May 2013

Three Essays on Quantile Regression

Liang Wang
University of Wisconsin-Milwaukee

THREE ESSAYS ON QUANTILE REGRESSION

by

Liang Wang

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in

Economics

at

The University of Wisconsin-Milwaukee

May 2013

ABSTRACT

THREE ESSAYS ON QUANTILE REGRESSION

by

Liang Wang

The University of Wisconsin-Milwaukee, 2013
Under the Supervision of Professor Antonio F. Galvao

The first chapter studies identification, estimation, and inference of general unconditional treatment effects models with continuous treatment under the ignorability assumption. We show identification of dose-response functions under the assumption that selection to treatment is based on observables. We consider estimation of dose-response functions through moment restriction models with generalized residual functions which are possibly non-smooth, and propose a semiparametric two-step estimator. This general formulation includes average and quantile treatment effects as special cases. The asymptotic properties of the estimator are derived. We also develop statistical inference procedures and show the validity of a bootstrap approach to implement these methods in practice. Monte Carlo simulations demonstrate that the test statistics have good finite sample properties. Finally, we apply the proposed methods to estimate unconditional average and quantile effects of mothers' weight gain and age on birthweight.

The second chapter develops a new minimum distance quantile regression (MD-QR) estimator for panel data models with fixed effects. We establish consistency and derive the limiting distribution of the MD-QR estimator for panels with a large number of cross-sections and time-series. The limit theory allows for both sequential and joint limits. The proposed estimator is efficient in the class of minimum distance estimators. In addition, the MD-QR estimator is computationally fast, especially

for large cross-sections. Monte Carlo simulations are conducted to evaluate finite sample performance. Finally, we illustrate the use of the estimator with a simple application to the investment equation model.

The third chapter proposes tests for slope homogeneity across individuals in quantile regression fixed effects panel data models. The tests are based on the Swamy statistic. We establish the asymptotic null distribution of the tests under large panel data, with sequential and joint limits. Monte Carlo experiments show good performance of the proposed tests in finite samples in terms of size and power. Finally, we test and reject the hypothesis of homogeneous speed of capital structure adjustment across firms using a panel dataset.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First, I would like to thank Dr. Antonio F. Galvao for advising me on this project.

Also, I would like to thank the committee members Dr. Scott Adams, Dr. Chuan Goh, Dr. Suyong Song, and Dr. Zhijie Xiao.

Finally, I would like to thank Dr. Ted Juhl for valuable discussions.

Chapter 1

Uniformly Semiparametric Efficient Estimation of Treatment Effects with a Continuous Treatment

1.1 Introduction

This chapter studies identification, estimation, and inference of general unconditional treatment effect (TE) models with a continuous dose of treatment. We consider estimating the parameters of interest, dose-response functions (DRF), through moment restriction models in which generalized residual functions are possibly nonsmooth. In this general formulation, the DRF include mean and quantile functions as special cases, and consequently average treatment effects (ATE) and quantile treatment effects (QTE) are direct applications of the methods developed in this chapter.

In this chapter, the ignorability assumption is used to achieve identification of parameters of interest. The ignorability assumption states that "given a set of observed covariates, each individual is randomly assigned either to the treatment group or to the control group"; see Firpo (2007). This condition has been largely employed in TE literature, see e.g. Rubin (1977), Barnow et al. (1980), Heckman et al. (1998), Dehejia and Wahba (1999), Firpo (2007), Flores (2007), Angrist and Pischke (2009), and Cattaneo (2010) for a review.

Based on the identification condition, we construct a two-step estimation procedure. The implementation of the estimator in practice is simple. In the first step, one estimates a ratio of two conditional distributions, which is similar to a propensity score. In the second step, an optimization problem is solved. It is important to note that, once the identification is achieved and a DRF is estimated, other parameters of interest based on these functions can be estimated with little additional effort. For example, one can easily estimate TE, which are defined as differences of the DRF evaluated at different levels of treatment. In addition, one could estimate the entire curve of potential outcomes or the DRF.

Mild sufficient conditions are provided for the two-step estimator to have desired asymptotic properties, namely, consistency, weak convergence, and semiparametric efficiency. In particular, we show that the two-step estimator of a DRF is uniformly consistent over a set of treatment. Different from the binary or multi-valued treatment models, in which case pointwise results are equivalent to uniform results, when treatment levels are an interval $\mathcal{T}$, the uniform results are stronger than pointwise results, and consequently, only pointwise results are often not adequate for inferences. In addition, we show that the estimator converges weakly to a Gaussian process, and that it is uniformly semiparametric efficient. For the latter derivation we use the method of Bickel et al. (1993).

Technically, the derivations of the asymptotic properties for the proposed estimator are independently interesting. Because the semiparametric model considered encompasses continuous treatment levels and nuisance parameters, both of which are infinit dimensional, existing results available in the literature are not directly applicable. Therefore, an additional contribution of this chapter is to provide sufficient conditions for consistency and weak convergence of generic moment restriction estimators (Z-estimators) with possibly non-smooth functions and a nuisance parameter, when both the parameter of interest and the nuisance parameter are possibly infinite dimensional. These general results are used to prove the asymptotic properties of the two-step estimator discussed above. In this general setting, the data need not be independent and identically distributed (i.i.d.). These results extend those of Chen et al. (2003) in that the parameter of interest is not in a Euclidean space but in a generic Banach space. Moreover, the results extend Theorem 3.3.1 of van der Vaart and Wellner (1996) in that a possibly infinite dimensional nuisance parameter needs to be estimated in the first step. This is an important innovation because it facilitates the derivation of the limiting results for general Z-estimators, and can be utilized in future works for other statistical models.

In addition, we develop statistical inference procedures based on the two-step estimator. In particular, we conduct inference on a DRF uniformly over the treatment levels. We propose testing procedures for the hypothesis of the equality of a DRF and any given function. The test statistics used are Kolmogorov and Cramér-von Mises types, which detect any deviation of the null hypothesis. Since the parameter of interest is infinite dimensional and the weak limit of these statistics are not standard, we compute critical values using a bootstrap method. We provide sufficient conditions under which the bootstrap is valid, and discuss an algorithm for practical implementation. The proof of the validity of the bootstrap is also an extension of that in Chen et al. (2003).

We conduct Monte Carlo simulations to evaluate finite sample performance of the test statistics. The simulations show that the Cramér-von Mises type test statistic has good empirical size and high power against a few alternatives. In addition, the result is improved when the sample size increases, and is not sensitive to the selected numbers of bootstrap.

To illustrate the proposed methods, we consider an empirical application to a birthweight study using data from the National Vital Statistics System of Centers for Disease Control and Prevention. We estimate unconditional average and quantile dose-birthweight function for mothers' weight gain during pregnancy as well as age separately. The empirical results document important heterogeneity in the dose-birthweight functions for mothers' weight gain during pregnancy and age across quantiles. The findings provide evidence that, in general, more weight gain during pregnancy leads to higher birthweight. However, the treatment effects differ at different levels of weight gain. For a given quantile of interest, positive impacts are larger for low and high weight gains while relatively lower in the middle range of weight gain. The quantile dose-birthweight functions of the mother's age on birthweight is downward-sloping. In addition, for a given age, this impact becomes more severe for lower parts of the distribution of birthweight. Although intuitive, this result complements the existing results in the literature.[1]

1.1.1 Literature and Outline

There is large and growing literature on unconditional TE, most of which focuses on models with discrete (usually binary) treatment levels. Hahn (1998), Heckman et al. (1998), and Imbens et al. (2006) study efficient estimation of the average treatment effect nonparametrically. To estimate the average treatment effect, Hirano et al. (2003) estimate propensity scores nonparametrically first while Abadie and Imbens (2006) apply matching methods. In addition, Li et al. (2009) propose "efficient estimation of average treatment effects with mixed categorical and continuous data." The study of unconditional average TE has been extended to the quantile framework by Firpo (2007) with a two-step estimator that is semiparametric efficient. This method, as that of Hirano et al. (2003), is based on nonparametric estimation of propensity score in the first step. There is also literature on multi-valued treatment effect models. Imbens (2000) shows that the multi-valued counterpart of the propensity score theorem of Rosenbaum and Rubin (1983) still holds. Imbens (2000) and Lechner (2001) discuss the unconditional mean treatment effect. Cattaneo (2010) extends the literature and proposes semiparametric efficient estimation

[1]Previous approaches to estimating birthweight outcome using quantile regressions have employed reduced form models and, therefore, cannot be interpreted as causal effects. For instance, Abrevaya (2001) (see also Koenker and Hallock (2001) and Chernozhukov and Fernandez-Val (2011)) used "federal natality data and found that various observables have significantly stronger associations with birthweight at lower quantiles of the birthweight distribution."

of a family of multi-valued DRF which are implicitly defined by sets of possibly over-identified non-smooth moment conditions under the ignorability condition.

However, literature on study of continuous TE is relatively sparse. Among others, Hirano and Imbens (2004) and Imai and van Dyk (2004) develop the generalized propensity score for continuous treatment models, and Flores (2007) develops nonparametric estimators for the ADRF, its maximizer, and its global maximum under the ignorability assumption. Also, Florens et al. (2008) consider the identification of average TE using control functions. More recently, Lee (2012) studies unconditional distribution of potential outcomes with continuous treatments as a partial mean process with generated regressors. Despite this sparsity, many questions of interest in applied research involve continuous treatments. For example, in the study of TE of mothers' weight gain during pregnancy as well as mother's age on birthweight, the weight gain in pounds and age are continuous variables.

This chapter contributes to the existing TE literature by studying continuous treatments and considering general forms of dose response for TE models, which include both ATE and QTE as special cases. Thus, this chapter extends the literature on ATE and QTE for discrete and multi-valued doses of treatment (see e.g. Heckman and Vytlacil (2005), Firpo (2007), Cattaneo (2010)) to continuous doses of treatment. We point out that the extension from the finite to continuous treatment levels is non-trivial. In fact, since the parameters of interest are now infinite dimensional, the results need to be uniform on the set of treatment levels. In addition, we extend the literature on continuous treatments, which, to our knowledge, only allows for ATE (see, e.g., Flores (2007)), to general (possibly non-smooth) DRF, with QTE being an important example. This is a important innovation because the extension to the non-smooth cases are important in practice and technically challenging.

The remaining of the chapter is organized as follows. Section 1.2 provides identification conditions of the continuous treatment model and proposes a two-step estimator. Section 1.3 studies the asymptotic properties of the two-step estimator. Section 1.4 provides Monte Carlo simulation results and Section 1.5 illustrates the two-step estimator with an application to the estimation of dose-birthweight functions. Section 1.6 concludes the chapter. The proofs of the main results are collected in the Appendix A3.

Notations: Let E and $\mathbb{E}$ denote the expectation and sample average, respectively. Let $\rightsquigarrow$, $\xrightarrow{p}$, and $\xrightarrow{p^*}$ denote weak convergence, convergence in probability, and convergence in outer probability, respectively.

1.2 The Model, Identification, and Estimation

In this chapter, we assume that a random sample of size n is available. The objective is to learn how an outcome variable of an agent changes as the dose of some treatment variable varies. The dose is denoted by t, where $t \in \mathcal{T}$, an interval in $\mathbb{R}$, and the outcome variable is denoted by $Y(t)$. More specifically, for each $t \in \mathcal{T}$, $Y(t)$ is the outcome when the dose of treatment is t. When t varies in $\mathcal{T}$, a random process $Y(t)$ is defined. The random process $Y(t)$ indexed by $t \in \mathcal{T}$ denotes potential outcomes under treatment levels in $\mathcal{T}$. However, one cannot observe the random process $Y(t)$ for all $t \in \mathcal{T}$. Rather, only a single $Y(t_0)$ can be observed, where t_0 is the realization of a random variable T. Therefore, the observed outcome is the random variable

$$Y = Y(T) = \int_{t \in \mathcal{T}} Y(t)\, d\mathbf{1}\{t \geq T\},$$

where $\mathbf{1}\{\cdot\}$ is the indicator function.

Ideally we would like to estimate the value of the DRF at t_0 using the sample with $T = t_0$. However, in general, due to the self-selection problem, bias can be introduced by direct use of the sample counterparts to calculate treatment effects.

To illustrate this point, we consider the estimation of average treatment effects as an example. For any $t_1 < t < t_2$, since

$$\begin{aligned}\underbrace{\mathrm{E}[Y|T=t_2]-\mathrm{E}[Y|T=t_1]}_{\text{Observed difference in birthweight}} = & \underbrace{\mathrm{E}[Y(t_2)-Y(t_1)|T=t]}_{\text{Average treatment effect on the treated}} \\ & + \underbrace{\mathrm{E}[Y(t_2)|T=t_2]-\mathrm{E}[Y(t_2)|T=t]}_{\text{Selection bias 1}} \\ & + \underbrace{\mathrm{E}[Y(t_1)|T=t]-\mathrm{E}[Y(t_1)|T=t_1]}_{\text{Selection bias 2}},\end{aligned}$$

we have

$$\begin{aligned}\underbrace{\mathrm{E}[Y|T=t_2]-\mathrm{E}[Y|T=t_1]}_{\text{Observed difference in birthweight}} = & \underbrace{\mathrm{E}[Y(t_2)-Y(t_1)]}_{\text{Average treatment effect on the treated}} \\ & + \underbrace{\mathrm{E}_t[\mathrm{E}[Y(t_2)|T=t_2]-\mathrm{E}[Y(t_2)|T=t]]}_{\text{Average of selection bias 1}} \\ & + \underbrace{\mathrm{E}_t[\mathrm{E}[Y(t_1)|T=t]-\mathrm{E}[Y(t_1)|T=t_1]]}_{\text{Average of selection bias 2}}.\end{aligned}$$

This simple example indicates that, due to the existence of averages of the selection biases 1 and 2, it is impossible to directly use the sample counterparts to calculate treatment effects. To solve this problem, it is common in the literature to assume the existence of a set of random variables $\boldsymbol{X}$ conditional on which $Y(t)$ is independent from T for all $t \in \mathcal{T}$. In such a case,

$$\begin{aligned}\mathrm{E}[Y|\boldsymbol{X},T=t_2]-\mathrm{E}[Y|\boldsymbol{X},T=t_1] &= \mathrm{E}[Y(t_2)|\boldsymbol{X},T=t_2]-\mathrm{E}[Y(t_1)|\boldsymbol{X},T=t_1] \\ &= \mathrm{E}[Y(t_2)|\boldsymbol{X}]-\mathrm{E}[Y(t_1)|\boldsymbol{X}] \\ &= \mathrm{E}[Y(t_2)-Y(t_1)|\boldsymbol{X}],\end{aligned}$$

which has a causal interpretation. This is the ignorability condition and it is discussed in more detail below. Finally, we need to combine the results for each $\boldsymbol{X}$

and obtain an unconditional treatment effect. In this case, using the law of iterated expectation, this unconditional expectation can be recovered.

The objective of this chapter is to study ADRF and QDRF. From the corresponding DRF it is straightforward to recover the average treatment effect (ATE) and quantile treatment effect (QTE), respectively. To accomplish this aim we develop a general framework for generic moment restriction estimators (Z-estimators) with possibly non-smooth functions. For each $t \in \mathcal{T}$, the parameter of interest $\beta(t) \in B \subset \mathbb{R}$ is assumed to uniquely solve the identifying conditions as

$$\mathrm{E}[m(Y(t); \beta(t))] = 0,$$

where $m(\cdot)$ is a generalized residual function, which we discuss in more details in condition **I.I** stated below. Then the DRF is defined as the parameters of interest, $\beta(t)$, that solve the moment condition. As we will see below, ADRF and QDRF result from choosing specific forms of $m(\cdot)$.

Now we state assumptions on the general model to achieve identification of the parameters of interest.

I.I For each $t \in \mathcal{T}$, $\beta_0(t)$ uniquely solves $\mathrm{E}[m(Y(t); \beta(t))] = 0$, where $m : \mathbb{R} \times B \mapsto \mathbb{R}$ is measurable.

I.II For all $t \in \mathcal{T}$, we have

1 $Y(t) \perp T | \boldsymbol{X}$;

2 $f_{0T|\boldsymbol{X},Y}(t|\boldsymbol{x}, y) > 0$ for $t \in \mathcal{T}$, $\boldsymbol{x} \in \mathcal{X}$ and $y \in \mathcal{Y}$.

I.III Assume that

1 There exists a function $e(y)$ with $\int e(y)\,dy < \infty$ such that

$$|m(y; \beta(t_0)) f_{T,Y|\boldsymbol{X}}(t_0 + \Delta t, y|\boldsymbol{x})| \leq e(y);$$

2 $\mathrm{E}[m(Y; \beta(t_0))|\boldsymbol{X}, T = t_0] = \lim_{\Delta t \downarrow 0} \mathrm{E}[m(Y; \beta(t_0))|\boldsymbol{X}, T \in [t_0, t_0 + \Delta t]]$. Also the interval $\mathcal{T}$ is right open.

Condition **I.I** is an identification condition were $Y(t)$ observable. The parameter of interest, $\beta(t)$, is defined by this moment condition. However, this condition cannot be used directly to estimate $\beta(t)$ because our data are not experimental and $Y(t)$ are not observable for all $t \in \mathcal{T}$. Therefore, condition **I.II.1**, the assumption of ignorability, is fundamental. According to condition **I.II.1**, although the assignment of the treatment level is not random, it is random within subpopulations characterized by $\boldsymbol{X}$. This assumption has been used, among others, by Dehejia and Wahba (1999) and Heckman et al. (1998). Condition **I.II.2** states that for all $\boldsymbol{x} \in \mathcal{X}$ and $y \in \mathcal{Y}$, the density of treatment levels is positive. In our model the triple $(\boldsymbol{X}, Y, T)$ is observable, and a random sample of size n can be obtained. Condition **I.III** allows us to change orders of a limit and an integral. Also, that the set $\mathcal{T}$ is right open is without loss of generality. If we would like to have $\mathcal{T}$ to be right closed, we shrink the interval $[t_0 - \Delta t, t_0]$ to obtain t_0.

The following two examples show that ADRF and QDRF are special cases of $\beta(t)$ in condition **I.I**.

