larger than those at low and high quantiles.

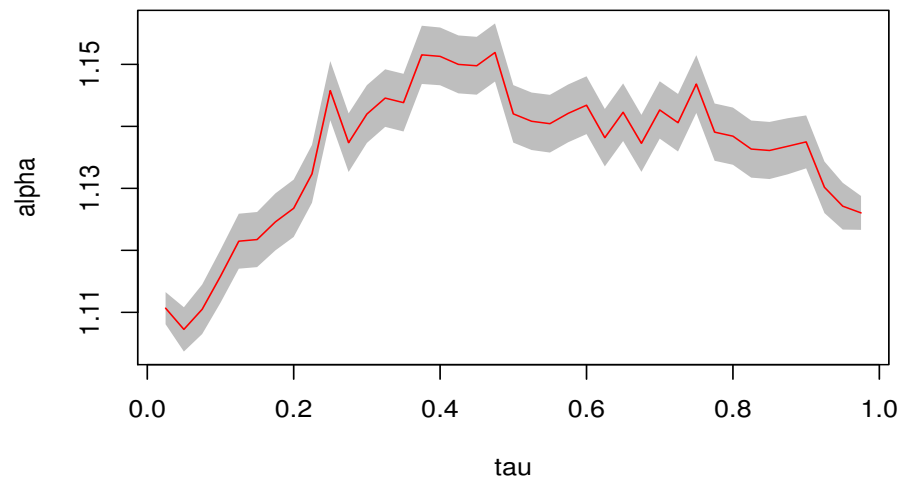


Figure 2. The MD-IVQR estimators at different quantiles. (The shaded area in the figure represents the 95% confidence interval).

To investigate the impact of oil-rich countries on our estimation results, we exclude them from our analysis as oil price fluctuations may be influenced by their oil production. According to the BP Statistical Review of World Energy 2017, OPEC’s proven oil reserves in 2016 accounted for approximately 71.5% of the world’s total. Additionally, we divide the 166 countries in our dataset into two groups: OPEC and non-OPEC. As shown in Figure 3, the estimated income elasticity of medical expenditure for non-OPEC countries is around 1.11–1.15, consistent with the findings presented in Figure 2. Furthermore, similar to our previous results, the income elasticity of health expenditures for low and high quantiles is lower than that for medium quantiles. For OPEC countries, the estimated income elasticity of medical expenditures is mostly below 1. However, the estimates increase with the quantile.

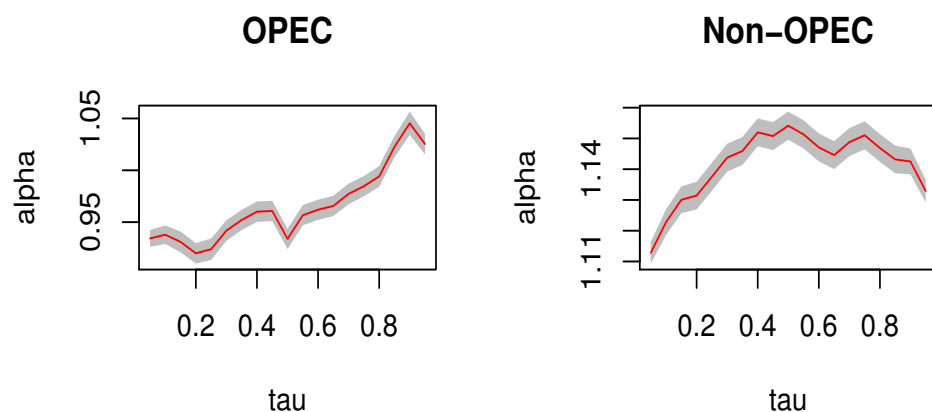


Figure 3. The MD-IVQR estimators for OPEC and non OPEC at different quantiles. (The shaded area in the figure represents the 95% confidence interval).

To account for different stages of economic development across countries, we divide the 166 countries in our dataset into two groups: OECD and non-OECD, representing developed and developing economies, respectively. As shown in Figure 4, the estimated income elasticity of health expenditure for OECD countries is approximately 1.14–1.18, whereas for non-OECD economies, it is around 1.04–1.08. Our findings suggest that OECD countries have a relatively higher elasticity of health expenditure compared to non-OECD countries, indicating that the income elasticity of medical expenditure may increase as

a country's economic development improves. This result is consistent with the findings of Farag et al. [37], who found that healthcare spending is less responsive to changes in income in low-income countries and more responsive to changes in middle and high-income countries.

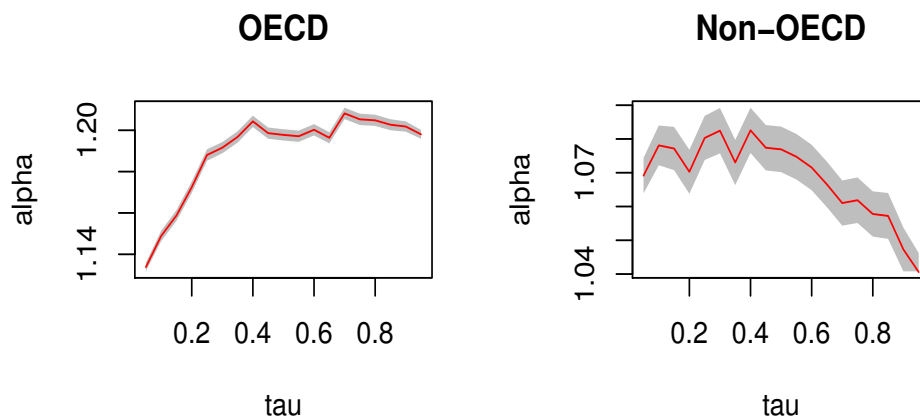


Figure 4. The MD-IVQR estimators for OECD and non OECD at different quantiles. (The shaded area in the figure represents the 95% confidence interval).

6. Conclusions

In this paper, we propose a minimum distance instrumental variable quantile regression (MD-IVQR) estimator for panel data with fixed effects, to address the biased parameter estimation problem caused by endogenous variables and to simplify cumbersome computations due to large N and T . The MD-IVQR estimator is defined as a weighted average of conventional individual instrumental variable quantile regression slope estimators, where the weights are given by the inverses of the corresponding individual variance-covariance matrices. This estimator is a modification of the MD-QR that can be constructed from a series of conventional instrumental variable quantile regressions. The MD-IVQR estimator combines the advantages of the MD-QR estimator and the IVQR estimator, making it computationally convenient and simple to implement in many typical applications.

We give the asymptotic properties of the MD-IVQR estimator in large samples, considering both sequential and simultaneous limits. For the sequential limits asymptotics, we first let T tend to infinity, and then N . However, for the simultaneous asymptotics, we strengthen conditions and require T to tend to infinity faster than $N^2 \log N$. Monte Carlo experiments show that the estimation results with the use of IV perform much better for α . Furthermore, the computing time of the MD-IVQR is similar to that of IVFEQR when N and T are not large (less than 150). However, when N and T are large enough, IVFEQR consumes significantly more computing time than MD-IVQR.