Example (Average)**.** *We first discuss the identification of the ADRF. Setting $m(Y(t); \mu(t)) = Y(t) - \mu(t)$ and letting the first moment to equal to zero, we can obtain $\mu_0(t) = \mathrm{E}[Y(t)]$, the unconditional ADRF. From this it is easy to recover the ATE, which is given by $ATE(t, t') = \mu_0(t) - \mu_0(t')$.*

Example (Quantile)**.** *QDRF is another special case of our general model. Let $m(Y(t); q_\tau(t)) = \tau - \mathbf{1}\{Y(t) < q_\tau(t)\}$, we obtain $q_{\tau 0}(t) \in \inf\{q : F_{Y(t)}(q) \geq \tau\}$, the* *unconditional τth QDRF, where $F_{Y(t)}$ is the distribution function of $Y(t)$. From the QDRF, one can estimate the QTE as $QTE(t,t') = q_{\tau 0}(t) - q_{\tau 0}(t')$. Note that, as is in Firpo (2007) and Cattaneo (2010), in this chapter the QTE is defined as the difference of the τth quantile at different levels of treatment. This definition does not require the assumption of rank preservation, and it is regarded as "a convenient way to summarize interesting aspects of marginal distributions of potential outcomes. However, if rank preservation holds, then the simple differences in quantiles turn out to be the QTE." (Firpo (2007))*

The identification result is presented in the following theorem. For notational convenience, denote $\boldsymbol{u} := (\boldsymbol{x}^\top, y)^\top$ and $\boldsymbol{U} := (\boldsymbol{X}^\top, Y)^\top$.

Theorem 1. *Under conditions* **I.I**–**I.III**, *we have*

$$\mathrm{E}[m(Y(t);\beta(t))] = \mathrm{E}\left[m(Y;\beta(t))\pi_0(\boldsymbol{U};t)\right] \tag{1.1}$$

for each $t \in \mathcal{T}$, where $\pi_0(\boldsymbol{u};t) := \frac{f_{T|\boldsymbol{X},Y}(t|\boldsymbol{x},y)}{f_{T|\boldsymbol{X}}(t|\boldsymbol{x})}$. Consequently,

$$\mathrm{E}\left[m(Y;\beta(t))\pi_0(\boldsymbol{u};t)\right] = 0 \tag{1.2}$$

if and only if $\beta(t) = \beta_0(t)$.

Proof. See Appendix A3. □

The result in equation (1.1) allows identification of the DRF. The left hand side of equation (1.1) is used to define $\beta(t)$, which involves the unobservable $Y(t)$. Consequently, it cannot be used to estimate $\beta(t)$. Nevertheless, the right hand side of equation (1.1) is expressed in terms of observable $(\boldsymbol{X}, Y, T)$, and therefore, can be used to estimate $\beta(t)$. Note that $Y(t)$ is not observable while Y is. The intuition behind the result is that the existence of $\boldsymbol{X}$ delivers identification the parameter of interest. That is, conditional on observed covariates $\boldsymbol{X}$, each individual is randomly assigned to a treatment level.

Remark 1. *The result in Theorem 1 indeed has a similar format as the equation in (2) of Cattaneo (2010) after we transform the latter. We begin with*

$$\mathrm{E}\left[\frac{\mathbf{1}\{T=t\}m(Y;\beta(t))}{p_t(\boldsymbol{X})}\right]=0.$$

By the law of iterated expectation, the left hand side of the previous equation equals

$$\mathrm{E}\left[\frac{m(Y;\beta(t))}{p_t\left(\boldsymbol{X}\right)}\mathrm{E}\left[\mathbf{1}\left\{T=t\right\}|\boldsymbol{X},Y\right]\right]. \tag{1.3}$$

Noting that

$$\mathrm{E}\left[\mathbf{1}\left\{T=t\right\}|\boldsymbol{X},Y\right]=\mathrm{P}\left(T=t|\boldsymbol{X},Y\right)$$

and, by definition, $p_t\left(\boldsymbol{X}\right)=\mathrm{P}\left(T=t|\boldsymbol{X}\right)$*, equation (1.3) equals*

$$\mathrm{E}\left[m(Y;\beta)\frac{\mathrm{P}\left(T=t|\boldsymbol{X},Y\right)}{\mathrm{P}\left(T=t|\boldsymbol{X}\right)}\right].$$

Thus, our result simply "replaces" the conditional probabilities by conditional densities.

Given the identification condition in equation (1.2) of **Theorem 1**, we are able to estimate the parameters of interest. We propose a Z-estimator that involves two steps estimation as follows.[2]

Step 1 Estimate $\pi_0(\boldsymbol{U};t)$ parametrically or nonparametrically and obtain an estimator $\hat{\pi}$.

[2] We work directly with the estimating equations. However, estimation could be carried with GMM methods.

Step 2 Find a zero $\hat{\beta}(t)$ for each $t \in \mathcal{T}$

$$\frac{1}{n}\sum_{i=1}^{n} m(Y_i; \beta(t))\hat{\pi}(\boldsymbol{U}_i; t) = 0. \tag{1.4}$$

The estimator $\hat{\beta}(t)$ is defined as the zero of the equation above[3].

The identification conditions and the estimator are illustrated below using the previous two examples.

Example (Average, continued)**.** *The identification condition for* $\mu_0(t)$ *is*

$$\mathrm{E}[(Y - \mu_0(t))\pi_0(\boldsymbol{U}; t)] = 0.$$

An estimator of $\mu_0(t)$ *is*

$$\hat{\mu}(t) = \left(\frac{1}{n}\sum_{i=1}^{n} \hat{\pi}(\boldsymbol{U}_i; t)\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \hat{\pi}(\boldsymbol{U}_i; t) Y_i. \tag{1.5}$$

Example (Quantile, continued)**.** *The identification condition for* $q_{\tau 0}(t)$ *is*

$$\mathrm{E}[(\tau - \mathbf{1}\{Y < q_{\tau 0}(t)\})\pi_0(\boldsymbol{U}; t)] = 0.$$

An estimator of $q_{\tau 0}(t)$ *is*

$$\hat{q}_\tau(t) = \arg\min_q \frac{1}{n}\sum_{i=1}^{n} \hat{\pi}_0(\boldsymbol{U}_i; t)\rho_\tau(Y_i - q), \tag{1.6}$$

where $\rho_\tau(u) := u(\tau - \mathbf{1}\{u < 0\})$ *is the check function as in Koenker and Bassett (1978).*

[3] The zero does not have to be exact. We only need a solution that approximately solves the equation, which is common in literature; see, e.g., He and Shao (1996) and He and Shao (2000).

1.3 Asymptotic Properties

In this section, we first derive the uniform consistency and the weak limit, and show the uniform semiparametric efficiency of the general two-step estimator described above. Second, we apply the general results and derive the corresponding asymptotic properties of ADRF and QDRF. In addition, we discuss estimation of the nuisance parameter, π_0, and inference based on the two-step estimator.

To establish these results, we first extend existing theoretical results on moment condition restriction estimators (Z-estimators). **Lemmas 1** and **2** in Appendix A1 show consistency and weak convergence of the two-step estimator of the infinite dimensional parameters; the function in the moment condition defining the parameter does not need to be continuous. These results are an extension of Chen et al. (2003) in that the parameter of interest is not in a Euclidean space but in a generic Banach space. Moreover, the results extend Theorem 3.3.1 of van der Vaart and Wellner (1996) in that a possibly infinite dimensional nuisance parameter needs to be estimated in the first step.

We then proceed by applying **Lemmas 1** and **2** to show uniform consistency and weak convergence of the estimator of DRF, $\hat{\beta}(t)$, in $\ell^\infty(\mathcal{T})$. Applications of these results to the specific cases of ADRF and QDRF are also provided. The uniform semiparametric efficiency is based on the pointwise semiparametric efficiency and the weak convergence of the estimator to a tight random process. We then provide ways of estimating the nuisance parameter π_0. As for the inference, we focus on hypothesis testing based on a Kolmogorov and a Cramér-von Mises statistic.

1.3.1 Consistency

Consistency is a desired property for most estimators. In this chapter, different from the discrete or multi-valued treatment models, the treatment levels take values on an interval $\mathcal{T}$. Therefore, the consistency results are established uniformly over the set $\mathcal{T}$. For the general two-step estimator given in equation (1.4) to be uniformly consistent, we state the following sufficient conditions.

C.I $|\mathbb{E}[m(Y;\hat{\beta}(t))\hat{\pi}(\boldsymbol{U};t)]|_\infty = o_p(1)$.

C.II $|\mathrm{E}[m(Y;\beta_n(t))\pi_0(\boldsymbol{U};t)]| \to 0$ implies $|\beta_n(t) - \beta_0(t)|_\infty \to 0$ for any sequence $\beta_n(t)$.

C.III $\sup_{\beta\in\mathcal{B}} |\mathrm{E}[m(Y;\beta(t))]|_\infty < M < \infty$ for some $M > 0$.

C.IV $|\hat{\pi} - \pi_0|_\infty = o_p(1)$.

C.V $\sup_{\beta(t)\in\ell^\infty(\mathcal{T}),\pi\in\Pi_{\delta_n}} |\mathbb{E}[m(Y;\beta(t))\pi(\boldsymbol{U};t)] - \mathrm{E}[m(Y;\beta(t))\pi(\boldsymbol{U};t)]|_\infty = o_{p^*}(1)$ for $\delta_n \downarrow 0$, or

C.V' $\{\psi_{1,\beta,t} : \beta \in \ell^\infty(\mathcal{T}), t \in \mathcal{T}\}$ and $\{\psi_{2,\pi,t} : \pi \in \Pi_{\delta_n}, t \in \mathcal{T}\}$ are Glivenko-Cantelli with respective envelopes F_1 and F_2 such that $\mathrm{E}F_1F_2 < \infty$, where $\psi_{1\beta,t} = m(Y;\beta(t))$ and $\psi_{2,\pi,t} = \pi(\boldsymbol{U};t)$.

Conditions **C.I** defines the Z-estimator and **C.II** is an identification condition for the Z-estimator. Pakes and Pollard (1989) and Chen et al. (2003) have similar assumptions. For a detailed discussion of this type of identification condition, see p. 45 of van der Vaart (1998). Condition **C.III** only requires the moment of the estimating equation to be finite. This is a standard assumption and analogous to condition 4 (b) in Cattaneo (2010). Condition **C.IV** imposes consistency estimation of the nuisance parameter. This is also a usual requirement, which is appears in equation (2.1) of Theorem 2 and equation (3.1) of Theorem 3 of Cattaneo (2010). We will discuss estimations of the nuisance parameter in Section 3.4 below. Condition **C.V** is a uniform law of large numbers, which is implied by condition **C.V'**. These conditions correspond to assumption 4 (a) of Cattaneo (2010), and are high level conditions, and we will provide more primitive conditions for specific cases, i.e., average and quantile. Now we state the consistency result for the estimator of the DRF.

Theorem 2. *Suppose that* $\mathrm{E}[m(Y, \beta_0(t))\pi_0(\boldsymbol{U}; t)] = 0$*, and that conditions* **C.I**–**C.V** *are satisfied. Then*

$$\sup_{t \in \mathcal{T}} |\hat{\beta}(t) - \beta_0(t)| = o_{p^*}(1).$$

Proof. See Appendix A3. □

Now we discuss the consistency of the two-step estimators of ADRF and QDRF given in equations (1.5) and (1.6), respectively. To establish the result for the ADRF, the following conditions are imposed.

AC.I There exists $0 < M_Y < \infty$ such that $\mathrm{E}[Y(t)] < M_Y$. Also, the parameter space for μ is a bounded sub-Banach space $\mathcal{M}$ of $\ell^\infty(\mathcal{T})$.

AC.II The function class $\{\psi_{2,\pi,t} : \pi \in \Pi_\delta, t \in \mathcal{T}\}$ is Glivenko-Cantelli, and has an envelope $F_2(y)$ such that $yF_2(y)$ that is integrable.

Condition **AC.I** requires that the expectation of $Y(t)$ be finite. Also the diameter of the parameter space is finite, which is a common condition for M-estimators, see e.g., Theorem 3 of Chen et al. (2003) or Wooldridge (2010). Hirano et al. (2003) assume the second moment of $Y(1)$ and $Y(0)$ to be finite, which is slightly stronger than our condition. This mainly is because they do not explicitly describe conditions for the consistency of their estimator. Cattaneo (2010) has the same second moment restriction as well. Condition **AC.II** is a high level condition on the nuisance parameters, and will be discussed in more detail below. Nevertheless, there are many functional classes satisfiy this condition. Examples include the smooth function class in Example 19.9 of van der Vaart (1998) for sufficiently smooth functions and sufficiently small tail probabilities. Uniform consistency for the two-step estimator of the ADRF is summarized in the following corollary.

Corollary 1 (Average)**.** *The two-step estimator of ADRF is consistent, i.e.,* $|\hat{\mu}(t) - \mu_0(t)|_\infty = o_{p^*}(1)$*, provided conditions* **AC.I**–**AC.II** *and* **C.IV** *are satisfied.*

Proof. See Appendix A3. □

For the uniform consistency of the QDRF estimator over $t \in \mathcal{T}$, the following conditions are imposed.

QC.I Uniformly in t, the densities $f_{Y(t)}(y)$ is bounded above and $f_{Y(t)}(q_{\tau 0}(t)) > 0$. Also, for any $\delta > 0$, $\inf_{|q - q_{\tau 0}|_\infty > \delta} |\mathrm{E}[(\tau - 1\{Y < q\})\pi_0(\boldsymbol{U}; t)]_\infty > \epsilon_\delta$ fir some $\epsilon_\delta > 0$.

QC.II There exists $0 < M_\pi < \infty$ such that $\pi_0(\boldsymbol{u}; t) < M_\pi$.

QC.III The function class $\{\psi_{2,\pi,t} : \pi \in \Pi_\delta, t \in \mathcal{T}\}$ is Glivenko-Cantelli, and have an envelope $F_2(y)$ such that $F_2(y)$ that is integrable.

Condition **QC.I** is an identification condition on the parameter of interest. This condition is the analogue to the general condition **C.I**, and it is similar to A.2 and A.3 of Angrist et al. (2006), and corresponds to Assumption 2 of Firpo (2007). Cattaneo (2010) uses a similar assumption for the quantile estimation. Condition **QC.II** is a boundedness condition of the joint density of $(\boldsymbol{U}, T)$, and is the continuous treatment version of Assumption 1 (ii) of Firpo (2007). Condition **QC.III** is weaker than condition **AC.II** since only $F_2(y)$ is required to be integrable. This is because $\tau - 1\{\cdot\}$ is already uniformly bounded. The consistency result for the estimator of QDRF is provided in the following corollary.

Corollary 2 (Quantile)**.** *For a given quantile of interest, the two-step estimator of the QDRF is consistent, i.e.,* $|\hat{q}_\tau(t) - q_{\tau 0}(t)|_\infty = o_{p^*}(1)$*, provided conditions* **QC.I**–**QC.III** *and* **C.IV** *are satisfied.*

Proof. See Appendix A3. □

1.3.2 Weak Convergence

In this section, we apply the general results of **Lemma 2** to derive the limiting distribution of the general two-step estimator in (1.4). Later, we demonstrate the results for ADRF and QDRF estimators. To this end, we impose the following sufficient conditions.

G.I $|\mathbb{E}[m(Y;\hat{\beta}(t))\hat{\pi}(\boldsymbol{U};t)]|_\infty = o_p(1/\sqrt{n})$.

G.II The map $\beta \mapsto \mathrm{E}[m(Y;\beta)\pi_0(\boldsymbol{U};\cdot)]$ is Fréchet differentiable at β_0 with a continuously invertible derivative $\mathrm{Z}_1(\beta_0,\pi_0)$.

G.III $\mathrm{E}[m(Y;\beta(t))]$ is Lipschitz continuous at $\beta_0(t)$. In addition, $\sup_{\beta\in\mathcal{B}}|\mathrm{E}[m(Y;\beta(t))^2]|_\infty < M < \infty$ for some $M > 0$.

G.IV $|\hat{\pi} - \pi_0|_\infty = o_p(n^{-1/4})$.

G.V The functional classes $\{\psi_{1,\beta,t} : \beta \in \ell_\delta^\infty(\mathcal{T}), t \in \mathcal{T}\}$ and $\{\psi_{2,\pi,t} : \pi \in \Pi_\delta, t \in \mathcal{T}\}$ are uniformly bounded Donsker classes.

G.V' The functional classes $\{\psi_{1,\beta,t}\psi_{2,\pi,t} : \beta \in \ell_\delta^\infty(\mathcal{T}), \pi \in \Pi_\delta, t \in \mathcal{T}\}$ is a Donsker classes.

G.VI $\sqrt{n}\big(\mathrm{E}[m(Y;\beta_0(t))(\pi(\boldsymbol{U};t) - \pi_0(\boldsymbol{U};t))]|_{\pi=\hat{\pi}} + \mathbb{E}[m(Y;\beta_0(t))\pi_0(\boldsymbol{U};t)]\big)$ converges weakly to a tight random element $\mathbb{G}(t)$ in $\ell^\infty(\mathcal{T})$.

Condition **G.I** defines the Z-estimator, which is slightly stronger than condition **C.I** but still allows the right hand size to be zero only approximately. This type of $o_p(n^{-1/2})$ condition is also assumed in (i) of Theorem 3.3 of Pakes and Pollard (1989) and (2.1) of Theorem 2 of Chen et al. (2003). Condition **G.II** requires the model to be differentiable in β and the derivative is invertible, and corresponds to

(ii) of Theorem 3.3 of Pakes and Pollard (1989) and (2.2) of Theorem 2 of Chen et al. (2003). Condition **G.III** corresponds to assumption 6 (b) of Cattaneo (2010). Condition **G.IV** strengthens condition **C.IV** such that the estimator of the nuisance parameter converges at a rate faster than $n^{-1/4}$. A similar condition appears in equation (4.1) of Theorem 4 and equation (5.1) of Theorem 5 of Cattaneo (2010). Conditions **G.V** and **G.V'** are similar conditions to Assumption 6 (a) of Cattaneo (2010). Those conditions are high level ones and will be discussed below. Now we present the weak convergence result.

Theorem 3. *Suppose that* $|\mathrm{E}[m(Y;\beta_0(t))\pi_0(\boldsymbol{U};t)]|_\infty = 0$*, that* $|\hat{\beta} - \beta_0|_\infty = o_{p^*}(1)$*, and that conditions* **G.I**–**G.VI** *are satisfied. Then*

$$\sqrt{n}(\hat{\beta}(t) - \beta_0(t)) \rightsquigarrow \mathrm{Z}_1^{-1}(\beta_0(t), \pi_0(\boldsymbol{U};t))\mathbb{G}(t)$$

in $\ell^\infty(\mathcal{T})$*.*

Proof. See Appendix A3. □

The result given in Theorem 3 shows that the limiting distribution of the two-step DRF estimator is nonstandard. This result is due to the presence of the set of continuous treatments. However, if one fixes the treatment at $\bar{t}$, then the limiting distribution collapses to a simple normal distribution. In spite of this, below we provide inference methods for DRF over the set of treatments that are simple to implement in practice. In addition, this result has important applications to the two leading examples of ADRF and QDRF. For the weak convergence of the two-step estimator of the ADRF, we impose the following conditions.

AG.I The parameter space for μ_0 is a bounded sub-Banach space $\mathcal{M}$ of $\ell^\infty(\mathcal{T})$. In addition, $\mathrm{E}[Y(t)^2] < M_Y$ for some $0 < M_Y < \infty$.

AG.II The function class $\{\psi_{3,\pi,t} : \pi \in \Pi_\delta, t \in \mathcal{T}\}$ is Donsker, where $\psi_{3,\pi,t}(\boldsymbol{u}) = y\pi(\boldsymbol{u}; t)$.

AG.III $\sqrt{n}\mathrm{E}[(Y - \mu_0(t))(\pi(\boldsymbol{U}; t) - \pi_0(\boldsymbol{U}; t))]|_{\pi=\hat{\pi}}$ converges weakly.

Condition **AG.I** is standard and requires the parameter space to be bounded. Also the second moment of $Y(t)$ is bounded, which is used in Hirano et al. (2003) and Cattaneo (2010). Many function classes satisfy condition **AG.II**, i.e., the smooth function class discussed above. Condition **AG.III** is a high level condition on the nuisance parameter, and we provide an estimator that satisfies this condition.

Corollary 3 (Average, continued)**.** *The two-step estimator of the ADRF is $\sqrt{n}$-consistent and converges weakly in $\ell^\infty(\mathcal{T})$, provided conditions* **AG.I**–**AG.III** *and* **G.IV**.

Proof. See Appendix A3. □

To obtain the weak convergence of the QDRF estimator, equation (1.6), we impose the following conditions.

QG.I The function class $\{\psi_{2,\pi,t} : \pi \in \Pi_\delta, t \in \mathcal{T}\}$ is Donsker with a uniform bound.

QG.II $\sqrt{n}\mathrm{E}[(\tau - \mathbf{1}\{Y < q_{\tau 0}(t)\})(\pi(\boldsymbol{U}; t) - \pi_0(\boldsymbol{U}; t))]|_{\pi=\hat{\pi}}$ converges weakly.

Examples satisfying condition **QG.I** include smooth function classes. Condition **QG.II** is a high level condition and will be discussed in the section of the estimation of the nuisance parameter. This assumptions is similar to **AG.III** and is a version of condition **G.VI**. Now we state the weak convergence result for the estimator of QDRF.

Corollary 4 (Quantile, continued)**.** *The two-step estimator of QDRF is $\sqrt{n}$-consistent and converges weakly in $\ell^\infty(\mathcal{T})$, provided conditions* **QC.I**–**QC.II**, **QG.I**–**QG.II**, *and* **G.IV**.

Proof. See Appendix A3. □

1.3.3 Semiparametric Efficiency of the Two-Step Estimator

In this subsection we first calculate the efficient influence function of the parameter $\beta(t)$ in the following semiparametric model

$$\mathcal{F} = \{F_{\beta,\pi} : \beta \in \ell^\infty(\mathcal{T}), \pi \in \Pi\},$$

where F_{β_0,π_0} is the distribution function of the observed data. Then we provide sufficient conditions under which the proposed two-step estimator is uniformly semiparametric efficient.

Claim 1. *Suppose* $\Gamma_0(t) := \frac{\partial \mathrm{E}[m(Y(t);\beta_0(t))]}{\partial \beta(t)}$ *exists. For each* $t \in \mathcal{T}$*, the efficient influence function of the parameter* $\beta(t)$ *is*

$$\Psi_\beta(y, t, \boldsymbol{x}) = -\Gamma_0^{-1}(t)\psi(y, \boldsymbol{x}, t, \beta_0, \pi_0, e_0),$$

where $\psi(y, \boldsymbol{x}, t, \beta_0, \pi_0, e_0) = m(y; \beta_0(t))\pi_0(\boldsymbol{u}; t) - e_0(\boldsymbol{x}, \beta_0(t))(\pi_0(\boldsymbol{u}; t) - 1)$ *with* $e_0(\boldsymbol{x}, \beta(t)) = \mathrm{E}[m(Y; \beta(t))|\boldsymbol{X} = \boldsymbol{x}]$.

Proof. See Appendix A3. □

Based on the efficient influence function of $\beta(t)$, we show that the two-step estimator is uniformly semiparametric efficient provided the following condition

E. $\sqrt{n}\mathbb{E}[m(Y; \beta_0(t))\hat{\pi}(\boldsymbol{U}; t)] = \sqrt{n}\mathbb{E}[\psi(Y, \boldsymbol{X}, t, \beta_0, \pi_0, e_0)] + o_p(1)$.

Condition **E** is critical to the efficiency of the two-step estimator, and it is similar to its corresponding condition for the multi-valued model is condition (4.2) of Cattaneo (2010).

Theorem 4. *Assume that the conditions of* **Theorem 3** *and condition* **E** *hold. Then the two-step estimator is uniformly semiparametric efficient.*

Proof. See Appendix A3. □

This result guarantees that the two-step estimator is uniformly semiparametric efficient. Hypothesis testings based on this estimator are expected to be optimal.

1.3.4 Estimation of π_0

We have been assuming that the estimator $\hat{\pi}$ of the nuisance parameter π_0 satisfies various conditions. In this subsection we discuss the estimation of the nuisance parameter π_0, i.e., $\frac{f_{T|X,Y}(t|x,y)}{f_{T|X}(t|x)}$. The estimation of the nuisance parameter is a very important step for implementation of the proposed estimator in practice. It is common in literature to estimate nuisance parameters in two-step estimators by using a parametric model, see, e.g., Newey (1984), Murphy and Topel (1985), Newey and McFadden (1994), Chernozhukov and Hong (2002) and Wei and Carroll (2009). For estimators of π_0 to have the desirable properties, we impose the following assumptions.

N.I Assume $\pi = \pi(\boldsymbol{u}; t; \vartheta)$, where $\vartheta \in \mathbb{R}^{d_\vartheta}$ with d_ϑ being a positive integer. $\pi(\boldsymbol{u}; t; \vartheta)$ is a smooth function of ϑ with uniformly continuous, bounded, and square integrable first derivative, $\pi'(\boldsymbol{u}; t; \vartheta)$, with respect to ϑ.

N.II There exists $\hat{\vartheta}$ such that $\sqrt{n}(\hat{\vartheta} - \vartheta) \xrightarrow{d} N(0, \Im_\vartheta^{-1})$ with $\Im_\vartheta$ being nonsingular.

Condition **N.I** is a smoothness and boundedness condition for the function to be estimated, and condition **N.II** assumes that there is an estimator of the parameter that is asymptotically normal. Examples which satisfy condition **N.II** include maximum likelihood estimator.

Claim 2. *Under conditions* **N.I** *and* **N.II**, *conditions* **C.IV**–**C.V** *and* **G.IV**–**G.VI** *hold.*

Proof. See Appendix A3. □

Claim 2 can be applied to verify the conditions on the nuisance parameters for the average and quantile examples. Now we provide a few examples to demonstrate the estimation of π_0 in practice.

Example. *We estimate $f_{Y|\boldsymbol{X},T}(y|\boldsymbol{x},t)$ and $f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$ separately. For $f_{Y|\boldsymbol{X},T}(y|\boldsymbol{x},t)$, assume the relationship*

$$Y = g(\boldsymbol{X}, T; \boldsymbol{b}) + \epsilon,$$

and $\epsilon|\boldsymbol{X}, T \sim N(0, \sigma_\epsilon^2)$, where $g(\cdot)$ is some known function and $\boldsymbol{b}$ is an unknown parameter to be estimated. Using Nonlinear Least Square Estimation, we obtain estimators of the conditional mean and variance, and therefore, the conditional density of Y given $\boldsymbol{X}$ and T. Similarly, we estimate $f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$.

Example. *To estimate the conditional density $f_{Y|\boldsymbol{X},T}(y|\boldsymbol{x},t)$, we can also assume the following model*

$$\Lambda(Y, \lambda) = \Lambda(g(\boldsymbol{X}, T), \lambda) + \epsilon,$$

where $\epsilon|\boldsymbol{X}, T \sim N(0, \sigma_\epsilon^2)$, $g(\cdot)$ is some known function, and $\Lambda(\cdot)$ is the Box-Cox transformation function, which is defined as $\Lambda(Z, \lambda) = \log Z$ if $\lambda = 0$ and $= \frac{Z^\lambda - 1}{\lambda}$ otherwise. Using Maximum Likelihood Estimation, we obtain the unknown parameter λ and therefore the conditional density $f_{Y|\boldsymbol{X},T}(y|\boldsymbol{x},t)$. Similarly, we estimate $f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$.

Example. *A simple approach to estimate π_0 is to assume that (t, x, y) follow a known multivariate distribution, as a Normal distribution for instance. Then, Maximum Likelihood Estimation can be applied and the estimator $\hat{f}_{T,X,Y}(t, x, y)$ calculated, and then $\hat{\pi}$ can be obtained.*

1.3.5 Inference on the DRF

Inference is carried uniformly over the set of treatment levels, $\mathcal{T}$. Given the formulation for inference of DRF, inference for the treatment effects is straightforward. In particular, it is possible to derive manifold tests from the weak convergence results in Theorem 3.

Kolmogorov and Cramer-von Mises type tests can be used to test general hypotheses on $\beta(t)$, i.e., $H_{01} : \beta(t) = r(t)$ when $r \in \ell^{\infty}(\mathcal{T})$ is known. Thus, from the result in Theorem 3, under the null hypothesis H_{01},

$$V_n(t) := \sqrt{n}(\hat{\beta}(t) - r(t)) \rightsquigarrow \mathbb{G}(t).$$

More precisely, we propose the following test statistics:

$$\begin{aligned} T_{1n} &:= \sup_{t \in \mathcal{T}} |V_n(t)|, \\ T_{2n} &:= \int_{t \in \mathcal{T}} |V_n(t)| \, dt. \end{aligned}$$

They are a Kolmogorov type and a Cramér-von Mises type statistic, respectively. Now we present the limiting distributions of the test statistics under the null hypothesis.

Corollary 5. *Assume the conditions of Theorem 3 hold. Under* $H_{01} : \beta_0(t) = r(t)$, *as* $n \to \infty$,

$$T_{1n} \rightsquigarrow \sup_{t \in \mathcal{T}} |\mathbb{G}(t)|, \quad T_{2n} \rightsquigarrow \int_{t \in \mathcal{T}} |\mathbb{G}(t)| \, dt.$$

Proof. The assertion holds by Theorem 3 and the continuous mapping theorem. □

In addition to testing the hypothesis $\beta_0(t) = r(t)$ with known $r \in C(\mathcal{T})$, we could also test the hypothesis with unknown r, in which case, the estimation of

r is needed. Often, a $\sqrt{n}$-consistent estimator $\hat{r}$ is available, and under the null hypothesis H_{02} $\beta_0(t) = r(t)$, the test statistic becomes

$$\bar{V}_n(t) := \sqrt{n}(\hat{\beta}(t) - \hat{r}(t)) \rightsquigarrow \mathbb{G} - \mathbb{G}_r,$$

where $\mathbb{G}_r$ is the weak limit of $\sqrt{n}(\hat{r}(t) - r(t))$. Therefore, due to the estimation of $r(t)$, a drift component is introduced.

We propose the following test statistics:

$$\bar{T}_{1n} := \sup_{t\in\mathcal{T}} |\bar{V}_n(t)|,$$
$$\bar{T}_{2n} := \int_{t\in\mathcal{T}} |\bar{V}_n(t)|\, dt.$$

Now we display the limiting distributions of the test statistics under the null hypothesis.

Corollary 6. *Assume the conditions of Theorem 3 hold. Under $H_{02} : \beta_0(t) = r(t)$, as $n \to \infty$,*

$$\bar{T}_{1n} \rightsquigarrow \sup_{t\in\mathcal{T}} |\mathbb{G}(t) - \mathbb{G}_r|, \quad \bar{T}_{2n} \rightsquigarrow \int_{t\in\mathcal{T}} |\mathbb{G}(t) - \mathbb{G}_r|\, dt.$$

Proof. The assertion holds by Theorem 3 and the continuous mapping theorem. □

The weak limits in **Corollaries 5** and **6** are not standard. Therefore, to make practical inference we suggest the use of bootstrap techniques to approximate the limiting distribution. A formal justification for our simulation method, discussed below, is stated in **Lemma 3**, in Appendix A1. It is also an extension of that in Chen et al. (2003). A simple application of Corollaries 5 and 6 produces the bootstrap procedure for the ADRF or QDRF.

Given the above framework, inference for the treatment effects is simple. Using the inference of ATE from treatment level t_1 to t_2 as an example, the point estimate of the ATE is $\hat{\mu}(t_2) - \hat{\mu}(t_1)$, which has an asymptotic normal distribution with mean $\mu(t_2) - \mu(t_1)$ and variance calculable from the weak limit of $\hat{\mu}(t)$. Therefore, the inference can be done using standard method.

Implementation of Testing Procedures

Implementation of the proposed tests in practice is simple. To test H_{01} with known $r(t)$, one needs to compute the statistics T_{1n} or T_{2n}. Analogously, to test H_{02} one computes $\bar{T}_{1n}$ or $\bar{T}_{2n}$. The steps for implementing the tests are as following.

First, the estimates of $\beta(t)$ are computed by solving the problem in equation (1.4). For special cases of DRF, as ADRF and QDRF, one replaces equation (1.4) with (1.5) and (1.6) respectively. Second, T_{1n} or T_{2n} are calculated by centralizing $\hat{\beta}(t)$ at $r(t)$ and taking the maximum over t (for T_{1n}) or summing over t (for T_{2n}). For the general case, H_{02} with unknown $r(t)$, the tests can be implemented in the same fashion. The only adjustment is after estimating $\beta(t)$ one uses $\hat{r}(t)$ to compute $\bar{T}_{1n}$ or $\bar{T}_{2n}$.

After obtaining the statistic of test, it is necessary to compute the critical values. We propose the following scheme. We use the statistic of test $\bar{T}_{1n}$ as an example, but the procedure is the same for the other cases. Take B as a large integer. For each $b = 1, \ldots, B$:

(i) Obtain the resampled data $\{(Y_i^b, \boldsymbol{U}_i^b), i = 1, \ldots, n\}$.

(ii) Estimate the DRF $\hat{\beta}^b(t)$ and set $V_n^b(t) := \sqrt{n}(\hat{\beta}^b(t) - r(t))$

(iii) Compute the test statistic

$$\widehat{T}_{1n}^b = \max_{t \in \mathcal{T}} |V_n^b(t)|$$

Let $\hat{c}_{1-\alpha}^B$ denote the empirical $(1-\alpha)$-quantile of the simulated sample $\{\widehat{T}_{1n}^1, \ldots, \widehat{T}_{1n}^B\}$, where $\alpha \in (0, 1)$ is the nominal size. We reject the null hypothesis if T_{1n} is larger than $\hat{c}^B_{1-\alpha}$. In practice, the maximum in step (iii) is taken over a discretized subset of $\mathcal{T}$.

1.4 Monte Carlo

In this section we conduct Monte Carlo simulations. Our data generating process has treatment level $t \in [0,1]$. A sample of n i.i.d. random elements $(X_i, \epsilon_i(t), v_i(t))$ whose components are mutually independent are generated, where the independent white Gaussian noises $\epsilon_i(t)$ and $v_i(t)$ are represented by $(\epsilon_i(0), \epsilon_i(0.01), \cdots, \epsilon_i(0.99), \epsilon_i(1))$ and $(v_i(0), v_i(0.01), \cdots, v_i(0.99), v_i(1))$, respectively. The observed individual characteristics $X_i \sim Unif[-0.5, 0.5]$ and $Y_i(t) = 0.5 - |0.5 - t| + X_i + v_i(t)$ where independent innovations $v_i(0)$, $v_i(0.1)$, $\cdots$, $v_i(1) \sim N(0,1)$. The treatment assignment is generated by $T_i = \arg\max_{t \in \{0, 0.01, \cdots, 0.99, 1\}} H_{t,i}$, where $H_{t,i} = \sin(2\pi t) X_i + \epsilon_i(t)$ where independent innovations $\epsilon_i(0)$, $\epsilon_i(0.1)$, $\cdots$, $\epsilon_i(1) \sim N(0,1)$. We generate the data in such a way that the mean and median functions are $0.5 - |0.5 - t|$, an up-side-down symmetric check function. The level is highest in the middle range and decreases as t deviates from the middle. The number of replications is 2,000.

Our null and alternative hypotheses are summarized below.

$H_{m0}:$	$\mu_0(t) = 0.5 - |0.5 - t|$ for $t \in [0.2, 0.8]$
$H_{q0}:$	$q_{0.5,0}(t) = 0.5 - |0.5 - t|$ for $t \in [0.2, 0.8]$
$H_{m1}:$	$\mu_0(t) = t$ for $t \in [0.2, 0.8]$
$H_{q1}:$	$q_{0.5,0}(t) = t$ for $t \in [0.2, 0.8]$
$H_{m2}:$	$\mu_0(t) = t^2$ for $t \in [0.2, 0.8]$
$H_{q2}:$	$q_{0.5,0}(t) = t^2$ for $t \in [0.2, 0.8]$
$H_{m3}:$	$\mu_0(t) = 0.25 - (t - 0.5)^2$ for $t \in [0.2, 0.8]$
$H_{q3}:$	$q_{0.5,0}(t) = 0.25 - (t - 0.5)^2$ for $t \in [0.2, 0.8]$

On the one hand, the first alternative is a linear function while the second is an asymmetric quadratic function, which are quite different from the null. On the other hand, the third alternative is a quadratic function symmetric around 0.5 and attains its maximum at 0.5.

We use the Cramér-von Mises test for the simulations. Also, we use the method of Hall et al. (2004), a nonparametric method to estimate the conditional densities. We first show the biasedness of the estimator when sample sizes are 150 and 300. The bias is defined as the supreme of the pointwise biases and presented below.

	n=150	n=300
Mean	0.071	0.059
Median	0.068	0.056

As expected, the bias decreases as sample size increases. Now we present the empirical size and power below.

		Size	Power for H_1	Power for H_2	Power for H_3
n=150	Mean	0.03	1.00	1.00	0.93
B=150	Median	0.02	1.00	1.00	0.55
n=150	Mean	0.03	1.00	1.00	0.93
B=300	Median	0.02	0.99	1.00	0.53
n=300	Mean	0.02	1.00	1.00	1.00
B=150	Median	0.01	1.00	1.00	0.88
n=300	Mean	0.02	1.00	1.00	1.00
B=300	Median	0.01	1.00	1.00	0.87
n=300	Mean	0.02	1.00	1.00	1.00
B=500	Median	0.01	1.00	1.00	0.88

In the simulations, we evaluate the empirical size and power for a variety of sample sizes and numbers of bootstrap. We observe that the sizes are close to the level of significance, 5%, and the power is high for the alternative hypotheses H_1 and H_2. To study the impact of sample size and number of bootstrap on the power of the test, we test for H_3, which is closer to the null hypothesis. It turns out that there is power gain from increasing the sample size. However, the power is not sensitive to the number of bootstrap, implying that smaller number of bootstrap is satisfactory and using larger number of bootstrap is not necessary.