Finally, we apply the proposed method to analyze the income elasticity of health expenditure. Our results indicate that the estimated income elasticity of medical expenditure at each quantile is greater than 1.

Author Contributions: Conceptualization, L.T. (Li Tao) and L.T. (Lingnan Tai); methodology, L.T. (Li Tao); software, L.T. (Li Tao); writing—original draft preparation, L.T. (Li Tao), L.T. (Lingnan Tai) and M.Q.; writing—review and editing, L.T. (Li Tao), L.T. (Lingnan Tai), M.Q. and M.T.; visualization, L.T. (Li Tao); supervision, M.Q. and M.T.; funding acquisition, L.T. (Li Tao) and M.T. All authors have read and agreed to the published version of the manuscript.

Funding: Tao's work was supported by the R&D Program of Beijing Municipal Education Commission (No. SM202210037009) and the Youth Science Fund for Beijing Wuzi University (2023XJQN04); Tian's work was partially supported by the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (22XNL016).

Data Availability Statement: The healthcare expenditure dataset in this study is available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs

Appendix A.1. Consistency of $\widehat{\boldsymbol{\theta}}_{MD-IVQR}$ under Sequential Asymptotics

For convenience we collect important definitions below. Let $\boldsymbol{\vartheta}_i = (\eta'_i, \boldsymbol{\beta}', \boldsymbol{\gamma}')'$, $\mathbf{x}_{it}^* = (1, \mathbf{x}'_{it}, \boldsymbol{\omega}'_{it})'$. Define

$$\begin{aligned} \mathbb{M}_{iT}(\boldsymbol{\alpha}, \boldsymbol{\vartheta}_i, \tau) &= \frac{1}{T} \sum_{t=1}^T \rho_\tau(y_{it} - \mathbf{d}'_{it}\boldsymbol{\alpha} - \mathbf{x}_{it}^* \boldsymbol{\vartheta}_i), \quad M_i(\boldsymbol{\alpha}, \boldsymbol{\vartheta}_i, \tau) = EM_{iT}(\boldsymbol{\alpha}, \boldsymbol{\vartheta}_i, \tau), \\ \widehat{\boldsymbol{\vartheta}}_i(\boldsymbol{\alpha}, \tau) &\equiv (\widehat{\eta}_i(\boldsymbol{\alpha}, \tau), \widehat{\boldsymbol{\beta}}'_i(\boldsymbol{\alpha}, \tau), \widehat{\boldsymbol{\gamma}}'_i(\boldsymbol{\alpha}, \tau))' \equiv \arg \inf \mathbb{M}_{iT}(\boldsymbol{\alpha}, \boldsymbol{\vartheta}_i, \tau), \\ \boldsymbol{\vartheta}_i(\boldsymbol{\alpha}, \tau) &\equiv (\eta_i(\boldsymbol{\alpha}, \tau), \boldsymbol{\beta}'(\boldsymbol{\alpha}, \tau), \boldsymbol{\gamma}'(\boldsymbol{\alpha}, \tau))' \equiv \arg \inf M_i(\boldsymbol{\alpha}, \boldsymbol{\vartheta}_i, \tau), \\ \Delta_{iT}(\boldsymbol{\alpha}, \boldsymbol{\vartheta}_i, \tau) &= \mathbb{M}_{iT}(\boldsymbol{\alpha}, \boldsymbol{\vartheta}_i, \tau) - \mathbb{M}_{iT}(\boldsymbol{\alpha}_0, \boldsymbol{\vartheta}_i(\boldsymbol{\alpha}), \tau), \\ W_{iT}[\boldsymbol{\alpha}] &:= \widehat{\boldsymbol{\gamma}}'_i(\boldsymbol{\alpha}, \tau) \widehat{A}(\boldsymbol{\alpha}) \widehat{\boldsymbol{\gamma}}_i(\boldsymbol{\alpha}, \tau), \quad W[\boldsymbol{\alpha}] := \boldsymbol{\gamma}'(\boldsymbol{\alpha}, \tau) A(\boldsymbol{\alpha}) \boldsymbol{\gamma}(\boldsymbol{\alpha}, \tau), \\ \widehat{\boldsymbol{\alpha}}(\tau) &\equiv \arg \inf_{\boldsymbol{\alpha} \in \mathcal{A}} W_{iT}[\boldsymbol{\alpha}], \quad \boldsymbol{\alpha}^* \equiv \arg \inf_{\boldsymbol{\alpha} \in \mathcal{A}} W[\boldsymbol{\alpha}], \\ \widehat{\boldsymbol{\vartheta}}_i(\tau) &= (\widehat{\eta}_i(\tau), \widehat{\boldsymbol{\beta}}'_i(\tau), \widehat{\boldsymbol{\gamma}}'_i(\tau))' \equiv \widehat{\boldsymbol{\vartheta}}_i(\widehat{\boldsymbol{\alpha}}_i(\tau), \tau), \\ \boldsymbol{\vartheta}_i(\tau) &= (\eta_i(\tau), \boldsymbol{\beta}'(\tau), 0)' \equiv \boldsymbol{\vartheta}_i(\boldsymbol{\alpha}(\tau), \tau). \end{aligned}$$

The following Lemma shows the consistency of the IVQR estimator and its large sample property.

Lemma A1. Let $\tilde{\boldsymbol{\theta}}_i = (\boldsymbol{\alpha}', \boldsymbol{\beta}', \eta_i) = (\boldsymbol{\alpha}', \tilde{\boldsymbol{\beta}}'_i)'$, and $\widehat{\boldsymbol{\theta}}_i = (\widehat{\boldsymbol{\alpha}}', \widehat{\boldsymbol{\beta}}', \widehat{\eta}_i) = (\widehat{\boldsymbol{\alpha}}', \widehat{\boldsymbol{\beta}}'_i)'$ is the instrumental quantile regression estimator for each individual time-series data. Under conditions (A1)–(A7), for each i , $\widehat{\boldsymbol{\theta}}_i \xrightarrow{p} \tilde{\boldsymbol{\theta}}_i$; and