The simulations show that, although we cannot show the weak convergence of the Z-estimator if we use kernel estimation in the first step, in practice, we nevertheless can use the kernel estimation as one of a few alternative methods in the first step.

1.5 Applications to the Study of Dose-Birthweight Functions

In this section, we illustrate the use of the two-step estimator with a study of dose-birthweight functions. Recently birthweight has been shown to be the foremost telltale of infant health. Unhealthy births have large economic costs in both immediate medical costs and longer care costs.

Infants are classified as low birthweight (LBW) when weighing less than 2.5 kilograms at birth. There is empirical evidence showing that the direct medical costs of LBW are rather high. Almond et al. (2005) document that the hospital costs for newborns are elevated: "the expected costs of delivery and initial care of a baby weighing one kilogram at birth can exceed $100,000 (in year 2000 dollars). The costs remain elevated even among babies weighing 2–2.1 kilograms; an additional pound (454 grams) of weight is still associated with a $10,000 difference in hospital charges for inpatient services."[4] Also, at lower birthweights, the infant death rate is higher.

On the other hand, problems associated with high birthweight have become more recognizable. Abrevaya and Dahl (2008) state "babies weighing more than 4 kilograms (defined as high birthweight (HBW)) and especially those weighing

[4]Appurtenant expenditures, such as radiological, pharmaceutical, respiratory, and laboratory fees, greatly extend the costs of intensive care for LBW infants. (See, e.g., Behrman et al. (2007)).

more than 4.5 kilograms (classified as very high birthweight) are more likely to require cesarean-section births, have higher infant mortality rates, and develop health problems later in life." Recent research also suggests giving birth to infants over 4.5 kilograms carries significant risks to both the infant and the mother; see, e.g., Cesur and Kelly (2010) and Webb (2011) for more detailed discussions. Fetal disorders such as shoulder dystocia, stillbirth, Erb's palsy, jaundice, and respiratory distress have been found to be more common in HBW infants in addition to greater levels of obesity later.

Other studies on QTE include Abadie et al. (2002) and Chernozhukov and Hansen (2005), which study treatment effects when they may not be monotonic along the outcome distribution. Chernozhukov and Hansen (2006, 2008) investigate a class of instrumental quantile regression methods for structural and TE models. Imbens and Newey (2009) extend Newey et al. (1999), Pinkse (2000), and Blundell and Powell (2003) to identification and estimation of a family of parameters, termed structural quantile functions, and apply to the continuous treatments case. Moreover, there is emerging literature on average treatment effects for continuous variables, which includes Hirano and Imbens (2004) and Flores (2007).

1.5.1 Data

The United States natality data from the National Vital Statistics System (NVSS) of Centers for Disease Control and Prevention (CDC) document nearly all births in registered areas. The data in this study are the 2004 public use natality data of Wisconsin.

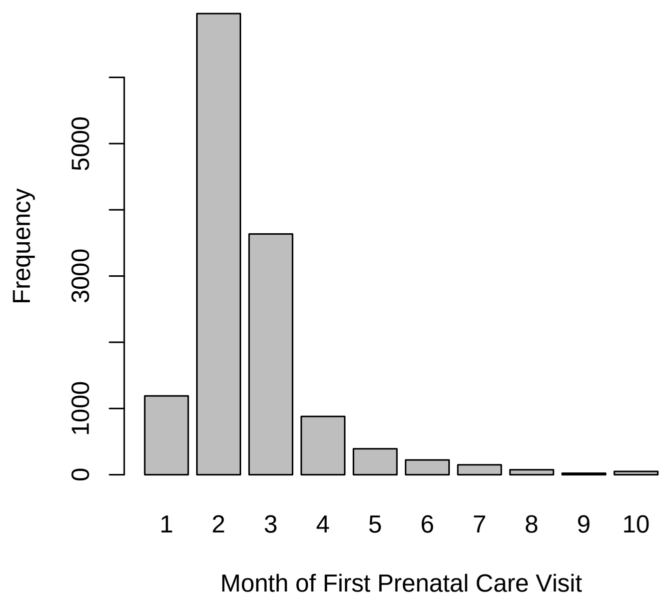
For this study, we consider only live, singleton births (without missing values of any characteristics used in the study) to new, white mothers that are not older than 45, with less than five years of college, whose counties of occurrence (birthing) and residence are the same. Birthweights have been found to differ across different ethnicities, numbers of babies at delivery, and so on. By using a more homogeneous sample, we can focus more on the effects of birth inputs on the birthweights. This results in a sample of 13,581 births. We emphasize that all the inferences done using the sample should be applicable only for the subpopulation represented by the sample choice.

Now we summarize the variables of interest. Out of 13,581 births, there are 6,508 females (proportion 0.4792), and 7,732 mothers are married (proportion 0.5693). Table 1.1 displays statistics for birthweight (measured in kilograms), the mother's age, the mother's weight gain during pregnancy (WG), number of cigarettes per day (Cigarettes), number of prenatal care visits (No. Care), and the mother's years of education for the sample. And Figure 1.1 shows the distribution of the month of first time prenatal care visits.

Our dataset is similar to "1st child" Washington and Arizona datasets of Abrevaya and Dahl (2008) for the variables that are directly comparable. The number of observations are 45,067 and 56,201 for Washington and Arizona, respectively. For example, the averages of the birthweights of Washington and Arizona data are 3.44 and 3.34 kilograms, respectively, while the average of the birthweights in Wisconsin is 3.35 kilograms. The averages of the ages of Washington and Arizona are 25.27 and 25.23, respectively, while that of Wisconsin is 24.88. The averages of number of prenatal care visits and education in Wisconsin are slightly lower than those of Washington and Arizona. The proportion of male infants are similar for all the births in the three states, but the proportion of married mothers in Wisconsin is much lower than those of Washington and Arizona.

[Table 1.1 and Figure 1.1 about here]

Figure 1.1: Distribution of the Months of First Prenatal Care Visit



Note: The number 10 means “did not have prenatal care”.

Table 1.1: Summary Statistics

	Birthweight	Age	WG	Cigarettes	No. Care	Education
Min.	0.26	14.00	0.00	0.00	0.00	0.00
1st Qu.	3.06	20.00	25.00	0.00	10.00	12.00
Median	3.37	24.00	33.00	0.00	12.00	13.00
Mean	3.35	24.88	34.20	1.10	11.80	13.02
3rd Qu.	3.69	29.00	42.00	0.00	14.00	15.00
Max	5.67	45.00	95.00	40.00	49.00	16.00
SD.	0.54	5.45	13.69	3.25	3.40	2.27

1.5.2 Estimation of Nuisance Parameter π_0

The estimation strategy of nuisance parameters $\pi_0(\boldsymbol{u};t) := \frac{f_{0T|\boldsymbol{X},Y}(t|\boldsymbol{x},y)}{f_{0T|\boldsymbol{X}}(t|\boldsymbol{x})}$ in (1.5) and (1.6) is the same for both the treatment effects of the mother's age and weight gain during pregnancy. In this section, we describe the details of the estimation procedure using the first treatment effect model, i.e., the mother's age, as an example.

The mothers' ages in our sample range from 14 to 45 years old. Therefore, it is natural to treat the mother's age as a continuous variable in the interval [13,46]. For the estimation of conditional distribution, we assume the relationship $\log\left(\frac{T_i-13}{46-T_i}\right) = \boldsymbol{X}_i^\top\boldsymbol{\theta}_0 + \epsilon_i$, where ϵ_i is independent of $\boldsymbol{X}_i$ and has density $N(0,\sigma_0^2)$. The choice of $\boldsymbol{X}$ is described in the following subsections. The log-ratio form of the dependent variable makes mothers' age to be limited to [13, 46]. This strategy is similar to the one in the logit model where the probability is confined to [0, 1]. Therefore, $\frac{T_i-13}{46-T_i} =: \eta_i$ follows log-normal distribution log-$N(\boldsymbol{X}_i^\top\boldsymbol{\theta}_0, \sigma_0^2)$. The density of $T|\boldsymbol{X}$ is obtained by calculating the distribution function,

$$F_{0T|\boldsymbol{X}}(t|\boldsymbol{x}) = \mathrm{P}(T \le t|\boldsymbol{X}=\boldsymbol{x}) = \mathrm{P}\left(\eta \le \frac{t-13}{46-t}|\boldsymbol{X}=\boldsymbol{x}\right)$$
$$= F_{\eta|\boldsymbol{X}}\left(\frac{t-13}{46-t}|\boldsymbol{x}\right) = \Phi\left(\frac{\log\frac{t-13}{46-t} - \boldsymbol{x}^\top\boldsymbol{\theta}_0}{\sigma_0}\right)$$
$$f_{0T|\boldsymbol{X}}(t|\boldsymbol{x}) = \phi\left(\frac{\log\frac{t-13}{46-t} - \boldsymbol{x}^\top\boldsymbol{\theta}_0}{\sigma_0}\right)\frac{1}{\sigma_0}\left(\frac{1}{t-13} + \frac{1}{46-t}\right),$$

where Φ and ϕ are distribution and density functions of a standard normal random variable. For the conditional density of $T|\boldsymbol{X}, Y$, we assume the relationship $\log\left(\frac{T_i-13}{46-T_i}\right) = \boldsymbol{U}_i^\top \boldsymbol{\vartheta}_0 + \varepsilon_i$, where ε_i is independent of $\boldsymbol{U}_i$ and has density $N(0, \varsigma_0)$. Therefore, the conditional distribution function

$$f_{0T|\boldsymbol{U}}(t|\boldsymbol{u}) = \phi\left(\frac{\log\frac{t-13}{46-t} - \boldsymbol{u}^\top \boldsymbol{\vartheta}_0}{\varsigma_0}\right)\frac{1}{\varsigma_0}\left(\frac{1}{t-13} + \frac{1}{46-t}\right).$$

Hence, we have

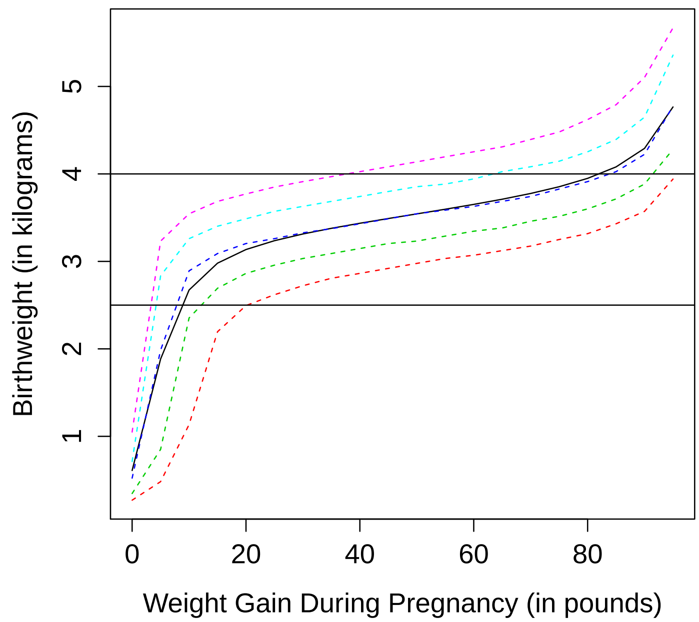
$$f_0(\boldsymbol{u}, t) = \frac{f_{0T|\boldsymbol{U}}(t|\boldsymbol{u})}{f_{0T|\boldsymbol{X}}(t|\boldsymbol{x})} = \frac{\phi\left(\frac{\log\frac{t-13}{46-t} - \boldsymbol{u}^\top \boldsymbol{\vartheta}_0}{\varsigma_0}\right)\sigma_0}{\phi\left(\frac{\log\frac{t-13}{46-t} - \boldsymbol{x}^\top \boldsymbol{\theta}_0}{\sigma_0}\right)\varsigma_0}.$$

1.5.3 Empirical Results

Mothers' Weight Gain during Pregnancy

The results regarding the mothers' weight gain during pregnancy show evidence that, after controlling for a mother's characteristics chosen (i.e., age, marital status, years of education, number of cigarettes per day, and the month of first prenatal care visit), in general, greater weight gain during pregnancy leads to higher birthweight. Figure 1.2 reports the estimates of the average and selected quantiles of the birthweight for different levels of the mother's weight gain during pregnancy. From the figure, we see that the slopes are relatively larger for low or high weight gain. The shape of the curves resembles a simple cubic function with steeper slopes at the extremes. This implies weight gaining generates higher birthweights at low and high levels of weight gain. For low weight gains, the impact on the birthweight is higher for upper quantiles and relatively mild for low quantiles. However, for the middle range of weight gain, all the curves are relatively parallel. The disaggregated plots with 90% confidence bands are shown in Figure 1.3. The confidence bands in general are relatively wider at the extremes of weight gain due to the sparsity of the data at

Figure 1.2: Mothers' Weight Gain during Pregnancy and Level of Birthweight



The top and bottom horizontal lines represent the thresholds of high and low birthweight, respectively. The solid curve is the average of birthweights and the dashed curves are the 90%, 75%, 50%, 25%, and 10% quantiles of birthweights.

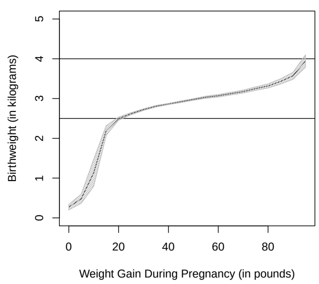
the extremes.

Table 1.2 describes treatment effects for selected weight treatment effects. It is divided according to the weight gain interval effects. The first part contains 20 pound effects. The second contains 40 pound effects, and so on until an 80 pound interval. The results show that the impact of gaining weight is positive.

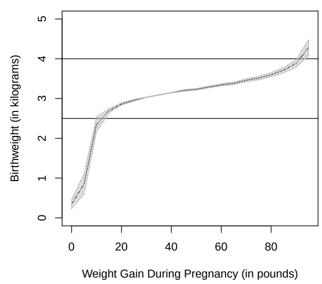
In summary, to produce an infant with healthy birthweight, mothers should gain weight between approximately 20 to 40 pounds. The average birthweight is below 2.5 kilograms for mothers with weight gain less than around 10 pounds and is above 4 kilograms for mothers with weight gain more than around 80 pounds. It seems optimal for pregnant women to gain between 20 to 40 pounds to lower the chances of having LBW or HBW infants.

Figure 1.3: Mothers' Weight Gain during Pregnancy and Level of Birthweight with 90% Confidence Bands

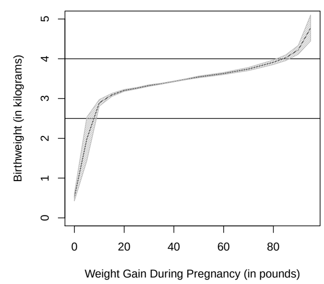
(a) 10% Quantile



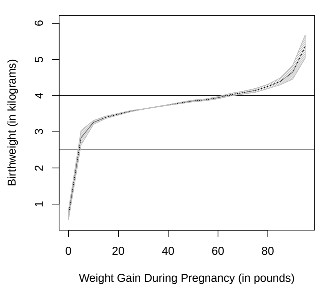
(b) 25% Quantile



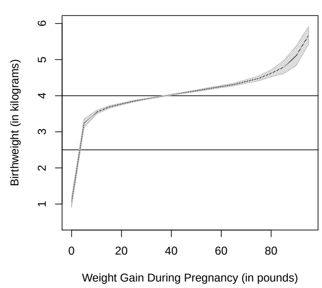
(c) 50% Quantile



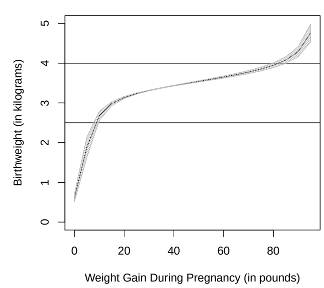
(d) 75% Quantile



(e) 90% Quantile



(f) Average



The top and bottom horizontal lines represent the thresholds of high and low birthweight, respectively.

Table 1.2: Treatment Effects of Mothers' Weight Gain During Pregnancy

WG change	10% Qt.	25% Qt.	50% Qt.	75% Qt.	90% Qt.	Average
0–20	2.23	2.52	2.68	2.77	2.72	2.53
SD.	0.05	0.07	0.05	0.08	0.07	0.05
20–40	0.37	0.28	0.23	0.25	0.25	0.30
SD.	0.04	0.07	0.06	0.08	0.08	0.06
40–60	0.21	0.20	0.20	0.20	0.23	0.22
SD.	0.05	0.07	0.06	0.09	0.09	0.07
60–80	0.25	0.25	0.28	0.31	0.37	0.30
SD.	0.05	0.08	0.07	0.10	0.11	0.08
0–40	2.59	2.80	2.91	3.03	2.98	2.83
SD.	0.03	0.02	0.02	0.02	0.02	0.02
20–60	0.57	0.48	0.43	0.46	0.48	0.52
SD.	0.04	0.03	0.03	0.03	0.03	0.03
40–80	0.45	0.45	0.48	0.51	0.60	0.51
SD.	0.05	0.05	0.05	0.05	0.07	0.05
0–60	2.80	3.00	3.11	3.23	3.20	3.04
SD.	0.02	0.02	0.01	0.02	0.02	0.01
20–80	0.82	0.74	0.71	0.77	0.85	0.81
SD.	0.03	0.03	0.03	0.04	0.06	0.03
0–80	3.05	3.25	3.39	3.54	3.57	3.34
SD.	0.02	0.02	0.02	0.02	0.05	0.02

Mothers' Age

The QDRF of the mother's age on birthweight is downward sloping. For a given age, this negative impact becomes more severe for lower parts of the distribution of birthweights. Although intuitive, this result complements existing results in the literature with three advantages. First, our results can be interpreted as causal effects. Second, we estimate the unconditional quantile and mean of the birthweight for a range of the mothers' age. Third, unlike using regression framework, our results show that the treatment effects are not confined to be constant or a linear function of ages.

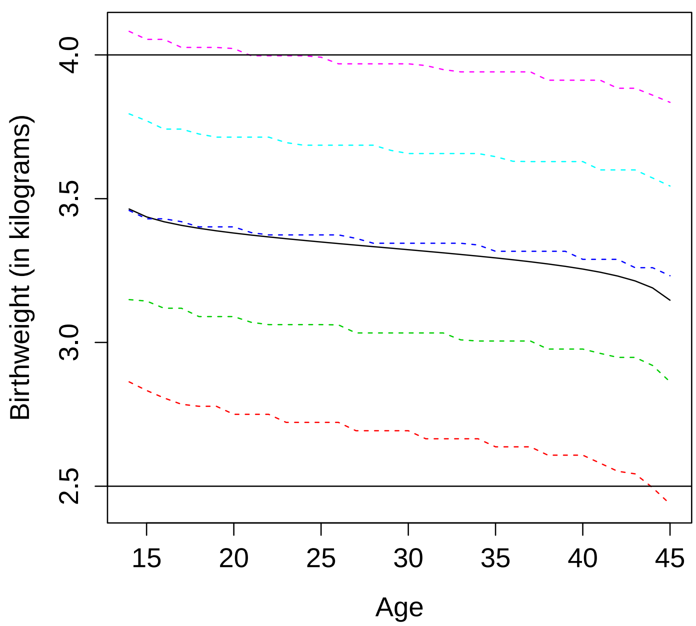
In the current study of the mothers' age, we control for marital status, years of education, and number of cigarettes per day during pregnancy. It is important to note that although we are controlling for some characteristics of mothers, we are estimating the unconditional treatment effects. The empirical results for treatment effects of the mother's age on birthweight reveal that the treatment effect is negative; that is, as expected, the birthweight decreases as the mother's age increases. Figure 1.4 plots the point estimates for the mean, 10%, 90%, and the three quartiles of birthweights for mothers' ages from 14 to 45. This impact of the mother's age on birthweight is negative for all the quantiles. However, for a given age, this impact becomes more severe for lower parts of the distribution of birthweight. In particular, the impact is very prominent for the 10% quantile of mothers after 40 years old. The estimated average birthweight is downward sloping, and more negative at high ages, which is different from the median and is probably capturing the effect of the low quantile. On the other hand, the median birthweight is robust to this feature.

From the disaggregated figures (Figure 1.5) one can see that the 90% confidence bands are narrower in the middle ages because there are more data for that age range. In contrast, we can see that the confidence interval for 10% quantile at the age of 45 is relatively wide.

Table 1.3 describes the treatment effects for selected age treatment effects. The table is divided according to the age interval effects. The first part contains 5 year effects. The second part contains 10 year effects, and so on, until a 30 year interval. Most of them are statistically significant, and negative values show evidence that aging is negatively related to birthweight. Finally, the effect is larger (in absolute values) for the low part of the distribution of birthweights; for example, for a mother aged 25 to 35 years the treatment effect is -0.08 at 10% and -0.05 at 90%.

In general, there are certain risks of having a baby when the mother is too young or too old. Although on average the birthweight is within the "healthy range" between 2.5 and 4 kilograms, our estimates show that mothers younger than 20 years are likely to have HBW infants, while mothers older than 44 years are likely

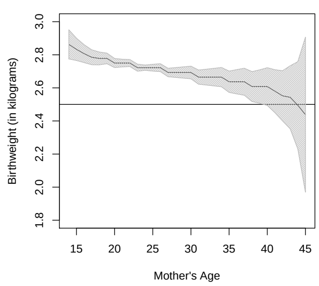
Figure 1.4: Mothers' Age and Level of Birthweight



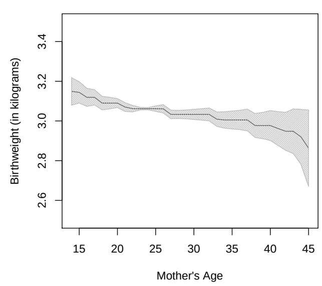
The top and bottom horizontal lines represent the thresholds of high and low birthweight, respectively. The solid curve is the average of birthweights and the dashed curves are the 90%, 75%, 50%, 25%, and 10% quantiles of birthweights.

Figure 1.5: Mothers' Age and Level of Birthweight with 90% Confidence Bands

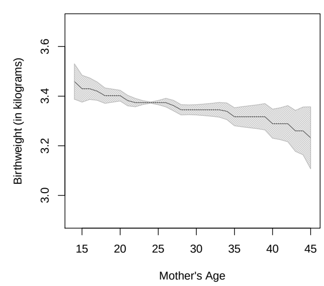
(a) 10% Quantile



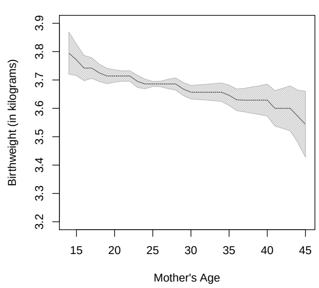
(b) 25% Quantile



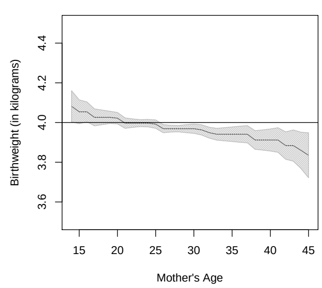
(c) 50% Quantile



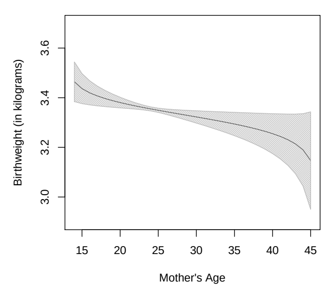
(d) 75% Quantile



(e) 90% Quantile



(f) Average



The thresholds for low and high birthweights are 2.5 kilograms and 4 kilograms, respectively.

to have LBW infants. Therefore, it may be prudent for women who plan to have a baby to do so approximately between 20 and 44 years of age. To prevent female teenagers from having unexpected babies, more education and other forms of help may be needed.

1.6 Conclusion

In this chapter, we first study the identification of a dose-response function with continuous treatment levels. In empirical studies, we usually have observational data. Agents can choose the levels of treatment they desire. Under the ignorability assumption, we derive moment conditions by which parameters of interest are identified with observational data. Based on the moment conditions, we propose a two-step estimator. Sufficient conditions are provided for the estimator to be consistent, converge weakly, and be semiparametric efficient. We study hypothesis testing procedures based on the two-step estimator. More specifically, we are interested in testing the null hypothesis that a DRF $\beta(t) = r(t)$ with $t \in \mathcal{T}$ for known or unknown $r(t)$. Because the parameters are infinite dimensional and the weak limits of test statistics are not standard, we use the bootstrap method when conducting inferences. Finally, we apply our estimation methods to the study of unconditional effects of mothers' weight gain during pregnancy and age on infants' birthweight, illustrating the usefulness of the new estimator.

Table 1.3: Treatment Effects of Mothers' Age

Age Change	10% Qt.	25% Qt.	50% Qt.	75% Qt.	90% Qt.	Average
15–20	-0.08	-0.05	-0.03	-0.06	-0.03	-0.06
SD.	0.03	0.02	0.02	0.02	0.02	0.02
20–25	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03
SD.	0.05	0.03	0.03	0.03	0.04	0.04
25–30	-0.03	-0.03	-0.03	-0.03	-0.02	-0.03
SD.	0.06	0.05	0.04	0.04	0.05	0.05
30–35	-0.06	-0.03	-0.03	-0.01	-0.03	-0.03
SD.	0.08	0.06	0.05	0.05	0.06	0.06
35–40	-0.03	-0.03	-0.03	-0.02	-0.03	-0.04
SD.	0.11	0.08	0.07	0.07	0.07	0.08
40–45	-0.17	-0.11	-0.06	-0.08	-0.08	-0.11
SD.	0.32	0.15	0.11	0.10	0.10	0.16
15–25	-0.11	-0.08	-0.06	-0.08	-0.06	-0.09
SD.	0.02	0.02	0.01	0.01	0.02	0.01
20–30	-0.06	-0.06	-0.06	-0.06	-0.05	-0.06
SD.	0.03	0.03	0.02	0.02	0.03	0.03
25–35	-0.08	-0.06	-0.06	-0.04	-0.05	-0.06
SD.	0.05	0.04	0.03	0.03	0.03	0.04
30–40	-0.08	-0.06	-0.06	-0.03	-0.06	-0.07
SD.	0.08	0.06	0.05	0.04	0.05	0.06
35–45	-0.20	-0.14	-0.08	-0.10	-0.11	-0.15
SD.	0.30	0.13	0.09	0.08	0.08	0.13
15–30	-0.14	-0.11	-0.08	-0.11	-0.09	-0.11
SD.	0.02	0.01	0.01	0.01	0.01	0.01
20–35	-0.11	-0.08	-0.08	-0.07	-0.08	-0.09
SD.	0.03	0.03	0.02	0.02	0.02	0.03
25–40	-0.11	-0.08	-0.08	-0.06	-0.08	-0.09
SD.	0.06	0.04	0.03	0.03	0.03	0.05
30–45	-0.25	-0.17	-0.11	-0.11	-0.13	-0.18
SD.	0.28	0.11	0.07	0.07	0.07	0.12
15–35	-0.20	-0.14	-0.11	-0.12	-0.11	-0.14
SD.	0.02	0.02	0.01	0.01	0.01	0.01
20–40	-0.14	-0.11	-0.11	-0.08	-0.11	-0.13
SD.	0.05	0.03	0.03	0.02	0.02	0.03
25–45	-0.28	-0.20	-0.14	-0.14	-0.16	-0.20
SD.	0.27	0.10	0.07	0.06	0.06	0.11
15–40	-0.23	-0.17	-0.14	-0.14	-0.14	-0.18
SD.	0.03	0.02	0.02	0.02	0.02	0.02
20–45	-0.31	-0.23	-0.17	-0.17	-0.19	-0.23
SD.	0.25	0.09	0.06	0.05	0.05	0.09
15–45	-0.40	-0.28	-0.20	-0.23	-0.22	-0.29
SD.	0.22	0.07	0.04	0.04	0.04	0.07

Chapter 2

Efficient Minimum Distance Estimator for Quantile Regression Fixed Effects Panel Data

2.1 Introduction

Quantile regression is a valuable method of statistical analysis. Particularly, conditional quantile methods are used to analyze how treatments influence the entire outcome distributions of interest. Lately, there has been an increase in literature on estimation and inference for fixed effects quantile regression panel data models (FE-QR). The FE-QR is designed to control for unobserved individual effects while exploring a range of covariate effects, and therefore provides a more complete method for the analysis of panel data. Koenker (2004) proposes a general approach for estimation of FE-QR, which treats the fixed effects (FE) as common for all conditional quantiles. Kato et al. (2012) rigorously derive the asymptotic properties of the FE-QR estimator and establish sufficient conditions for consistency and asymptotic normality. For other recent developments, see e.g. Abrevaya and Dahl (2008), Graham et al. (2009), Powell (2010), Canay (2011), Ponomareva (2011), and Rosen (2012).

However, despite these favorable asymptotic properties, there are certain drawbacks regarding the implementation of the FE-QR procedure in practice. The first is computational. In least squares applications the usual strategy would be to transform both the dependent and independent variables to deviations from individual means, and then compute the parameters of interest from the transformed data. For quantile regression this decomposition of projections is not available and one is required to deal directly with the full problem. In typical applications the number of individuals can be large and FE-QR estimator involves optimization with a large number of parameters to be estimated, which makes the problem computationally cumbersome. Also, the computation of the variance-covariance matrix for inference becomes utterly impracticable. Despite the favorable large sample properties of the quantile estimators, practical inference using FE-QR is difficult to implement in practice.

The present chapter attempts to address both the computational difficulties and this finite sample problem without sacrificing the desirable asymptotic properties of the FE-QR strategy. We propose a novel efficient minimum distance quantile regression (MD-QR) estimator. The MD-QR is defined as the weighted average of the individual QR slope estimators, with weights given by the inverses of the corresponding individual variance-covariance matrices. We provide sufficient conditions for consistency and asymptotic normality of the MD-QR when the number of individuals, n, and the number of time periods, T, grow to infinity. The limit theory allows for both sequential limits, where $T \to \infty$ followed by $n \to \infty$, and joint limits, where $T, n \to \infty$ simultaneously.

There are certain advantages of the MD-QR over the FE-QR. First, the MD-QR estimator presents an important advantage for applied researchers since it is computationally attractive. It is easy to implement in practice, interpret, and replicate. Estimation of FE-QR models for large samples can be very cumbersome. Instead of estimating all the individual specific intercepts and the slope parameters *simultaneously* as FE-QR, the MD-QR only solves regression quantiles for each individual. This procedure is especially economical for large cross-sections, and therefore computationally appealing. For the same reason, it is easier to estimate the variance-covariance matrices. Thus, it is easy to carry inference for large sample sizes. The numerical experiments confirm this assertion.

Second, the MD-QR is more efficient than the FE-QR. In fact, the MD-QR is the most efficient estimator in the class of minimum distance estimators, since its weights are the inverses of the corresponding variance-covariance matrices of the estimated parameters. To complete the argument, we show that the asymptotic linear part of the FE-QR estimator is a weighted average of the QR slope estimators. Therefore, FE-QR also belongs to the class of minimum distance estimators. Indeed, the asymptotic variance-covariance matrix of FE-QR is larger or equal than that of the MD-QR in the sense that the difference between the first and the second variance-covariance matrices is positive semi-definite. We conduct Monte Carlo simulations to evaluate the finite sample properties of the MD-QR and FE-QR estimators. The results show evidence that both estimators are approximately unbiased. In addition, the results indicate that MD-QR is more efficient than the FE-QR. We also compare both methods in terms of computation speed, and for relatively large samples the MD-QR is substantially faster. Finally, to illustrate the use of the proposed methods in practice, we apply the developed methods to Fazzari et al. (1988) investment equation model, where a firm's investment is the dependent variable and a proxy for investment demand (Tobin's q) and cash flows are independent variables. We document important heterogeneity in investment models. The results uncover strong evidence of substantial heterogeneity in the sensitivity of investment to cash flow across the conditional distribution of investment.

Another important contribution of the present work is to introduce the use of sequential limits to study QR panel data models with FE. Phillips and Moon (1999, 2000) discuss and develop a framework for analysis of mean regression panel data asymptotics with double indexed process for both sequential and joint limit theories. In this chapter, we apply these methods to the study of both FE-QR and MD-QR. Therefore, in addition to the results for the MD-QR with both sequential and joint limits, we also provide new results for consistency and asymptotic normality of the FE-QR under sequential asymptotics.[1] The first insight from this analysis is that the use of sequential limits substantially simplifies the formal derivation of the asymptotic theory for both MD-QR and FE-QR. The second significant finding is that, for the MD-QR estimator, the limiting distribution obtained under sequential limits is shown to be equal to that derived under joint limits. In other words, under the standard assumption available in the literature on the sample rate for joint limits, the limiting distribution of the MD-QR under sequential limits is equal to that under joint limits. We show that the same result also holds for FE-QR estimator.

The minimum distance (MD) estimator dates back to Berkson (1944), Neyman (1949), Taylor (1953), and Ferguson (1958), who, among others, aimed to produce computationally tractable and efficient substitutes of maximum likelihood estimators. Although very simple conceptually, MD estimation has been ingeniously used by many scholars in statistics and econometrics since Malinvaud (1970) and Rothenberg (1973). The literature on MD is vast, hence we only list a limited set of examples: Amemiya (1974, 1976, 1978), Nagaraj and Fuller (1991), Lee (1992), Koenker et al. (1994), Newey and McFadden (1994), Lehmann and Casella (1998), Moon and Schorfheide (2002), and Lee (2010). The MD estimation is a flexible methodology and has also been applied to panel data problems, examples, among others, include Chamberlain (1982, 1984), Ahn and Schmidt (1995), Hsiao et al. (2002), Hsiao (2003), Lee et al. (2012), and Moon et al. (2012).

This chapter is organized as follows: Section 2.2 introduces the model and the estimators. Section 2.3 studies the asymptotic properties of the MD-QR and FE-QR when data are independent across individuals and independent and identically distributed (i.i.d.) within each individual. Section 2.4 relaxes the i.i.d. assumption to stationary β-mixing. The finite sample properties are examined in Section 2.5.

[1]The asymptotic distribution of the FE-QR under joint limits has been derived in Kato et al. (2012).

Section 2.6 provides an illustration, and Section 2.7 concludes the chapter.

2.2 The Model and the Estimator

In this chapter, we consider a quantile regression (QR) panel data model with fixed effects (FE) as

$$Q_\tau(y_{it}|\boldsymbol{x}_{it}, \alpha_{i0}(\tau)) = \alpha_{i0}(\tau) + \boldsymbol{x}_{it}^\top \boldsymbol{\beta}_0(\tau) \equiv \boldsymbol{X}_{it}^\top \boldsymbol{\theta}_{i0}(\tau),\ t = 1, ...T,\ i = 1, ..., n, \quad (2.1)$$

where y_{it} is the response variable, $\boldsymbol{x}_{it}$ is a k dimensional vector of explanatory variables, $\boldsymbol{\beta}_0(\tau)$ is the common slope coefficients, $\alpha_{i0}(\tau)$ is the fixed effect parameter, $\boldsymbol{\theta}_{i0}(\tau) = (\alpha_{i0}(\tau), \boldsymbol{\beta}_0(\tau)^\top)^\top$, and $\boldsymbol{X}_{it}^\top = (1, \boldsymbol{x}_{it}^\top)$. The parameters can depend on the quantile index $\tau \in (0, 1)$, however, since τ is fixed throughout the chapter, we suppress such dependence on τ for simplicity when there is no confusion.

To estimate the QR model, Koenker (2004) considers the individual dummy variables estimator, which is a natural analog of the dummy variables estimator for the standard FE mean regression model. The fixed effects quantile regression (FE-QR) estimator is the defined as follows

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) := \underset{(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathcal{A}^n \times \mathcal{B}}{\arg\min} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \rho_\tau(y_{it} - \alpha_i - \boldsymbol{x}_{it}^\top \boldsymbol{\beta}),$$

where $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_n)^\top$, $\rho_\tau(u) := \{\tau - \mathbf{1}(u \leq 0)\}u$ is the check function as in Koenker and Bassett (1978), $\mathcal{A}$ is a compact subset of $\mathbb{R}$, $\mathcal{A}^n$ is the product of n copies of $\mathcal{A}$, and $\mathcal{B}$ is a compact subset of $\mathbb{R}^k$.

The fixed effect parameter α_i raises the incidental parameter problem as first noted by Neyman and Scott (1948). To overcome this drawback, it has become standard in the panel QR literature, to employ a large n and T asymptotics (see Koenker (2004) and Kato et al. (2012)). Motivated by the large time-series dimension requirement in this literature, we propose a simple to implement and efficient minimum distance QR estimator for panels with fixed effects.

We consider a minimum distance quantile regression (MD-QR) estimator, $\hat{\boldsymbol{\beta}}_{MD}$, defined as follows

$$\hat{\boldsymbol{\beta}}_{MD} = \left(\sum_{i=1}^{n} V_i^{-1} \right)^{-1} \sum_{i=1}^{n} V_i^{-1} \hat{\boldsymbol{\beta}}_i, \tag{2.2}$$

where $\hat{\boldsymbol{\beta}}_i$ is the slope coefficient estimator from each individual quantile regression problem using the time series data, and V_i denotes the associated variance-covariance matrix of $\hat{\boldsymbol{\beta}}_i$ for each individual.