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_i - \tilde{\boldsymbol{\theta}}_i) \xrightarrow{d} N(0, \tilde{V}_i), \quad \tilde{V}_i = (K'_i, L'_i)' S_i (K'_i, L'_i),$$

where, for $\tilde{\mathbf{x}}_i = [W'_i, \tilde{\mathbf{x}}'_i]'$ and $\epsilon_i = Y_i - D'_i \boldsymbol{\alpha}(\tau) - \tilde{\mathbf{x}}'_i \tilde{\boldsymbol{\beta}}_i(\tau)$, $S_i = \tau(1 - \tau)E[\tilde{\mathbf{x}}_i \tilde{\mathbf{x}}'_i]$, $K_i = (J'_{i,\boldsymbol{\alpha}} H_i J_{i,\boldsymbol{\alpha}})^{-1} J'_{i,\boldsymbol{\alpha}} H_i$, $H_i = \tilde{J}'_{i,\boldsymbol{\gamma}} A[\boldsymbol{\alpha}(\tau)] \tilde{J}_{i,\boldsymbol{\gamma}}$, $L_i = \tilde{J}'_{i,\tilde{\boldsymbol{\beta}}_i} M_i$, $M_i = I - J_{i,\boldsymbol{\alpha}} K_i$, $J_{i,\boldsymbol{\alpha}} = E[f_{\epsilon_i}(0|\tilde{\mathbf{x}}_i, W_i, D_i) \tilde{\mathbf{x}}_i D'_i]$, and $[\tilde{J}'_{i,\tilde{\boldsymbol{\beta}}_i}, \tilde{J}_{i,\boldsymbol{\gamma}}]$ is a partition of $J_{i,\psi} := (E[f_{\epsilon_i}(0|\tilde{\mathbf{x}}_i, W_i) \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}'_i])^{-1}$ such that $\tilde{J}'_{i,\tilde{\boldsymbol{\beta}}_i}$ is a $\dim(\tilde{\boldsymbol{\beta}}_i) \times \dim(\tilde{\boldsymbol{\beta}}_i, \boldsymbol{\gamma})$ matrix and $\tilde{J}_{i,\boldsymbol{\gamma}}$ is a $\dim(\boldsymbol{\gamma}) \times \dim(\tilde{\boldsymbol{\beta}}_i, \boldsymbol{\gamma})$ matrix.

Proof of Lemma A1. Lemma A1 implies the Proposition 2 of Chernozhukov and Hansen [11]. We verify the conditions for each individual i . Conditions R1 and R2 of Chernozhukov and Hansen [11] are implied by condition (A1) and condition (A2); condition R3 of Chernozhukov and Hansen [11] is implied by condition (A7); conditions R4-R6 of Chernozhukov and Hansen [11] are implied by conditions (A4)–(A6). Therefore, the lemma follows. \square

Proof of Theorem 1. As Lemma A1 turns out, $\widehat{\boldsymbol{\theta}}_i$ is consistent, $\widehat{\boldsymbol{\theta}}_i \xrightarrow{p} \tilde{\boldsymbol{\theta}}_i$. We can obtain that $\widehat{\boldsymbol{\theta}}_i \xrightarrow{p} \boldsymbol{\theta}$ as $T \rightarrow \infty$ for each i . Moreover, by condition (A10), $\widehat{V}_i \xrightarrow{p} V_i$ for each i , it follows that for fixed N , as $T \rightarrow \infty$

$$\widehat{\boldsymbol{\theta}}_{MD-IVQR} = \left(\sum_{i=1}^N \widehat{V}_i^{-1} \right)^{-1} \sum_{i=1}^N \widehat{V}_i^{-1} \widehat{\boldsymbol{\theta}}_i \xrightarrow{p} \left(\sum_{i=1}^N V_i^{-1} \right)^{-1} \sum_{i=1}^N V_i^{-1} \boldsymbol{\theta} = \boldsymbol{\theta}.$$

Thus, we can obtain that $\widehat{\boldsymbol{\theta}}_{MD-IVQR} \xrightarrow{p} \boldsymbol{\theta}$ as $(T, N)_{seq} \rightarrow \infty$. \square

Remark A1. As mentioned in Galvao and Wang [7], it is worth noting that condition (A10) is not really necessary. The convergence of \widehat{V}_i to any specific value is not crucial, as long as $\tilde{\boldsymbol{\theta}}_i$ remains consistent as T approaches infinity. This is because the equality on the right-hand side would hold under such circumstances.

Appendix A.2. Consistency of $\widehat{\boldsymbol{\theta}}_{MD-IVQR}$ under Joint Asymptotics

In order to establish Theorem 2, we first prove the following lemma.

Lemma A2. Under conditions (A1)–(A7), $\max_{1 \leq i \leq N} \|\hat{\theta}_i(\tau) - \theta(\tau)\| \xrightarrow{P} 0$ as $(N, T) \rightarrow \infty$ and $\frac{\log N}{T} \rightarrow 0$.

Proof of Lemma A2. Step 1 (Identification) Chernozhukov and Hansen [10,11] showed that $(\alpha(\tau), \tilde{\beta}_i(\tau))$ uniquely solves the limit problem $E[\tau - 1I(Y_i < \alpha D_i + X_i \beta + \eta_i) \tilde{x}_i(\tau)] = 0$ for each i and τ . That is to say, for each i and τ , $\alpha^*(\tau) = \alpha(\tau)$ and $\tilde{\beta}_i(\alpha^*(\tau), \tau) = \tilde{\beta}_i(\tau)$.

Step 2 (Consistency) For each α , recall $\bar{\theta}_i(\alpha) \equiv \arg \inf M_i(\alpha, \theta_i)$ and fix any $\delta > 0$. Let $B_i(\delta) := \{\theta : \|\theta - \bar{\theta}_i(\alpha)\|_1 \leq \delta\}$, the ball with center $\bar{\theta}_i(\alpha)$ and radius δ . For each $\theta_i \notin B_i(\delta)$, define $\tilde{\theta}_i = r_i \theta_i + (1 - r_i) \bar{\theta}_i(\alpha)$, where $r_i = \frac{\delta}{\|\theta_i - \bar{\theta}_i(\alpha)\|_1}$. So $\tilde{\theta}_i \in \partial B_i(\delta) := \{\theta : \|\theta - \bar{\theta}_i(\alpha)\|_1 = \delta\}$, the boundary of $B_i(\delta)$. Due to the convex nature of the objective function, we can deduce that

$$\begin{aligned} & r_i \{M_{iT}(\alpha, \theta_i) - M_{iT}(\alpha, \bar{\theta}_i(\alpha))\} \\ & \geq M_{iT}(\alpha, \tilde{\theta}_i) - M_{iT}(\alpha, \bar{\theta}_i(\alpha)) \\ & = E[\Delta_{iT}(\alpha, \tilde{\theta}_i)] + (\Delta_{iT}(\alpha, \tilde{\theta}_i) - E[\Delta_{iT}(\alpha, \tilde{\theta}_i)]) \\ & > \epsilon_\delta + (\Delta_{iT}(\alpha, \tilde{\theta}_i) - E[\Delta_{iT}(\alpha, \tilde{\theta}_i)]), \end{aligned} \tag{A1}$$

the last inequality is derived from the identity established by Knight [41], in conjunction with Condition (A8). Thus, we obtain the inclusion relation

$$\begin{aligned} & \{ \max_{1 \leq i \leq N} \sup_{\alpha} \|\hat{\theta}_i(\alpha) - \bar{\theta}_i(\alpha)\|_1 > \delta \} \\ & \stackrel{(a)}{\subseteq} \{ M_{iT}(\alpha, \theta_i) \leq M_{iT}(\alpha, \bar{\theta}_i(\alpha)), 1 \leq \exists i \leq N, \exists \theta_i \text{ s.t. } \sup_{\alpha} \|\theta_i - \bar{\theta}_i(\alpha)\|_1 > \delta \} \\ & \stackrel{(b)}{\subseteq} \bigcup_{i=1}^N \{ \sup_{\|\theta_i - \bar{\theta}_i(\alpha)\|_1 \leq \delta} |\Delta_{iT}(\alpha, \theta_i) - E[\Delta_{iT}(\alpha, \theta_i)]| \geq \epsilon_\delta \}. \end{aligned}$$

The validity of Relation (a) is established by the definition of $\hat{\theta}_i(\alpha) \equiv \arg \inf M_{iT}(\alpha, \theta_i)$. Relation (b) holds according to the inequality on the rightmost side of line (A1). Hence, we can conclude that

$$\begin{aligned} & P\{ \max_{1 \leq i \leq N} \sup_{\alpha} \|\hat{\theta}_i(\alpha) - \bar{\theta}_i(\alpha)\|_1 > \delta \} \\ & \leq P\{ \bigcup_{i=1}^N \sup_{\|\theta_i - \bar{\theta}_i(\alpha)\|_1 \leq \delta} |\Delta_{iT}(\alpha, \theta_i) - E[\Delta_{iT}(\alpha, \theta_i)]| \geq \epsilon_\delta \} \\ & \leq \sum_{i=1}^N P \sup_{\|\theta_i - \bar{\theta}_i(\alpha)\|_1 \leq \delta} |\Delta_{iT}(\alpha, \theta_i) - E[\Delta_{iT}(\alpha, \theta_i)]| \geq \epsilon_\delta \} \\ & \leq N \max_{1 \leq i \leq N} P\{ \sup_{\|\theta_i - \bar{\theta}_i(\alpha)\|_1 \leq \delta} |\Delta_{iT}(\alpha, \theta_i) - E[\Delta_{iT}(\alpha, \theta_i)]| \geq \epsilon_\delta \}. \end{aligned}$$

Hence, if we are able to demonstrate that

$$\max_{1 \leq i \leq N} P\{ \sup_{\|\theta_i - \bar{\theta}_i(\alpha)\|_1 \leq \delta} |\Delta_{iT}(\alpha, \theta) - E[\Delta_i(\alpha, \theta)]| \geq \epsilon_\delta \} = o(1/N).$$

the proof of the lemma is completed. Without loss of generality, we may assume that $\bar{\theta}_i(\alpha) = 0$ for each α and i . Then, $B_i(\delta)$ is independent of i and write $B_i(\delta) = B(\delta)$ for simplicity. In addition, the function $g(\theta) := \rho_\tau(u - \mathbf{x}'\theta) - \rho_\tau(u)$ has the following property, $|g(\theta_1) - g(\theta_2)| \leq CM\|\theta_1 - \theta_2\|$ for some fixed $C > 0$ by the identity of Knight (1998). Let $\kappa := CM$.

Since the closed ball $B(\delta)$ is compact, there exist K balls with centers $\vartheta^j, j = 1, \dots, K$, and radius $\frac{\epsilon}{3K}$ such that the collection of them covers B_δ . Therefore, for any $\vartheta \in B_\delta$, there is some $j \in 1, \dots, K$ such that

$$\begin{aligned} & |\Delta_{iT}(\alpha, \vartheta) - E\Delta_i(\alpha, \vartheta)| - |\Delta_{iT}(\alpha, \vartheta^j) - E\Delta_i(\alpha, \vartheta^j)| \\ & \leq |\Delta_{iT}(\alpha, \vartheta) - E\Delta_i(\alpha, \vartheta) - \Delta_{iT}(\alpha, \vartheta^j) + E\Delta_i(\alpha, \vartheta^j)| \\ & \leq |\Delta_{iT}(\alpha, \vartheta) - \Delta_{iT}(\alpha, \vartheta^j)| + |E\Delta_i(\alpha, \vartheta) - E\Delta_i(\alpha, \vartheta^j)| \\ & \leq CM\frac{\epsilon}{3K} + CM\frac{\epsilon}{3K} \\ & \leq \frac{2\epsilon}{3}. \end{aligned}$$

It then follows that for any $\epsilon > 0$, $\sup_{\|\vartheta_i - \bar{\vartheta}_i(\alpha)\| \leq \delta} |\Delta_{iT}(\alpha, \vartheta) - E\Delta_i(\alpha, \vartheta)| \leq \max_{1 \leq i \leq K} |\Delta_{iT}(\alpha, \vartheta^j) - E\Delta_i(\alpha, \vartheta^j)| + \frac{2\epsilon}{3}$, and

$$\begin{aligned} & P\left\{ \sup_{\|\vartheta_i - \bar{\vartheta}_i(\alpha)\| \leq \delta} |\Delta_{iT}(\alpha, \vartheta) - E\Delta_i(\alpha, \vartheta)| > \epsilon \right\} \\ & \leq P\left\{ \max_{1 \leq i \leq K} |\Delta_{iT}(\alpha, \vartheta^j) - E\Delta_i(\alpha, \vartheta^j)| + \frac{2\epsilon}{3} > \epsilon \right\} \\ & = P\left\{ \max_{1 \leq i \leq K} |\Delta_{iT}(\alpha, \vartheta^j) - E\Delta_i(\alpha, \vartheta^j)| > \frac{\epsilon}{3} \right\} \\ & \leq \sum_{i=1}^K P\left\{ |\Delta_{iT}(\alpha, \vartheta^j) - E\Delta_i(\alpha, \vartheta^j)| > \frac{\epsilon}{3} \right\}. \end{aligned}$$

Since the data are independently and identically distributed (i.i.d.) within each individual, it follows that the inequality on the rightmost side is less than or equal to

$$2K \exp\left\{-\frac{\epsilon^2 T}{18M^2 C^2 \delta^2}\right\} = O(\exp(-T)),$$

by Hoeffding’s inequality. Because $\frac{\log N}{T} \rightarrow 0$, it follows that $O(\exp(-T)) = o(1/N)$. Thus, we have

$$P\left\{ \max_{1 \leq i \leq N} \sup_{\alpha} \|\hat{\vartheta}_i(\alpha) - \bar{\vartheta}_i(\alpha)\|_1 > \delta \right\} = 0.$$

That is

$$\max_{1 \leq i \leq N} \sup_{\alpha} \|\hat{\vartheta}_i(\alpha) - \bar{\vartheta}_i(\alpha)\| \xrightarrow{P} o(1)(*)$$

i.e., $\max_{1 \leq i \leq N} \sup_{\alpha} \|\hat{\gamma}_i(\alpha) - \gamma(\alpha)\| \xrightarrow{P} o(1)$, which implies $\max_{1 \leq i \leq N} \sup_{\alpha} |W_{iT}(\alpha) - W(\alpha)| \xrightarrow{P} o(1)$ where $W(\alpha)$ is continuous in α over \mathcal{A} . It therefore follows by the consistency argument for extremum estimators that $\max_{1 \leq i \leq N} \hat{\alpha}_i \xrightarrow{P} \alpha$, which by (*) implies that $\max_{1 \leq i \leq N} \|\hat{\eta}_i - \eta_i\| \xrightarrow{P} 0$, $\max_{1 \leq i \leq N} \|\hat{\beta}_i - \beta\| \xrightarrow{P} 0$ and $\max_{1 \leq i \leq N} \|\hat{\gamma}_i(\hat{\alpha}_i) - 0\| \xrightarrow{P} 0$. Therefore, $\max_{1 \leq i \leq N} \|\hat{\theta}_i - \bar{\theta}_i\| \xrightarrow{P} 0$ as $(N, T) \rightarrow \infty$ and the lemma follows. \square

Having Lemma A2 at our disposal, we proceed to prove Theorem 2.