However, in applications, the estimator $\hat{\boldsymbol{\beta}}_{MD}$, defined in equation (2.2), is infeasible unless V_i is known for every individual. The feasible estimator is defined with each V_i replaced by its corresponding consistent estimators $\hat{V}_i$, such that the MD-QR is given by

$$\hat{\boldsymbol{\beta}}_{MD} = \left(\sum_{i=1}^{n} \hat{V}_i^{-1} \right)^{-1} \sum_{i=1}^{n} \hat{V}_i^{-1} \hat{\boldsymbol{\beta}}_i. \tag{2.3}$$

The feasible two-step estimator can be implemented by obtaining, in the first step, consistent estimates for the slope coefficients and their associated variance-covariance matrices, $\hat{V}_i$. One can obtain such estimates from the standard quantile regression algorithm for each individual separately. In the second step, to compute the MD-QR estimator, the estimated V_i's are substituted into (2.2) and treated as if they were known to obtain (2.3).

In effect, the proposed MD-QR estimator is a product of a simple optimization problem. The MD-QR is derived from the following set of restrictions

$$\iota_n \otimes \boldsymbol{\beta}_0 - \boldsymbol{\gamma}_0 = \mathbf{0},$$

where ι_n denotes an n-vector of ones, and $\boldsymbol{\gamma}_0 = (\boldsymbol{\beta}_{10}^\top, ..., \boldsymbol{\beta}_{n0}^\top)^\top$ is a vector containing the slope coefficients from each individual. The $n \cdot k \times 1$ vector $\hat{\boldsymbol{\gamma}}$ contains the n stacked auxiliary parameter vectors. The vector ι_n imposes $(n-1) \cdot k$ restrictions on $\boldsymbol{\gamma}_0$, and $\boldsymbol{\beta}_0$ is $k \times 1$ vector of parameters of interest from (2.1). The idea is that $\boldsymbol{\gamma}_0$ consists of "reduced form" parameters from each individual regression, $\boldsymbol{\beta}_0$ consists of "structural" parameters, and ι_n gives the mapping from structure to reduced form. In other words, this restriction says that all the slope coefficients from different individuals are the same. Thus, the quantile regression fixed effects minimum distance estimator we propose results from the following minimization

$$\hat{\boldsymbol{\beta}} := \arg\min_{\boldsymbol{\beta} \in \mathcal{B}} (\hat{\boldsymbol{\gamma}} - \iota \otimes \boldsymbol{\beta})^\top W (\hat{\boldsymbol{\gamma}} - \iota \otimes \boldsymbol{\beta}),$$

where $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\beta}}_1^\top, ..., \hat{\boldsymbol{\beta}}_n^\top)^\top$, and W is a positive definite matrix. Under independence across individuals, the optimization simplifies to

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in \mathcal{B}} \sum_{i=1}^{n} (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})^\top W_i (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}),$$

and the closed form solution of the MD-QR estimator is given by

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} W_i \right)^{-1} \sum_{i=1}^{n} W_i \hat{\boldsymbol{\beta}}_i, \tag{2.4}$$

which is given in equation (2.3) by replacing W_i with the inverse of the estimated variance-covariance matrix of the individual regression parameters, $\hat{V}_i^{-1}$.

Notice that by varying W_i in (2.4) one obtains the class of minimum distance estimators, denoted by $\mathcal{M}$. A natural question is whether $\hat{\boldsymbol{\beta}}_{MD}$ is an "optimal" combination of the regression quantiles. In other words, does $\hat{\boldsymbol{\beta}}_{MD}$ have the smallest asymptotic variance-covariance matrix in $\mathcal{M}$? It turns out that the optimal weights are the inverse of the asymptotic variance-covariance matrices of the slope regression quantiles. For a proof of the inverse of the covariance matrices being optimal weights, see Rao (1965, p. 48), Serfling (1980, p. 126), or Hsiao (2003, p. 65). Thus, the answer to the posed question is that the estimator defined in (2.2) is most efficient estimator among all the estimators that are linear combinations of the regression quantiles.

Given the efficiency property of the $\hat{\boldsymbol{\beta}}_{MD}$, it is important to show that MD-QR is more efficient than the standard FE-QR. To accomplish this, we need to demonstrate that $\hat{\boldsymbol{\beta}}_{FE}$ also belongs to $\mathcal{M}$, and therefore, cannot be more efficient than the optimal $\hat{\boldsymbol{\beta}}_{MD}$. Below we show that both the FE-QR, $\hat{\boldsymbol{\beta}}_{FE}$, and the MD-QR, $\hat{\boldsymbol{\beta}}_{MD}$, are (asymptotically) within the class of minimum distance estimators, $\mathcal{M}$.

Clearly, $\hat{\boldsymbol{\beta}}_{MD}$ belongs to $\mathcal{M}$, with $W_i = V_i^{-1}$. To derive the result that $\hat{\boldsymbol{\beta}}_{FE}$ also belongs to $\mathcal{M}$, let $F_i(u|\boldsymbol{X})$ be the conditional distribution function of $u_{it} := y_{it} - \boldsymbol{X}_{it}^{\top}\boldsymbol{\theta}_{i0}$ given $\boldsymbol{X}_{it} = \boldsymbol{X}$, and have conditional density $f_i(u|\boldsymbol{X})$. Denote $\boldsymbol{\gamma}_i := \mathrm{E}[f_i(0|\boldsymbol{x}_{it})\boldsymbol{x}_{it}]/f_i(0)$ where $f_i(u)$ is the marginal distribution of u_{it}, and $\Gamma_i := \mathrm{E}[f_i(0|\boldsymbol{x}_{it})\boldsymbol{x}_{it}(\boldsymbol{x}_{it}^{\top} - \boldsymbol{\gamma}_i^{\top})]$. Now we show that $\hat{\boldsymbol{\beta}}_{FE}$ is a linear combination of weighted regression quantiles for each individual with $W_i = \Gamma_i$. Under regularity conditions and if the diverging rates of n and T satisfy the conditions of Kato et al. (2012), the FE-QR estimator has the representation

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0 &= \left(\frac{1}{n}\sum_{i=1}^{n}\Gamma_i\right)^{-1}\left\{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\tau - \mathbf{1}(u_{it} \le 0)\right](\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)\right\} + o_p(1) \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}\Gamma_i\right)^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\Gamma_i\Gamma_i^{-1}\frac{1}{T}\sum_{t=1}^{T}\left[\tau - \mathbf{1}(u_{it} \le 0)\right](\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)\right\} + o_p(1) \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}\Gamma_i\right)^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\Gamma_i(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0)\right\} + o_p(1).
\end{aligned}$$

Therefore, $\hat{\boldsymbol{\beta}}_{FE} = \left(\frac{1}{n}\sum_{i=1}^{n}\Gamma_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\Gamma_i\hat{\boldsymbol{\beta}}_i + o_p(1)$. Thus, it follows that (asymptotically) $\hat{\boldsymbol{\beta}}_{FE}$ is a linear combination (weighted average) of QR slope estimator for each individual. As described previously, V_i^{-1}'s are the optimal weights to combine $\hat{\boldsymbol{\beta}}_i$, therefore Γ_i's cannot provide better weights. Hence, $\hat{\boldsymbol{\beta}}_{FE}$ cannot be asymptotically more efficient than $\hat{\boldsymbol{\beta}}_{MD}$.

2.3 Asymptotic Theory: i.i.d. within Individuals

In this section, we investigate the asymptotic properties of the MD-QR estimator when both T and n go to infinity, both sequentially and simultaneously. For comparison and completeness we also provide parallel results for the FE-QR estimator. The results for sequential asymptotics of the FE-QR are new in the literature.

The sequential asymptotics is defined as T diverging to infinity first, and then n. In the definition of the simultaneous asymptotics, T and n tend to infinity at the same time. We do not specify the exact relationship between n and T, although we maintain that T depends on n. For notational simplicity, we suppress this dependence. For a detailed discussion on sequential and simultaneous asymptotics for panel data, see Phillips and Moon (1999, 2000). In what follows, we adopt the following notation: $(T,n)_{seq} \to \infty$ means that first $T \to \infty$ and then $n \to \infty$, while $(T,n) \to \infty$ means T and n tend to infinity simultaneously. To obtain the desired results we make the following assumptions.

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

A2 There is a constant M such that $\max_{1\le t\le T, 1\le i\le n} ||\boldsymbol{X}_{it}|| < M$.

A3 The parameter sets $\mathcal{A}$ and $\mathcal{B}$ are compact. For each $\delta > 0$,

$$\epsilon_\delta := \inf_{1\le i\le n} \inf_{||\boldsymbol{\theta}||=\delta} E\left[\int_0^{\boldsymbol{X}_{it}^\top \boldsymbol{\theta}} \{F_i(s|\boldsymbol{X}_{it}) - \tau\} ds\right] > 0, \tag{2.5}$$

where $F_i(s|\boldsymbol{X}_{it})$ is the distribution function of the innovations conditional on the covariates.

A4 The conditional density $f_i(u|\boldsymbol{X})$ is continuously differentiable for each $\boldsymbol{X}$ and i. There exist $0 < C_L \leq C_U < \infty$ such that $f_i(u|\boldsymbol{X}) \leq C_U$ uniformly over $(u, \boldsymbol{X})$ and $i \geq 1$, and $f_i(0|\boldsymbol{X}) \geq C_L$ uniformly over $\boldsymbol{X}$ and $i \geq 1$; and there exists $C_f > 0$ such that $|f_i^{(1)}(u|\boldsymbol{X})| \leq C_f$.

A5 There exists $\delta_\Omega > 0$ such that the smallest eigenvalue of $\widetilde{\Omega}_i \geq \delta_\Omega$, where $\tilde{\Omega}_i = \mathrm{E}[\boldsymbol{X}_{it}\boldsymbol{X}_{it}^\top]$.

Condition A1 assumes that the data are independent across individuals, and i.i.d. within each individual. This condition is usual in the literature and the same as assumption A1 in Kato et al. (2012). Condition A2 assumes that the covariates are uniformly bounded. This is also common in QR literature, and is imposed in A3 of Koenker (2004) and condition (a) of Theorem 1 of Chernozhukov and Hong (2002), among others. The compactness of the parameter set in condition A3 is usual and is also assumed in Hahn and Newey (2004) and Fernandez-Val (2005). Inequality (2.5) in Assumption A3 is an identification condition, the same as A3 of Kato et al. (2012). The first three assumptions are used to guarantee the uniform consistency of the regression quantiles across individuals. The uniform bound in A2 also guarantees that $\widetilde{\Omega}_i$'s exist and have a uniform bound. Condition A4 restricts the smoothness and the boundedness of the density of the innovations conditional on the covariates and its derivatives. Condition A5 assures that $\widetilde{\Omega}_i^{-1}$ are bounded uniformly across i. Also, the uniform upper and lower bounds of the continuous density functions together with assumptions A2 and A5 guarantee that both $\widetilde{\Gamma}_i = \mathrm{E}f(0|\boldsymbol{X}_{it})\boldsymbol{X}_{it}\boldsymbol{X}_{it}^\top$ and their inverses are uniformly bounded across i. Therefore, the variance-covariance matrices of the regression quantiles $\widetilde{V}_i$ and their inverses are uniformly bounded. Note that, unless $n \to \infty$, the uniform conditions across i are not really restrictive.

In applications, the variance-covariance matrices are unknown and need to be estimated. For the limiting theory where T and n tend to infinity sequentially, we make the following assumption.

A6 $\hat{V}_i = V_i + o_p(1)$ for each i as $T \to \infty$.

In the situation when n and T tend to infinity simultaneously, we impose the following condition.

A6' $\hat{V}_i = V_i + O_p(T^{-1/2}h_n^{-1/2})$ for some $h_n \downarrow 0$ uniformly across i and $\lim_{n\to\infty} \frac{n \log n}{T h_n} = 0$ as $n \to \infty$.

Examples that satisfy condition A6 are suggested by Hendricks and Koenker (1991) and Powell (1991). An example satisfying A6' is

$$\tau(1-\tau)\left(\frac{1}{T}\sum_{t=1}^{T} K_{h_n}(\hat{u}_{it})\boldsymbol{X}_{it}\boldsymbol{X}_{it}^{\top}\right)^{-1} \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_{it}\boldsymbol{X}_{it}^{\top}\left(\frac{1}{T}\sum_{t=1}^{T} K_{h_n}(\hat{u}_{it})\boldsymbol{X}_{it}\boldsymbol{X}_{it}^{\top}\right)^{-1}$$

where $\hat{u}_{it} = y_{it} - \hat{\alpha}_i - \boldsymbol{x}_{it}^{\top}\hat{\boldsymbol{\beta}}$ and $K_{h_n}(\cdot)$ is defined in Kato et al. (2012). For a study of the convergence rate of the Powell's kernel estimator, see Kato (2012).

Throughout the section, we study the properties of the feasible version of the MD-QR estimator, which is defined in equation (2.3). The consistency of the MD-QR and the FE-QR estimators are presented in Theorems 1 and 2, respectively. All proofs are collected in the Appendix B.

Theorem 1.

1. *Under conditions A1–A3 and A6,* $\hat{\boldsymbol{\beta}}_{MD} \xrightarrow{p} \boldsymbol{\beta}_0$ *as* $(T,n)_{seq} \to \infty$.

2. *Under conditions A1–A3 and A6',* $\hat{\boldsymbol{\beta}}_{MD} \xrightarrow{p} \boldsymbol{\beta}_0$ *as* $(T,n) \to \infty$ *and* $\frac{\log n}{T} \to 0$.

Theorem 2.

1. *Under conditions A1–A3,* $\hat{\boldsymbol{\beta}}_{FE} \xrightarrow{p} \boldsymbol{\beta}_0$ *as* $(T,n)_{seq} \to \infty$.

2. *Under conditions A1–A3,* $\hat{\boldsymbol{\beta}}_{FE} \xrightarrow{p} \boldsymbol{\beta}_0$ *as* $(T,n) \to \infty$ *and* $\frac{\log n}{T} \to 0$.

One can notice that the conditions required for consistency of the MD-QR estimator are similar to those of the FE-QR. This shows the close relationship between these two estimators. In addition, these results show that both estimators are consistent for the parameters of interest, i.e., the QR slope coefficients in (2.1).

Next we provide results regrading the asymptotic normality of the MD-QR and the FE-QR estimators, under sequential and simultaneous limits. The results are collected in Theorems 3 and 4 for the MD-QR and FE-QR, respectively.

Theorem 3. *Let* $V := \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} V_i^{-1}\right)^{-1}$.

1. *Under conditions A1–A6, as* $(T,n)_{seq}\to\infty$,

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{MD}-\boldsymbol{\beta}_0)\xrightarrow{d} N(0,V).$$

2. *Under conditions A1–A5 and A6', as* $(T,n)\to\infty$ *and* $\frac{n^2(\log n)}{T}\left|\log\frac{(\log n)^{0.5}}{T^{0.5}}\right|^2\to 0$,

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{MD}-\boldsymbol{\beta}_0)\xrightarrow{d} N(0,V).$$

Remark 2. *One important conclusion from Theorem 3 is that the asymptotic variance of the feasible MD-QR estimator is the same as that of the infeasible MD-QR estimator.*

Theorem 4. *Let* $\Gamma := \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Gamma_i$, *and* $\mathcal{V} := \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[(\boldsymbol{x}_{it}-\boldsymbol{\gamma}_i)(\boldsymbol{x}_{it}-\boldsymbol{\gamma}_i)^\top]$.

1. *Under conditions A1–A5, as* $(T,n)_{seq}\to\infty$,

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}_0)\xrightarrow{d} N(0,\tau(1-\tau)\Gamma^{-1}\mathcal{V}\Gamma^{-1}).$$

2. *Under conditions A1–A5, as* $(T,n)\to\infty$ *and* $\frac{n^2(\log n)}{T}\left|\log\frac{(\log n)^{0.5}}{T^{0.5}}\right|^2\to 0$,

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}_0)\xrightarrow{d} N(0,\tau(1-\tau)\Gamma^{-1}\mathcal{V}\Gamma^{-1}).$$

There are several important remarks to be observed from these results. First, from Theorems 3 and 4, one can see that both $\hat{\boldsymbol{\beta}}_{MD}$ and $\hat{\boldsymbol{\beta}}_{FE}$ are $\sqrt{nT}$ consistent. Moreover, under joint limits (part 2 of both Theorems 3 and 4), to ensure the asymptotic normality, the requirements on the diverging rates of n and T are equal for both estimators, and the same as that available in the literature, $n^2/T \to 0$ (see Kato et al. (2012)).

Second, another interesting insight from the theorems is that, given the required conditions, the limiting behavior of both estimators are identical under the different limits. In particular, parts 1 and 2 of Theorem 3 show that the MD-QR has the same limiting distribution for both sequential and joint limits. The analogous result is shown in Theorem 4 for the FE-QR estimator, that is, FE-QR has same limiting distribution under sequential and joint limits. This result is somewhat surprising since a joint limit usually gives a more robust result than a sequential limit as stated in Phillips and Moon (2000). We conjecture that the equality of the limiting distributions under different limits is a result of the very stringent conditions required to achieve asymptotic normality under joint limits.

Third, the mathematical proofs are greatly simplified for the sequential limits (as shown in the Appendix B). In all the theorems above, we provide results for both sequential and simultaneous asymptotics. For the sequential limits asymptotics, we let T tend to infinity, and then n. This view of double indexes asymptotics simplifies the proofs substantially, and provides valuable insights of the results, although sequential limits could give deceptive asymptotic results. In contrast, the view of simultaneous asymptotics is more general, although it is significantly more difficult to obtain even with more stringent assumptions. In our case, the cost is related to the requirements on the diverging rates of T and n. In the results listed above, we note that for the simultaneous asymptotics, roughly speaking, the best can be done is to require T tending to infinity faster than $n^2 \log n$. Thus, given the stringent requirement on the growth rate of T under the joint limits, we believe the use of sequential asymptotics is an important tool and provides useful approximations for QR panel fixed effects analysis, and dramatically decreases the complexity of the proofs. The scope for the use of sequential asymptotics in extensions of QR panel models with FE is large; for instance, censored, duration, and survival models are examples of central models that remain to be formally developed.

Finally, from the discussion in the previous section on optimal weights for MD estimation, $\hat{\boldsymbol{\beta}}_{MD}$ is an appealing alternative estimator for QR models with fixed effects, in the sense that $\hat{\boldsymbol{\beta}}_{FE}$ cannot be (asymptotically) more efficient than $\hat{\boldsymbol{\beta}}_{MD}$. Therefore, MD-QR estimator, together with its associated inference, is a compelling alternative for practitioners since it is the most efficient estimator among the minimum distance estimators class to which FE-QR estimator belongs.

Inference for the MD-QR estimator is simple. We now discuss the estimator of the asymptotic covariance matrix. An easy-to-implement consistent estimator of the variance-covariance matrix V described in Theorem 3 is $\left(\frac{1}{n}\sum_{i=1}^{n} V_i^{-1}\right)^{-1}$. Since V_i is not known in general, consistent estimators $\hat{V}_i$ used for the computation of the MD-QR estimator could be plugged in. Therefore, a consistent estimator $\hat{V} = \left(\frac{1}{n}\sum_{i=1}^{n} \hat{V}_i^{-1}\right)^{-1}$. One can notice that this matrix is simple to compute and inference is specially easy for large cross-sections case. Inference for the FE-QR is described in Kato et al. (2012).