Proof of Theorem 2. It follows that

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{MD-IVQR} - \boldsymbol{\theta} &= \left(\sum_{i=1}^N \widehat{V}_i^{-1}\right)^{-1} \sum_{i=1}^N \widehat{V}_i^{-1}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}) \\ &= \left(\sum_{i=1}^N \widehat{V}_i^{-1}\right)^{-1} \sum_{i=1}^N \widehat{V}_i^{-1}(o_p(1)) \\ &= o_p(1). \end{aligned}$$

The last equality holds because $\max_{1 \leq i \leq N} \|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\| = o_p(1)$ by Lemma A2. \square

Appendix A.3. Asymptotic Normality of $\widehat{\boldsymbol{\theta}}_{MD-IVQR}$ under Sequential Asymptotics

Proof of Theorem 3. As the Proposition 2 of Chernozhukov and Hansen [11], for each individual i ,

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i) \xrightarrow{d} N(0, \widetilde{V}_i), \quad \widetilde{V}_i = (K_i', L_i')' S_i (K_i', L_i').$$

Define $\Xi_1 = [I_{k \times k} | 0_{k \times 1}]$, where $k = \dim(\boldsymbol{\alpha}) + \dim(\boldsymbol{\beta}) = \dim(\boldsymbol{\theta})$. We begin by fixing N and allowing T to approach infinity. Consequently, it follows that

$$\begin{aligned} \sqrt{NT}(\widehat{\boldsymbol{\theta}}_{MD-IVQR} - \boldsymbol{\theta}) &= \left(\frac{1}{N} \sum_{i=1}^N \widehat{V}_i^{-1}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \widehat{V}_i^{-1} \Xi_1 \sqrt{T}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i) \\ &\xrightarrow{d} \left(\frac{1}{N} \sum_{i=1}^N V_i^{-1}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N V_i^{-1} \Xi_1 N(0, \widetilde{V}_i) \\ &= \left(\frac{1}{N} \sum_{i=1}^N V_i^{-1}\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N N(0, V_i^{-1}). \end{aligned}$$

The second line in the display above holds because $\sqrt{T}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i) \xrightarrow{d} N(0, \widetilde{V}_i)$ and $\widehat{V}_i \xrightarrow{p} V_i$ for each i as $T \rightarrow \infty$ by Condition (A10) and Slutsky’s theorem. The third line holds because $\Xi_1 \sqrt{T}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i) = \sqrt{T}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta})$.

Now let N tend to infinity, and we obtain $\left(\frac{1}{N} \sum_{i=1}^N V_i^{-1}\right)^{-1} \xrightarrow{p} V$. Moreover, by Lyapunov Central Limit Theorem, it follows that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N N(0, V_i^{-1}) \xrightarrow{d} N(0, V^{-1}).$$

Hence, by Slutsky Theorem, we obtain the desired result

$$\sqrt{NT}(\widehat{\boldsymbol{\theta}}_{MD-IVQR} - \boldsymbol{\theta}) \xrightarrow{d} N(0, V).$$

\square

Appendix A.4. Asymptotic Normality of $\widehat{\boldsymbol{\theta}}_{MD-IVQR}$ under Joint Asymptotics

Let $\boldsymbol{q}_i = (\boldsymbol{\alpha}', \boldsymbol{\beta}', \eta_i, \gamma')'$, $\boldsymbol{x}_{it}^* = (\boldsymbol{d}_{it}', \boldsymbol{x}_{it}', 1, \boldsymbol{w}_{it}')'$, denote $\mathbb{G}_{iT}(\boldsymbol{q}_i) := \frac{1}{T} \sum_{t=1}^T \psi_\tau(y_{it} - \boldsymbol{x}_{it}^{*'} \boldsymbol{q}_i) \boldsymbol{x}_{it}^*$, $G_i(\boldsymbol{q}_i) := E(\mathbb{G}_{iT}(\boldsymbol{q}_i))$, where $\varphi_\tau := \tau - \mathbf{1}\{u \leq 0\}$. According to Lemma B.2 of Chernozhukov and Hansen [10], we know that for each i and for any $\sup_{\tau} \|\widehat{\boldsymbol{q}}_i - \boldsymbol{q}_i\| = o_p(1)$, it is the case that $\sup_{\tau} \|\mathbb{G}_{iT}(\widehat{\boldsymbol{q}}_i) - \mathbb{G}_{iT}(\boldsymbol{q}_i)\| - [G_i(\widehat{\boldsymbol{q}}_i) - G_i(\boldsymbol{q}_i)] = o_p(1)$. Then the following Lemma gives the order of $\max_{1 \leq i \leq N} [\mathbb{G}_{iT}(\widehat{\boldsymbol{q}}_i) - \mathbb{G}_{iT}(\boldsymbol{q}_i)] - [G_i(\widehat{\boldsymbol{q}}_i) - G_i(\boldsymbol{q}_i)]$.

Lemma A3. If $\max_{1 \leq i \leq N} \|\hat{q}_i - q_i\| = O_p(\delta_N)$, where $\lim_{N \rightarrow \infty} \delta_N = 0$, then under then under Conditions (A1)–(A9) and (A10'), we have $\max_{1 \leq i \leq N} [\mathbb{G}_{iT}(\hat{q}_i) - \mathbb{G}_{iT}(q_i)] - [G_i(\hat{q}_i) - G_i(q_i)] = O_p(d_N)$, where $d_N = \frac{\log \delta_N}{T} \vee \sqrt{\frac{\delta_N |\log \delta_N|}{T}}$.

Proof of Lemma A3. Without loss of generality, we assume that $q_i = 0$ for all i . Let $g_\varrho(\mathbf{X}_{it}^*) = [\mathbf{1}\{u_{it} \leq 0\} - \mathbf{1}\{u_{it} - \tilde{\mathbf{x}}_{it}' \varrho \leq 0\}] \tilde{\mathbf{x}}_{it}'$, where $\mathbf{X}_{it}^* = (u_{it}, \tilde{\mathbf{x}}_{it}')$. Then, we have

$$\begin{aligned} \mathbb{G}_{iT}(\hat{q}_i) - \mathbb{G}_{iT}(0) &= \frac{1}{T} \sum_{t=1}^T (\tau - \mathbf{1}\{y_{it} - \tilde{\mathbf{x}}_{it}' \hat{q}_i \leq 0\}) \tilde{\mathbf{x}}_{it}' - \frac{1}{T} \sum_{t=1}^T (\tau - \mathbf{1}\{y_{it} \leq 0\}) \tilde{\mathbf{x}}_{it}' \\ &= -\frac{1}{T} \sum_{t=1}^T [\mathbf{1}\{u_{it} - \tilde{\mathbf{x}}_{it}' \hat{q}_i \leq 0\} - \mathbf{1}\{u_{it} \leq 0\}] \tilde{\mathbf{x}}_{it}' \\ &= \frac{1}{T} \sum_{t=1}^T g_{\hat{q}_i}(\mathbf{X}_{it}^*). \end{aligned}$$