2.4 Asymptotic Theory: Extensions to Dependent Data

The i.i.d. requirement for data with in each individual is relaxed to stationary and β-mixing. We still maintain that the data are independent across individuals. The following conditions are imposed for the MD estimator to have desirable properties.

B1 For each i, $\{(y_{it}, \boldsymbol{X}_{it}), t \geq 1\}$ is stationary and β-mixing time series with β-mixing coefficient $\beta_i(j)$. There exist constants $a \in (0,1)$ and $B > 0$ such that $\sup_{i \geq 1} \beta_i(j) \leq Ba^j$ for all $j \geq 1$.

B2 $f_{i,j}(u_1, u_{1+j}|\boldsymbol{X}_1, \boldsymbol{X}_{1+j})$ is uniformly bounded with respect to all the four variables, where $f_{i,j}(u_1, u_{1+j}|\boldsymbol{X}_1, \boldsymbol{X}_{1+j})$ is the conditional density of (u_1, u_{1+j}) given $(\boldsymbol{X}_{i,1}, \boldsymbol{X}_{i,1+j}) = (\boldsymbol{X}_1, \boldsymbol{X}_{1+j})$.

Condition B1 relaxes the assumption of i.i.d. within each individual to that of stationary β-mixing which is used in Kato et al. (2012) and is similar to Hahn and Kuersteiner (2011). Condition B2 is needed because the data are not i.i.d. and we need to impose a condition on the joint distributions.

The following two theorems are asymptotic results for stationary β-mixing data, and are extensions of the theorems in Section 3.

Theorem 5.

1. *Under conditions A2–A6 and B1–B2, we have* $\hat{\boldsymbol{\beta}}_{MD} \xrightarrow{p} \boldsymbol{\beta}_0$ *and*

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{MD} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, V).$$

   as $(T, n)_{seq} \to \infty$.

2. *Under conditions A2–A6, and B1–B2,* $\hat{\boldsymbol{\beta}}_{MD} \xrightarrow{p} \boldsymbol{\beta}_0$ *as* $(T, n) \to \infty$ *and* $\frac{\log n}{T} \to 0$. *In addition, if* $\frac{n^2(\log n)}{T} \left| \log \frac{(\log n)^{0.5}}{T^{0.5}} \right|^2 \to 0$, *then*

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{MD} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, V).$$

Theorem 6. *Let* $\Gamma := \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Gamma_i$, *and* $\dot{\mathcal{V}} := \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \text{Var}[T^{-1/2} \sum_{t=1}^{T} \{\tau - 1(u_{it} \leq 0)\}(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)]$.

1. *Under conditions A2–A5 and B1–B2,* $\hat{\boldsymbol{\beta}}_{FE} \xrightarrow{p} \boldsymbol{\beta}_0$ *and*

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \tau(1-\tau)\Gamma^{-1}\dot{\mathcal{V}}\Gamma^{-1}).$$

   as $(T, n)_{seq} \to \infty$.

2. *Under conditions A2–A5 and B1–B2,* $\hat{\boldsymbol{\beta}}_{FE} \xrightarrow{p} \boldsymbol{\beta}_0$ *as* $(T, n) \to \infty$ *and* $\frac{\log n}{T} \to 0$. *In addition, if* $\frac{n^2(\log n)}{T} \left|\log \frac{(\log n)^{0.5}}{T^{0.5}}\right|^2 \to 0$, *then*

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \tau(1-\tau)\Gamma^{-1}\dot{\mathcal{V}}\Gamma^{-1}).$$

The results in Theorem 5 require condition A6 or A6' to hold for stationary and β-mixing data. An example of the estimator is

$$\left(\frac{1}{T}\sum_{t=1}^{T} K_{h_n}(\hat{u}_{it})\boldsymbol{X}_{it}\boldsymbol{X}_{it}^{\top}\right)^{-1} \hat{\Omega}_{ni} \left(\frac{1}{T}\sum_{t=1}^{T} K_{h_n}(\hat{u}_{it})\boldsymbol{X}_{it}\boldsymbol{X}_{it}^{\top}\right)^{-1},$$

where

$$\hat{\Omega}_{ni} = \frac{\tau(1-\tau)}{T}\sum_{t=1}^{T}\boldsymbol{X}_{it}\boldsymbol{X}_{it}^{\top} + \sum_{1\le|j|\le m_n}\left(1 - \frac{|j|}{T}\right)\left[\frac{1}{T}\sum_{t=\max\{1,-j+1\}}^{\min\{T,T-j\}}\{\tau - 1(\hat{u}_{it} \le 0)\}\{\tau - 1(\hat{u}_{i,t+j} \le 0)\}\boldsymbol{X}_{it}\boldsymbol{X}_{it}^{\top}\right],$$

as discussed in Remark 3.2, the proof of Theorem 3.2 of Kato and Galvao (2010). Kato (2012) also studies the rate of convergence of the Powell's kernel estimator for stationary β-mixing data.

2.5 Monte Carlo Simulations

This section conducts simulations to investigate finite sample performance of the MD-QR estimator. For comparison purposes, we also report simulations for the FE-QR estimator. We consider location and location-scale models as listed below

1. Location shift model: $y_{it} = \eta_i + \beta x_{it} + \epsilon_{it}$,

2. Location-scale shift model: $y_{it} = \eta_i + \beta x_{it} + (1 + 0.1x_{it})\epsilon_{it}$,

where $x_{it} = 0.3\eta_i + z_{it}$, $z_{it} \sim$ i.i.d.χ^2_3, $\eta_i \sim$ i.i.d.$N(0, 1)$, and the innovations ϵ_{it} follow a general distribution F. The correlation between η_i and x_{it} makes the random effects estimators to be inconsistent. We consider some selected F distribution functions as standard normal $[N(0, 1)]$, t-distribution with three degrees of freedom $[t(3)]$, and χ^2-distribution with three degrees of freedom $[\chi^2(3)]$. The slope parameter of interest is $\beta = 1$. Thus, in the location shift model, the parameters in equation (2.1) are $\alpha_{i0} = \alpha_{i0}(\tau) = \eta_i + F^{-1}(\tau)$ and $\beta_0(\tau) = 1$. In the location-scale shift model, $\alpha_{i0} = \alpha_{i0}(\tau) = \eta_i + F^{-1}(\tau)$ and $\beta_0(\tau) = 1 + 0.1F(\tau)^{-1}$. We consider several sample sizes and quantiles, where $n \in \{50, 100, 200\}$, $T \in \{50, 75, 100\}$, and $\tau \in \{0.25, 0.5, 0.75\}$. The number of replications is 2,000 in all cases.

In the numerical study, we compute the MD-QR introduced above and the FE-QR. For comparison, we also report results for the MD-QR estimator generated with weights computed using the true sparsity function instead of the estimated sparsity in the corresponding variance-covariance matrix. It is important to present results for the MD-QR with the true sparsity to investigate whether the finite sample performance of the estimator is affected by estimation of the sparsity function in the weights.

We use the following abbreviations: "MDT" stands for MD-QR using the true sparsity; "MDE" stands for MD-QR using the estimated sparsity; and "FE" stands for the FE-QR estimator. For the MDE we estimate the sandwich variance-covariance matrix using kernel estimation, and set the bandwidth $h = 1.3h_{HS}$, where h_{HS} is the Hall-Sheather bandwidth. Tables 2.1–2.3 report the bias and standard deviation (SD) of the MDT, MDE, and FE estimators for the location model. Tables 2.4–2.6 report the analogous results for the location-scale model. Tables 2.7–2.12 report the average, over the number of replications, of the estimated standard deviations of each estimator.

2.5.1 Bias and SD Results

Table 2.1 reports the bias and SD of the estimators for the location shift model with $N(0,1)$ innovations. From the top panel, we see that the estimators are all approximately unbiased, particularly when $\tau = 0.5$. Comparing the MDT and MDE, we find that the bias of MDT is in general lower than or equal to that of MDE, which is due to the estimation effect of the sparsity. However, this effect disappears as the time dimension increases, and the results for MDE are very similar to those for MDT. In addition, as expected, we observe that the bias decreases as T increases for all the estimators but not as n increases. This is due to the incidental parameter problem discussed above. The SD's are reported in the bottom panel. The results support our claim that FE-QR cannot be more efficient than the MD-QR. MDT has same SD as FE, and for two cases MDE has smaller SD than FE ($\tau = 0.25$, $n = 200$, $T = 75$; and $\tau = 0.75$, $n = 200$, $T = 50$). Moreover, the SD decreases as either T or n increases. Although the size of n does not affect the bias of MD-QR and FE-QR estimators significantly, it does improve the SD of the estimators.

Tables 2.2 and 2.3 report the simulation results for the location shift model with $t(3)$ and $\chi^2(3)$ innovations, respectively. From the results, we observe similar patterns of the bias and SD for all the estimators. All estimators are approximately unbiased, with bias reducing as T increases. In addition, there is evidence of smaller SD for MDE relative to FE for $\tau = 0.75$ in the $\chi^2(3)$ case displayed in Table 2.3.

Table 2.1: Bias and SD of the QR Estimators of the Location Shift Model When the Innovations Are $N(0, 1)$

			$\tau = 0.25$			τ =0.5			τ =0.75		
	n	T	MDT	MDE	FE	MDT	MDE	FE	MDT	MDE	FE
Bias	50	50	0.002	0.006	0.001	0.000	0.000	-0.000	-0.002	-0.006	-0.000
		75	0.002	0.005	-0.000	0.000	-0.000	-0.000	-0.002	-0.005	0.000
		100	0.001	0.004	-0.000	-0.000	0.000	-0.000	-0.001	-0.004	0.000
	100	50	0.002	0.006	0.000	0.000	0.000	-0.000	-0.002	-0.006	-0.000
		75	0.001	0.005	0.000	-0.000	0.000	-0.000	-0.001	-0.005	-0.000
		100	0.001	0.005	-0.000	0.000	-0.000	-0.000	-0.001	-0.005	-0.000
	200	50	0.002	0.006	0.000	0.000	-0.000	0.000	-0.002	-0.006	-0.000
		75	0.002	0.005	0.000	-0.000	0.000	-0.000	-0.002	-0.005	-0.000
		100	0.001	0.005	0.000	0.000	0.000	0.000	-0.001	-0.005	-0.000
SD	50	50	0.011	0.011	0.011	0.010	0.011	0.010	0.011	0.011	0.011
		75	0.009	0.009	0.009	0.009	0.008	0.009	0.009	0.009	0.009
		100	0.008	0.008	0.008	0.007	0.007	0.008	0.008	0.008	0.008
	100	50	0.008	0.008	0.008	0.007	0.007	0.007	0.008	0.008	0.008
		75	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
		100	0.006	0.006	0.006	0.005	0.005	0.005	0.006	0.006	0.006
	200	50	0.006	0.006	0.006	0.005	0.005	0.005	0.006	0.005	0.006
		75	0.005	0.004	0.005	0.004	0.004	0.004	0.005	0.005	0.005
		100	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004

For the location-scale model, the results of bias and SD with $N(0, 1)$, $t(3)$, and $\chi^2(3)$ innovations are reported in Tables 2.4–2.6, respectively. Basically, they are parallel to those for the location shift model. In general, MDT and MDE estimators are approximately unbiased. The effects of estimation of sparsity disappear as sample size increase, and the SD's of the MD estimators are not larger than the corresponding FE.

2.5.2 The Estimators of SD

Table 2.7 reports the average of the estimated SD of MDT, MDE, and FE estimators for location shift model with $N(0, 1)$ innovations. We can see that the estimates of the SD for MDT and FE are very close to the SD in the lower panels of Table 2.1, which implies that the estimators for the standard errors are approximately unbiased. As for the MDE, the SD estimators slightly overestimate when $T = 50$.

Table 2.2: Bias and SD of the QR Estimators of the Location Shift Model When the Innovations Are $t(3)$

			τ =0.25			τ =0.5			τ =0.75		
	n	T	MDT	MDE	FE	MDT	MDE	FE	MDT	MDE	FE
Bias	50	50	-0.002	0.004	-0.000	-0.000	0.000	0.000	0.001	-0.004	-0.000
		75	-0.001	0.004	-0.000	-0.000	0.000	-0.000	0.002	-0.004	0.000
		100	-0.001	0.004	0.000	0.000	0.000	0.000	0.001	-0.004	-0.000
	100	50	-0.001	0.004	0.000	-0.000	-0.000	0.000	0.001	-0.004	-0.000
		75	-0.001	0.005	-0.000	0.000	-0.000	0.000	0.001	-0.004	-0.000
		100	-0.001	0.004	-0.000	-0.000	0.000	0.000	0.001	-0.004	0.000
	200	50	-0.001	0.005	-0.000	0.000	-0.000	0.000	0.001	-0.004	-0.000
		75	-0.001	0.004	0.000	-0.000	0.000	0.000	0.001	-0.005	0.000
		100	-0.001	0.004	0.000	0.000	-0.000	-0.000	0.001	-0.004	0.000
SD	50	50	0.015	0.013	0.014	0.012	0.012	0.012	0.015	0.014	0.014
		75	0.012	0.011	0.012	0.009	0.009	0.009	0.012	0.011	0.011
		100	0.010	0.010	0.010	0.008	0.008	0.008	0.010	0.010	0.010
	100	50	0.010	0.010	0.010	0.008	0.008	0.008	0.010	0.010	0.010
		75	0.008	0.008	0.008	0.007	0.007	0.006	0.008	0.008	0.008
		100	0.007	0.007	0.007	0.006	0.006	0.006	0.007	0.007	0.007
	200	50	0.007	0.007	0.007	0.006	0.006	0.006	0.007	0.007	0.007
		75	0.006	0.005	0.006	0.005	0.005	0.004	0.006	0.006	0.006
		100	0.005	0.005	0.005	0.004	0.004	0.004	0.005	0.005	0.005

But as T increases, the overestimation decreases, and almost disappears when $T = 200$. (We did simulations for $T = 200$. To save space, we do not report this result.) This reflects the requirement of the size of the time dimension for the estimation of the SD. The standard deviations of the estimated SD of MDT, MDE, and FE are all very close to zero, therefore we do not report them in tables to save space. Tables 2.8 and 2.9 report the average of the estimated SD of MDT, MDE, and FE estimators for location shift models with $t(3)$ and $\chi^2(3)$ innovations, respectively, and we observe similar patterns to Table 2.7. Results for the location-scale shift model are reported in Tables 2.10–2.12, respectively. The results are parallel to those for the location shift model. Overall, our simulations show that the estimators of the SD are approximately unbiased.

Table 2.3: Bias and SD of the QR Estimators of the Location Shift Model When the Innovations Are $\chi^2(3)$

			τ =0.25			τ =0.5			τ =0.75		
	n	T	MDT	MDE	FE	MDT	MDE	FE	MDT	MDE	FE
Bias	50	50	0.013	0.014	0.000	0.011	0.010	0.000	0.009	-0.008	0.001
		75	0.009	0.011	0.000	0.007	0.006	0.001	0.006	-0.013	0.000
		100	0.007	0.008	0.000	0.006	0.004	0.000	0.004	-0.016	-0.001
	100	50	0.013	0.014	0.001	0.011	0.010	0.000	0.009	-0.008	0.000
		75	0.009	0.010	0.000	0.008	0.006	0.000	0.006	-0.013	-0.000
		100	0.007	0.008	0.000	0.006	0.004	-0.000	0.006	-0.015	-0.000
	200	50	0.013	0.014	0.000	0.012	0.009	0.000	0.009	-0.008	0.001
		75	0.009	0.010	0.000	0.008	0.006	0.000	0.006	-0.013	-0.000
		100	0.007	0.008	0.000	0.006	0.004	-0.000	0.005	-0.015	-0.000
SD	50	50	0.015	0.016	0.015	0.022	0.022	0.022	0.033	0.031	0.034
		75	0.012	0.013	0.012	0.018	0.018	0.018	0.028	0.026	0.028
		100	0.011	0.011	0.011	0.015	0.015	0.015	0.024	0.022	0.024
	100	50	0.011	0.012	0.010	0.016	0.015	0.016	0.023	0.021	0.023
		75	0.009	0.009	0.008	0.012	0.013	0.013	0.019	0.018	0.020
		100	0.008	0.008	0.007	0.011	0.011	0.011	0.017	0.015	0.017
	200	50	0.007	0.008	0.007	0.011	0.011	0.011	0.017	0.015	0.017
		75	0.006	0.007	0.006	0.009	0.009	0.009	0.014	0.013	0.014
		100	0.005	0.005	0.005	0.008	0.008	0.008	0.012	0.011	0.012

2.5.3 Estimation Speed

Estimation of FE-QR models can be very cumbersome. It is important for applied researchers to have available estimators that are easy to implement and compute. The MD-QR is expected to be computationally attractive to practitioners. Thus, we report the computing time for the MD-QR and the FE-QR estimators and the associated standard deviations. Since the computing speed varies and depends on the computer and the software, the following table provides some parameters of the hardware and software we use.

Processor Speed:	2.93 GHz
Memory:	16 GB
Processor Interconnect Speed:	4.8 GT/s
`R` version:	2.14.1
`quantreg` version:	4.77

Table 2.4: Bias and SD of the QR Estimators of the Location-Scale Shift Model When the Innovations Are $N(0,1)$

			τ =0.25			τ =0.5			τ =0.75		
	n	T	MDT	MDE	FE	MDT	MDE	FE	MDT	MDE	FE
Bias	50	50	0.006	0.018	0.001	0.001	-0.001	-0.000	-0.006	-0.018	-0.001
		75	0.004	0.014	0.000	0.000	0.000	0.001	-0.004	-0.014	-0.001
		100	0.003	0.011	0.001	0.000	-0.000	-0.000	-0.003	-0.011	-0.001
	100	50	0.006	0.018	0.001	-0.000	-0.000	0.000	-0.006	-0.018	-0.001
		75	0.004	0.014	0.001	0.000	0.000	-0.000	-0.004	-0.014	-0.000
		100	0.003	0.010	0.001	-0.000	-0.000	-0.000	-0.003	-0.011	-0.000
	200	50	0.006	0.018	0.001	-0.000	0.000	0.000	-0.006	-0.018	-0.001
		75	0.004	0.014	0.001	-0.000	-0.000	0.000	-0.004	-0.013	-0.001
		100	0.003	0.011	0.001	0.000	-0.000	0.000	-0.003	-0.011	-0.001
SD	50	50	0.018	0.019	0.018	0.017	0.018	0.017	0.018	0.019	0.018
		75	0.015	0.015	0.015	0.014	0.014	0.014	0.015	0.015	0.015
		100	0.013	0.013	0.013	0.012	0.012	0.012	0.013	0.013	0.013
	100	50	0.013	0.013	0.013	0.012	0.012	0.012	0.013	0.013	0.013
		75	0.010	0.011	0.010	0.009	0.010	0.010	0.010	0.011	0.011
		100	0.009	0.009	0.009	0.008	0.009	0.008	0.009	0.009	0.009
	200	50	0.009	0.009	0.009	0.008	0.009	0.008	0.009	0.009	0.009
		75	0.007	0.008	0.008	0.007	0.007	0.007	0.007	0.008	0.007
		100	0.006	0.007	0.006	0.006	0.006	0.006	0.006	0.007	0.006

We report results from one replication of the above simulation where we estimate a location model for one particular quantile ($\tau = 0.5$) using different sample sizes. We compare the time in terms of three quantities: user, system, and elapsed. "User" represents the CPU time spent executing the user instructions of the calling process, and "system" is the CPU time charged by the system on behalf of the calling process. "Elapsed" is the most interesting and describes the time required for one replication of the simulation. The durations (in seconds) of the MD-QR and FE-QR estimations for sample sizes $n = T \in \{10, 50, 100, 250, 500, 1000\}$ are listed in the following table.