So, the claim to be proved becomes $\max_{1 \leq i \leq N} \sum_{t=1}^T [g_{\hat{q}_i}(\mathbf{X}_{it}^*) - E[g_{\hat{q}_i}(\mathbf{X}_{it}^*) | \varrho = \hat{q}_i]] = O_p(Td_n)$. Let $\tilde{g}_{i\delta_N} = \{g_{\hat{q}_i}(\mathbf{X}_{it}^*) - E[g_{\hat{q}_i}(\mathbf{X}_{it}^*)] : \|\varrho\| \leq \delta_N\}$. Thus, we need to show that $\max_{1 \leq i \leq N} EZ_i = O_p(Td_n)$, where $Z_i := \sum_{t=1}^T g(\mathbf{X}_{it}^*) \|\tilde{g}_{i\delta_N}\}$.

In order to use Proposition B.1. of Kato et al. [42], we verify the conditions for $\tilde{g}_{i\delta_N}$ of Proposition B.1. of Kato et al. [42]. First, it is pointwise measurable, and is bounded by $8M$. Since the $\mathfrak{g}_\infty := \{g_\varrho : \varrho \in \mathbb{R}^p\}$ class is a VC subgraph class and $\tilde{g}_{i\delta_N} \subset \{g - E[g(\mathbf{X}_{it}^*)] : g \in \mathfrak{g}_\infty\}$. By Lemma 2.6.15 of van der Vaart and Wellner [43], there exist constant $A \geq 3\sqrt{r}$ and $v \geq 1$ independent of i and N , for every $0 < \epsilon < 1$ and every probability measure Q , $N(8M\epsilon, \tilde{g}_{i\delta_N}, L_2(Q)) \leq (\frac{A}{\epsilon})^v$ holds. Moreover, $Eg_\varrho(\mathbf{X}_{it}^*)^2 = E[F_i(\tilde{\mathbf{x}}_{it}' \varrho_i | \tilde{\mathbf{x}}_{it}^*) - F_i(0 | \tilde{\mathbf{x}}_{it}^*) | \tilde{\mathbf{x}}_{it}^{*2}]$ is bounded by $8C_u M^3 \delta_N$. That is to say, $\tilde{g}_{i\delta_N}$ satisfies all the conditions of Proposition B.1. of Kato et al. [42] with $U = 8M + 8C_u T M^3 \delta_N$ and $\sigma^2 = 8C_u M^3 \delta_N$. So, we have $\max_{1 \leq i \leq N} EZ_i = O_p(Td_N)$. \square

Lemma A4. Under condition of Lemma A3, we have $\max_{1 \leq i \leq N} \|\hat{q}_i - q_i\| = O_p(\sqrt{\frac{\log N}{T}})$.

Proof of Lemma A4. $G_i(\hat{q}_i)$ around q_i and by Lemma 2.12 of van der Vaart [44], we have

$$G_i(\hat{q}_i) = G_i(q_i) + \frac{\partial G_i(q_i)}{\partial q_i} (\hat{q}_i - q_i) + O_p((\hat{q}_i - q_i)^2).$$

Notice that $\tilde{\gamma}_i = \frac{\partial G_i(q_i)}{\partial q_i}$, it then follows that

$$\begin{aligned} \hat{q}_i - q_i &= \tilde{\gamma}_i^{-1} (G_i(\hat{q}_i) - G_i(q_i) + O_p((\hat{q}_i - q_i)^2)) \\ &= \tilde{\gamma}_i^{-1} (-\mathbb{G}_{iT}(\hat{q}_i) - (G_i(\hat{q}_i) - G_i(q_i)) - (\mathbb{G}_{iT}(\hat{q}_i) - \mathbb{G}_{iT}(q_i)) + \mathbb{G}_{iT}(\hat{q}_i) + O_p((\hat{q}_i - q_i)^2)) \\ &= -\tilde{\gamma}_i^{-1} \mathbb{G}_{iT}(\hat{q}_i) - \tilde{\gamma}_i^{-1} [(\mathbb{G}_{iT}(\hat{q}_i) - \mathbb{G}_{iT}(q_i)) - (G_i(\hat{q}_i) - G_i(q_i))] + \tilde{\gamma}_i^{-1} \mathbb{G}_{iT}(\hat{q}_i) \\ &\quad + \tilde{\gamma}_i^{-1} O_p((\hat{q}_i - q_i)^2). \end{aligned}$$

Because of the computational property of instrument variable quantile regression estimators, $\mathbb{G}_{iT}(\hat{q}_i) = O_p(\frac{1}{T})$ for each i , therefore by Lemma 3 of Galvao and Wang [7], $\max_{1 \leq i \leq N} \|\hat{q}_i - q_i\|$ is bounded by

$$const. \times [\max_{1 \leq i \leq N} \|\mathbb{G}_{iT}(\hat{q}_i)\| + \max_{1 \leq i \leq N} \|(\mathbb{G}_{iT}(\hat{q}_i) - \mathbb{G}_{iT}(q_i)) - (G_i(\hat{q}_i) - G_i(q_i))\|] + O_p(\frac{1}{T})$$

with probability approaching one.

We first show that $\max_{1 \leq i \leq N} \|\mathbb{G}_{iT}(\hat{q}_i)\|$ is bounded. First, For any fixed $K > 0$,

$$P\{\max_{1 \leq i \leq N} \|G_i(q_i)\| > \sqrt{\frac{\log N}{T}}K\} \leq \sum_{i=1}^N P\{\|G_i(q_i)\| > \sqrt{\frac{\log N}{T}}K\} \leq 2N^{1-\frac{K^2}{8M^2}}$$

by Hoeffding’s inequality. Hence we obtain $\max_{1 \leq i \leq N} \|\mathbb{G}_{iT}(\hat{q}_i)\| = O_p(\sqrt{\frac{\log N}{T}})$.

Then, we show that

$$\max_{1 \leq i \leq N} \|(\mathbb{G}_{iT}(\hat{q}_i) - \mathbb{G}_{iT}(q_i)) - (G_i(\hat{q}_i) - G_i(q_i))\| = o_p(\sqrt{\frac{\log N}{T}}).$$

Like Lemma A3, without loss of generality, we set $q_i = 0$ for $1 \leq i \leq N$. Thus, it is sufficient to present that

$$\max_{1 \leq i \leq N} \|\frac{1}{T} \sum_{t=1}^T [g_{\hat{q}_i}(\mathbf{X}_{it}^*) - E[g_{\hat{q}}(\mathbf{X}_{it}^*)|_{q=\hat{q}_i}]]\| = o_p(\sqrt{\frac{\log N}{T}}).$$

The above claim is equivalent to that for any $\epsilon > 0$,

$$P(\max_{1 \leq i \leq N} \|\frac{1}{T} \sum_{t=1}^T g_{\hat{q}_i}(\mathbf{X}_{it}^*) - E[g_{\hat{q}}(\mathbf{X}_{it}^*)|_{q=\hat{q}_i}]\| > \epsilon \sqrt{\frac{\log N}{T}}) = o(1).$$

So, it is worthwhile to prove that

$$\max_{1 \leq i \leq N} P(\|\frac{1}{T} \sum_{t=1}^T g_{\hat{q}_i}(\mathbf{X}_{it}^*) - E[g_{\hat{q}}(\mathbf{X}_{it}^*)|_{q=\hat{q}_i}]\| > \epsilon \sqrt{\frac{\log N}{T}}) = o(\frac{1}{N}). \quad (\#)$$

Using Proposition B.2. of Kato et al. [42], by setting $s = \sqrt{2 \log N}$, we obtain

$$P\{Z_i \geq EZ_i + \sqrt{4 \log N \{8TC_u M^3 \delta_N + (16M + 16TC_u M^3 \delta_N)EZ_i\}} + \frac{2 \log N}{3} (8M + 8TC_u M^3 \delta_N)\} \leq \frac{1}{N^2},$$

where $Z_i := \|\sum_{t=1}^T g(\mathbf{X}_{it}^*)\|_{\tilde{g}_i \delta_N}$. Therefore,

$$P\{Z_i \geq \max_{1 \leq i \leq N} EZ_i + \sqrt{4 \log N \{8TC_u M^3 \delta_N + (16M + 16TC_u M^3 \delta_N) \max_{1 \leq i \leq N} EZ_i\}} + \frac{(16M + 16TC_u M^3 \delta_N) \log N}{3}\} \leq \frac{1}{N^2},$$

It is worth noting that Lemma A3 demonstrates that $\max_{1 \leq i \leq N} EZ_i \leq const. \times (\log \delta_N + \sqrt{T \delta_N \log \delta_N})$, thus for fixed $\epsilon > 0$, we can find δ_N sufficiently small and n_0 such that when $N \geq n_0$, $\max_{1 \leq i \leq N} EZ_i + \sqrt{4 \log N \{8TC_u M^3 \delta_N + (16M + 16TC_u M^3 \delta_N) \max_{1 \leq i \leq N} EZ_i\}} + \frac{2 \log N}{3} (8M + 8TC_u M^3 \delta_N) \leq \epsilon \sqrt{T \log N}$ holds. That is, for any $\epsilon > 0$, there is a sufficiently small $\delta > 0$ such that

$$P(\|\sum_{t=1}^T g(\mathbf{X}_{it}^*)\|_{\tilde{g}_i \delta_N} > \epsilon \sqrt{T \log N}) = o(\frac{1}{N}).$$

Given that the IVQR estimators exhibit uniform consistency, we can obtain (#). Therefore we have $\max_{1 \leq i \leq N} \|\hat{q}_i - q_i\| = O_p(\sqrt{\frac{\log N}{T}})$. \square

Remark A2. By Lemmas A3 and A4 we have proved, we have that

$$d_N = \frac{(\log N)^{1/4} \sqrt{|\log \delta_N|}}{T^3/4}.$$

Proof of Theorem 4. Let $\Xi := [I_{k \times k} | 0_{k \times (q+1)}]$, where $k = \dim(\alpha) + \dim(\beta) = \dim(\theta)$ and $q = \dim(\gamma)$, and consider the following

$$\begin{aligned} & \sqrt{NT}(\theta_{MD-IVQR} - \theta) \\ &= \left(\frac{1}{N} \sum_{i=1}^N \hat{V}_i^{-1}\right)^{-1} \left(\frac{\sqrt{N}}{N} \sum_{i=1}^N \hat{V}_i^{-1} \Xi \sqrt{T}(\hat{q}_i - q_i)\right) \\ &= \left(\left(\frac{1}{N} \sum_{i=1}^N \hat{V}_i^{-1}\right)^{-1} + O_p(T^{-1/2}h_N^{-1/2})\right) \left(\frac{\sqrt{N}}{N} \sum_{i=1}^N V_i^{-1} \Xi \sqrt{N}(q_i - q_i) + O_p(N^{-1/2}T^{-1/2}h_N^{-1/2})\right) \\ &= \left(\frac{1}{N} \sum_{i=1}^N V_i^{-1}\right)^{-1} \frac{\sqrt{N}}{N} \sum_{i=1}^N V_i^{-1} \Xi \sqrt{N}(\hat{q}_i - q_i) + O_p(N^{-1/2}T^{-1/2}h_N^{-1/2}). \end{aligned}$$

By Lemma A4, we have for each i ,

$$\begin{aligned} \hat{q}_i - \theta &= -\Xi \tilde{\gamma}_i^{-1} \mathbb{G}_{iT}(q_i) + O_p(d_N) + O_p\left(\frac{1}{T}\right) + \Xi \tilde{\gamma}_i^{-1} O_p((\hat{q}_i - q_i)^2) \\ &= -\Xi \tilde{\gamma}_i^{-1} \mathbb{G}_{iT}(\theta_i) + O_p(d_N). \end{aligned}$$

Therefore, it follows that

$$\frac{\sqrt{N}}{N} \sum_{i=1}^N V_i^{-1} \Xi \sqrt{T}(\hat{q}_i - q_i) = -\frac{\sqrt{N}}{N} \sum_{i=1}^N V_i^{-1} \Xi \sqrt{T} \tilde{\gamma}_i^{-1} \mathbb{G}_{iT}(q_i) + O_p(\sqrt{NT}d_N).$$

The second term is $o_p(1)$ according to the assumption regarding the relative rates of N and T in the theorem. As for the first term, applying the Lyapunov central limit theorem implies that it converges in distribution to $N(0, \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N V_i^{-1})$. Consequently, by utilizing Slutsky’s theorem and condition (A10’),

$$\sqrt{NT}(\hat{\theta}_{MD-IVQR} - \theta) \xrightarrow{d} N(0, V).$$

as $(N, T) \rightarrow \infty$ and $\sqrt{NT}d_N \rightarrow 0$. \square

References

1. Koenker, R. Quantile regression for longitudinal data. *J. Multivar. Anal.* **2004**, *1*, 74–89. [CrossRef]
2. Lamarche, C. Robust penalized quantile regression estimation for panel data. *J. Econom.* **2010**, *157*, 396–408. [CrossRef]
3. Gu, A.J.; Volgushev, S. Panel data quantile regression with grouped fixed effects. *J. Econom.* **2019**, *213*, 68–91. [CrossRef]
4. Galvao, A.F.; Parker, T.; Xiao, Z. Bootstrap Inference for Panel Data Quantile Regression. *arXiv* **2021**, arXiv:2111.03626.
5. Canay, I.A. A simple approach to quantile regression for panel data. *Economet. J.* **2011**, *14*, 368–386. [CrossRef]
6. Chen, L.; Huo, Y. A simple estimator for quantile panel data models using smoothed quantile regressions. *Economet. J.* **2021**, *24*, 247–263. [CrossRef]
7. Galvao, A.F.; Wang, L. Efficient minimum distance estimator for quantile regression fixed effects panel data. *J. Multivar. Anal.* **2015**, *133*, 1–26. [CrossRef]
8. Galvao, A.F.; Gu, J.; Volgushev, S. On the unbiased asymptotic normality of quantile regression with fixed effects. *J. Econom.* **2020**, *218*, 178–215. [CrossRef]
9. Chernozhukov, V.; Hansen, C. An IV Model of quantile treatment effects. *Econometrica* **2005**, *73*, 245–261. [CrossRef]

10. Chernozhukov, V.; Hansen, C. Instrumental quantile regression inference for structural and treatment effect models. *J. Econom.* **2006**, *132*, 491–525. [[CrossRef](#)]
11. Chernozhukov, V.; Hansen, C. Instrumental variable quantile regression: A robust inference approach. *J. Econom.* **2008**, *142*, 379–398. [[CrossRef](#)]
12. Koenker, R.; Bassett, G. Regression quantile. *Econometrica* **1978**, *46*, 33–50. [[CrossRef](#)]
13. Harding, M.; Lamarche, C. Quantile regression approach for estimating panel data models using instrumental variables. *Econ. Lett.* **2009**, *104*, 133–135. [[CrossRef](#)]
14. Galvao, A.F.; Montes-Rojas, G. *Instrumental Variables Quantile Regression for Panel Data with Measurement Errors*; Working Papers; Department of Economics, City University London: London, UK, 2009.
15. Galvao, A.F.; Montes-Rojas, G. Penalized quantile regression for dynamic panel Data. *J. Stat. Plan. Infer.* **2010**, *140*, 3476–3497. [[CrossRef](#)]
16. Galvao, A.F. Quantile regression for dynamic panel data with fixed effects. *J. Econom.* **2011**, *164*, 142–157. [[CrossRef](#)]
17. Harding, M.; Lamarche, C. Hausman-Taylor instrumental variable approach to the penalized estimation of quantile panel models. *Econ. Lett.* **2014**, *124*, 176–179. [[CrossRef](#)]
18. Zhang, Y.; Jiang, J.; Feng, Y. Penalized quantile regression for spatial panel data with fixed effects. *Commun. Stat.-Theory Methods* **2021**, *52*, 1287–1299. [[CrossRef](#)]
19. Besstremyannaya, G.; Golovan, S. Measuring heterogeneity with fixed effect quantile regression: Long panels and short panels. *Appl. Econom.* **2021**, *64*, 70–82.
20. Ponomareva, M. *Quantile Regression for Panel Data Models with Fixed Effects and Small T: Identification and Estimation*; Working Paper; Northwestern University: Evanston, IL, USA, 2010.
21. Luo, Y.; Lian, H.; Tian, M. Bayesian quantile regression for longitudinal data models. *J. Stat. Comput. Simul.* **2012**, *82*, 1635–1649. [[CrossRef](#)]
22. Galvao, A.F.; Lamarche, C.; Lima, L.R. Estimation of censored quantile regression for panel data with fixed effects. *J. Am. Stat. Assoc.* **2013**, *108*, 1075–1089. [[CrossRef](#)]
23. Chetverikov, D.; Larsen, B.; Palmer, C. IV quantile regression for group-level treatments, with an application to the distributional effects of trade. *Econometrics* **2016**, *84*, 809–833. [[CrossRef](#)]
24. Contoyannis, P.; Li, J. The dynamics of adolescent depression: An instrumental variable quantile regression with fixed effects approach. *J. R. Stat. Soc. B* **2017**, *180*, 907–922. [[CrossRef](#)]
25. Rao, C.R. *Linear Statistical Inference and Its Applications*, 2nd ed.; Wiley: New York, NY, USA, 1965.
26. Serfling, R.J. *Approximation Theorems of Mathematical Statistics*; Wiley: New York, NY, USA, 1980.
27. Hsiao, C. *Analysis of Panel Data*; Cambridge University Press: New York, NY, USA, 2003.
28. Koenker, R. Confidence intervals for regression quantiles. In *Asymptotic Statistics*; Physica-Verlag HD: Heidelberg, Germany, 1994; pp. 349–359.
29. Kato, K. Asymptotic normality of Powell’s kernel estimator. *Ann. Inst. Stat. Math.* **2012**, *64*, 255–273. [[CrossRef](#)]
30. Galvao, A.F. *Essays on Quantile Regression for Dynamic Panel Data Models*. Ph.D. Dissertation, University of Illinois at Urbana-Champaign, Urbana, IL, USA, 2009.
31. William, H. Greene. *Econometric Analysis*, 8th ed.; Pearson Education Limited: New York, NY, USA, 2018.
32. Koenker, R. *Quantile Regression*; Cambridge University Press: New York, NY, USA, 2005.
33. Newhouse, J.P. Medical care expenditures: A cross-national survey. *J. Hum. Resour.* **1977**, *12*, 115–125. [[CrossRef](#)]
34. Gerdtham, U.; Jönsson, B. International comparisons of health expenditure: Theory, data and econometric analysis. In *Handbook of Health Economics*; Elsevier: Amsterdam, The Netherlands, 2000; Volume 1, pp. 11–53.
35. Moran, J.R.; Simon, K.I. Income and the use of prescription drugs by the elderly. *J. Hum. Resour.* **2005**, 411–432. [[CrossRef](#)]
36. Murphy, K.M.; Robert, H.T. The value of health and longevity. *J. Political Econ.* **2006**, *5*, 871–904. [[CrossRef](#)]
37. Farag, M.; Nandakumar, A.K.; Wallack, S.; Hodgkin, D.; Gaumer, G.; Erbil, C. The income elasticity of health care spending in developing and developed countries. *Int. J. Health Care Financ. Econ.* **2012**, *12*, 145–162. [[CrossRef](#)]
38. Acemoglu, D.; Finkelstein, A.; Notowidigdo, M.J. Income and health spending: Evidence from oil price shocks. *Rev. Econ. Stat.* **2013**, *95*, 1079–1095. [[CrossRef](#)]
39. Tian, F.; Gao, J.; Yang, K. A quantile regression approach to panel data analysis of health care expenditure in OECD countries. *Health Econ.* **2013**, *27*, 1921–1944. [[CrossRef](#)]
40. Bruckner, M.; Ciccone, A.; Tesei, A. Oil price shocks, income, and democracy. *Rev. Econ. Stat.* **2013**, *94*, 389–399. [[CrossRef](#)]
41. Knight, K. Limiting distributions for L_1 regression estimators under general conditions. *Ann. Stat.* **2013**, *26*, 755–770.
42. Kato, K.; Galvao, A.F.; Montes-Rojas, G.V. Asymptotics for panel quantile regression models with individual effects. *J. Econom.* **2013**, *170*, 76–91. [[CrossRef](#)]
43. van der Vaart, A.W.; Wellner, J.A. *Weak Convergence and Empirical Processes*; Springer: New York, NY, USA, 1996.
44. van der Vaart, A.W. *Asymptotic Statistics*; Cambridge University Press: New York, NY, USA, 1998.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.