Table 2.5: Bias and SD of the QR Estimators of the Location-Scale Shift Model When the Innovations Are $t(3)$

			τ =0.25			τ =0.5			τ =0.75		
	n	T	MDT	MDE	FE	MDT	MDE	FE	MDT	MDE	FE
Bias	50	50	0.001	0.020	0.003	0.000	0.000	-0.000	-0.001	-0.021	-0.001
		75	0.001	0.016	0.000	0.000	0.000	0.000	-0.001	-0.018	-0.001
		100	-0.000	0.014	0.001	0.001	-0.001	-0.000	-0.000	-0.014	-0.001
	100	50	0.001	0.021	0.002	-0.000	0.000	0.000	-0.001	-0.021	-0.002
		75	0.000	0.017	0.001	0.000	0.000	0.000	-0.000	-0.017	-0.001
		100	0.000	0.014	0.001	-0.000	-0.000	-0.000	-0.000	-0.014	-0.001
	200	50	0.001	0.021	0.002	0.000	-0.000	0.000	-0.000	-0.021	-0.002
		75	0.000	0.017	0.001	0.000	-0.000	-0.000	-0.000	-0.017	-0.001
		100	0.000	0.015	0.001	-0.000	0.000	-0.000	-0.000	-0.014	-0.001
SD	50	50	0.023	0.024	0.023	0.019	0.020	0.018	0.023	0.023	0.022
		75	0.020	0.018	0.018	0.015	0.015	0.015	0.019	0.019	0.018
		100	0.016	0.016	0.016	0.013	0.013	0.013	0.016	0.016	0.016
	100	50	0.017	0.016	0.016	0.014	0.014	0.013	0.017	0.016	0.016
		75	0.013	0.013	0.013	0.011	0.011	0.010	0.013	0.013	0.013
		100	0.011	0.011	0.011	0.009	0.010	0.009	0.011	0.011	0.011
	200	50	0.012	0.011	0.012	0.010	0.010	0.009	0.011	0.011	0.011
		75	0.009	0.009	0.009	0.008	0.008	0.007	0.010	0.010	0.009
		100	0.008	0.008	0.008	0.007	0.007	0.006	0.008	0.008	0.008

	MD-QR			FE-QR		
	user	system	elapsed	user	system	elapsed
$n = T = 10$	0.04	0.00	0.04	0.01	0.00	0.01
$n = T = 50$	0.18	0.00	0.18	0.15	0.01	0.16
$n = T = 100$	0.39	0.01	0.40	1.99	0.07	2.06
$n = T = 250$	1.19	0.06	1.25	111.94	1.08	113.07
$n = T = 500$	3.10	0.61	3.71	5035.23	10.29	5045.61
$n = T = 1000$	9.92	6.88	59.29	144133.94	888.38	153419.38

From the table, one can see that as the sample sizes increase, the durations of computing each of the estimators increase, although the increment is mild for MD-QR, it is drastic for FE-QR. The large difference in the duration of computing the estimates comes from the fact that FE-QR solves a single larger optimization problem while MD-QR splits the data into smaller parts and estimate each of them

Table 2.6: Bias and SD of the QR Estimators of the Location-Scale Shift Model When the Innovations Are $\chi^2(3)$

			τ =0.25			τ =0.5			τ =0.75		
	n	T	MDT	MDE	FE	MDT	MDE	FE	MDT	MDE	FE
Bias	50	50	0.022	0.024	0.000	0.017	0.003	-0.003	0.003	-0.064	-0.006
		75	0.015	0.014	0.000	0.010	-0.002	-0.001	0.003	-0.063	-0.003
		100	0.011	0.011	0.000	0.009	-0.005	-0.001	0.001	-0.055	-0.002
	100	50	0.022	0.024	-0.000	0.015	0.002	-0.002	0.003	-0.062	-0.008
		75	0.015	0.014	-0.000	0.010	-0.002	-0.001	0.003	-0.061	-0.003
		100	0.011	0.011	-0.000	0.008	-0.004	-0.001	0.002	-0.056	-0.002
	200	50	0.022	0.023	-0.000	0.015	0.001	-0.002	0.004	-0.063	-0.006
		75	0.015	0.015	-0.000	0.011	-0.003	-0.001	0.003	-0.062	-0.004
		100	0.011	0.010	0.000	0.008	-0.004	-0.001	0.001	-0.055	-0.003
SD	50	50	0.025	0.026	0.024	0.036	0.038	0.035	0.054	0.051	0.055
		75	0.020	0.021	0.019	0.029	0.029	0.028	0.045	0.044	0.045
		100	0.017	0.018	0.017	0.025	0.026	0.025	0.040	0.040	0.040
	100	50	0.017	0.018	0.017	0.026	0.026	0.025	0.039	0.036	0.038
		75	0.014	0.015	0.014	0.020	0.021	0.021	0.032	0.031	0.032
		100	0.012	0.013	0.012	0.018	0.018	0.018	0.028	0.027	0.027
	200	50	0.012	0.013	0.012	0.018	0.019	0.017	0.027	0.026	0.027
		75	0.010	0.011	0.010	0.014	0.015	0.015	0.022	0.022	0.022
		100	0.008	0.009	0.008	0.013	0.013	0.013	0.019	0.019	0.020

individually.

A natural question that follows is whether the duration of the estimation is more sensitive to T or n. The following table shows the durations (in seconds) for various levels of T when $n = 100$, and for various levels of n when $T = 100$. From the table one can see that the durations of computing both of the estimators are much more sensitive to the size of n. Moreover, the sensitivity to the sample size is much lower for MD-QR than for FE-QR.

Overall, the computing time of MD-QR is considerably smaller than that of FE-QR estimator. When the sample size, especially n, is large there is a strong preference toward the MD-QR estimator. Thus, we expect the MD-QR to be very useful for applied scholars seeking to estimate QR panels with fixed effects under relatively large cross-section dimension.

Table 2.7: Average of the Estimated SD of the Location Shift Model When the Innovations Are $N(0,1)$

		τ =0.25			τ =0.5			τ =0.75		
n	T	MDT	MDE	FE	MDT	MDE	FE	MDT	MDE	FE
50	50	0.011	0.021	0.011	0.010	0.021	0.011	0.011	0.021	0.012
	75	0.009	0.014	0.009	0.008	0.015	0.009	0.009	0.014	0.009
	100	0.008	0.011	0.008	0.007	0.012	0.008	0.008	0.011	0.008
100	50	0.008	0.015	0.008	0.007	0.015	0.008	0.008	0.015	0.008
	75	0.006	0.010	0.007	0.006	0.011	0.006	0.006	0.010	0.007
	100	0.006	0.008	0.006	0.005	0.009	0.005	0.006	0.008	0.006
200	50	0.006	0.010	0.006	0.005	0.010	0.005	0.006	0.010	0.006
	75	0.005	0.007	0.005	0.004	0.008	0.004	0.005	0.007	0.005
	100	0.004	0.006	0.004	0.004	0.006	0.004	0.004	0.006	0.004

Table 2.8: Average of the Estimated SD of the Location Shift Model When the Innovations Are $t(3)$

		τ =0.25			τ =0.5			τ =0.75		
n	T	MDT	MDE	FE	MDT	MDE	FE	MDT	MDE	FE
50	50	0.014	0.025	0.014	0.011	0.025	0.012	0.014	0.025	0.014
	75	0.011	0.017	0.011	0.009	0.019	0.010	0.011	0.017	0.012
	100	0.010	0.013	0.010	0.008	0.015	0.008	0.010	0.013	0.010
100	50	0.010	0.018	0.010	0.008	0.018	0.008	0.010	0.018	0.010
	75	0.008	0.012	0.008	0.006	0.013	0.007	0.008	0.012	0.008
	100	0.007	0.010	0.007	0.006	0.011	0.006	0.007	0.010	0.007
200	50	0.007	0.013	0.007	0.006	0.013	0.006	0.007	0.013	0.007
	75	0.006	0.008	0.006	0.005	0.009	0.005	0.006	0.008	0.006
	100	0.005	0.007	0.005	0.004	0.007	0.004	0.005	0.007	0.005

2.6 Application

In this section we apply the developed estimators to Fazzari et al. (1988) investment equation model, where a firm's investment is the dependent variable, and a proxy for investment demand (Tobin's q) and cash flows are independent variables. As stated in Almeida et al. (2010), "following Fazzari et al. (1988), investment-cash-flow sensitivities became a standard metric in the literature that examines the impact of financing imperfections on corporate investment (Stein (2003)). These empirical sensitivities are also used for drawing inferences about efficiency in internal capital

Table 2.9: Average of the Estimated SD of the Location Shift Model When the Innovations Are $\chi^2(3)$

		τ =0.25			τ =0.5			τ =0.75		
n	T	MDT	MDE	FE	MDT	MDE	FE	MDT	MDE	FE
50	50	0.015	0.045	0.017	0.022	0.045	0.022	0.034	0.045	0.033
	75	0.012	0.030	0.013	0.018	0.033	0.018	0.028	0.031	0.027
	100	0.010	0.023	0.011	0.015	0.027	0.015	0.024	0.025	0.023
100	50	0.011	0.031	0.011	0.016	0.032	0.015	0.024	0.032	0.024
	75	0.009	0.021	0.009	0.013	0.023	0.013	0.020	0.022	0.020
	100	0.007	0.017	0.008	0.011	0.019	0.011	0.017	0.017	0.017
200	50	0.007	0.022	0.008	0.011	0.023	0.011	0.017	0.022	0.017
	75	0.006	0.015	0.006	0.009	0.017	0.009	0.014	0.015	0.014
	100	0.005	0.012	0.005	0.008	0.013	0.008	0.012	0.012	0.012

Table 2.10: Average of the Estimated SD of the Location-Scale Shift Model When the Innovations Are $N(0,1)$

		τ =0.25			τ =0.5			τ =0.75		
n	T	MDT	MDE	FE	MDT	MDE	FE	MDT	MDE	FE
50	50	0.018	0.029	0.018	0.017	0.029	0.017	0.018	0.029	0.018
	75	0.015	0.020	0.015	0.014	0.021	0.014	0.015	0.020	0.015
	100	0.013	0.016	0.013	0.012	0.018	0.012	0.013	0.016	0.013
100	50	0.013	0.020	0.013	0.012	0.020	0.012	0.013	0.020	0.013
	75	0.010	0.014	0.011	0.010	0.015	0.010	0.010	0.014	0.011
	100	0.009	0.011	0.009	0.008	0.012	0.008	0.009	0.011	0.009
200	50	0.009	0.014	0.009	0.008	0.014	0.008	0.009	0.014	0.009
	75	0.007	0.010	0.007	0.007	0.011	0.007	0.007	0.010	0.007
	100	0.006	0.008	0.006	0.006	0.009	0.006	0.006	0.008	0.006

markets (Lamont (1997); Shin and Stulz (1998)), the effect of agency on corporate spending (Hadlock (1998); Bertrand and Mullainathan (2005)), the role of business groups in capital allocation (Hoshi et al. (1991)), and the effect of managerial characteristics on corporate policies (Bertrand and Schoar (2003); Malmendier and Tate (2005))." Following the literature, the model in our application is

$$IK_{it} = \alpha_i + \theta q_{it} + \gamma CFK_{it} + u_{it}, \tag{2.6}$$

Table 2.11: Average of the Estimated SD of the Location-Scale Shift Model When the Innovations Are $t(3)$

		τ =0.25			τ =0.5			τ =0.75		
n	T	MDT	MDE	FE	MDT	MDE	FE	MDT	MDE	FE
50	50	0.022	0.035	0.023	0.018	0.035	0.020	0.022	0.035	0.023
	75	0.018	0.024	0.019	0.015	0.026	0.016	0.018	0.024	0.019
	100	0.016	0.020	0.016	0.013	0.021	0.013	0.016	0.019	0.016
100	50	0.016	0.025	0.016	0.013	0.025	0.014	0.016	0.025	0.016
	75	0.013	0.017	0.013	0.010	0.018	0.011	0.013	0.017	0.013
	100	0.011	0.014	0.011	0.009	0.015	0.009	0.011	0.014	0.011
200	50	0.011	0.017	0.011	0.009	0.017	0.009	0.011	0.017	0.011
	75	0.009	0.012	0.009	0.007	0.013	0.008	0.009	0.012	0.009
	100	0.008	0.010	0.008	0.006	0.011	0.007	0.008	0.010	0.008

Table 2.12: Average of the Estimated SD of the Location-Scale Shift Model When the Innovations Are $\chi^2(3)$

		τ =0.25			τ =0.5			τ =0.75		
n	T	MDT	MDE	FE	MDT	MDE	FE	MDT	MDE	FE
50	50	0.024	0.061	0.026	0.036	0.061	0.035	0.056	0.062	0.053
	75	0.020	0.041	0.021	0.029	0.045	0.029	0.045	0.044	0.045
	100	0.017	0.033	0.018	0.025	0.037	0.025	0.039	0.036	0.038
100	50	0.017	0.043	0.018	0.025	0.043	0.025	0.039	0.044	0.038
	75	0.014	0.029	0.014	0.020	0.032	0.020	0.032	0.031	0.032
	100	0.012	0.023	0.012	0.018	0.026	0.018	0.028	0.026	0.027
200	50	0.012	0.030	0.012	0.018	0.031	0.017	0.028	0.031	0.027
	75	0.010	0.021	0.010	0.014	0.023	0.014	0.023	0.022	0.023
	100	0.009	0.016	0.009	0.013	0.019	0.012	0.020	0.018	0.019

where the quantity $IK_{it} = I_{it}/K_{i,t-1}$ and $CFK_{it} = CF_{it}/K_{i,t-1}$ so that the base capital stock comes from the previous period, with I denoting investment, K capital stock, q the average Tobin's q, CF cash flow, α firm-specific fixed effect (FE), and u error term.

However, although there is a consensus in the literature about the inclusion of individual specific intercepts in investment equations, it is standard in the empirical literature to impose an homogeneous response of Tobin's q and cash flow by estimating conditional mean regression models. To investigate the potential different

Table 2.13: Duration of the Estimations

	MD-QR			FE-QR		
$n = 100$	user	system	elapsed	user	system	elapsed
$T = 10$	0.41	0.00	0.42	0.08	0.01	0.09
$T = 50$	0.45	0.01	0.46	0.44	0.03	0.48
$T = 100$	0.46	0.01	0.48	1.66	0.06	1.72
$T = 250$	0.54	0.03	0.57	5.64	0.16	5.81
$T = 500$	0.62	0.06	0.68	13.23	0.32	13.56
$T = 1000$	0.79	0.23	1.03	29.42	0.71	30.17
$T = 100$	user	system	elapsed	user	system	elapsed
$n = 10$	0.05	0.00	0.05	0.01	0.00	0.01
$n = 50$	0.25	0.01	0.25	0.15	0.02	0.17
$n = 100$	0.50	0.02	0.52	1.64	0.07	1.71
$n = 250$	0.99	0.07	1.06	39.76	0.40	40.21
$n = 500$	2.13	0.07	2.19	703.81	1.78	705.52
$n = 1000$	4.57	0.43	5.00	8082.06	8.45	8089.07

types of heterogeneity in investment models we use the QR framework developed in this chapter. QR panel data is used to analyze investment equations because it allows for individual fixed effects, and most importantly, it allows exploring a range of covariate effects. There are several compelling reasons to believe that the sensitivity of investment to cash flow varies across quantiles. Firms that have more volatile cash flow may not be as sensitive to cash flow in terms of investment. In particular, increased investment will increase future expenses. Firms with large variances in cash flow may respond in a more tempered way to positive changes in cash flow, believing that such changes can be mitigated by a negative shock in the future periods. Moreover, firms may also exhibit heterogeneity in their response to q_{it}.

We seek to estimate the following quantile regression version of the baseline equation described in (2.6). The conditional quantile functions are given by

$$Q_{IK_{it}}(\tau|\alpha_i, q_{it}, CFK_{it}) = \alpha_i(\tau) + \theta(\tau)q_{it} + \gamma(\tau)CFK_{it},$$

where the parameters of interest are $\theta(\tau)$ and $\gamma(\tau)$, which are allowed to depend on the quantile τ.

2.6.1 Data Description

The dataset is from COMPUSTAT and covers 1970 to 2010. We follow Almeida and Campello (2007) to collect the data. The sample consists of manufacturing firms with fixed capital of more than $ 5 million (with 1976 as the base year for the cpi), and the sample firms have growth of less than 100% in both assets and sales. Table 2.14 presents summary statistics for investment, cash flow, and q. For estimation and robustness check, we break the sample into cases were there are data available for the relevant firms between 35 and 40 years, more than 40 years, and more than 35 years. Each case in this breakdown allows us to estimate investment equations for each individual firm in the sample, and then average the results to compute the MD estimators. A comparison shows that these statistics are similar to those in Almeida and Campello (2007).

Table 2.14: Descriptive Statistics

	Variable	Obs.	Mean	Std. Dev.	Median	Min	Max
	Investment	5596	0.211	0.126	0.184	0.003	1.208
$35 \leq T < 40$	Cash Flow	5596	0.389	0.317	0.345	-2.509	5.042
	q	5596	1.130	0.879	0.910	0.323	16.026
	Investment	4320	0.208	0.111	0.188	0.004	1.141
$T \geq 40$	Cash Flow	4320	0.417	0.311	0.363	-2.351	4.022
	q	4320	1.147	1.002	0.931	0.304	18.342
	Investment	9916	0.210	0.120	0.186	0.003	1.208
$T \geq 35$	Cash Flow	9916	0.401	0.315	0.353	-2.509	5.042
	q	9916	1.137	0.934	0.918	0.304	18.342

2.6.2 Estimation Results

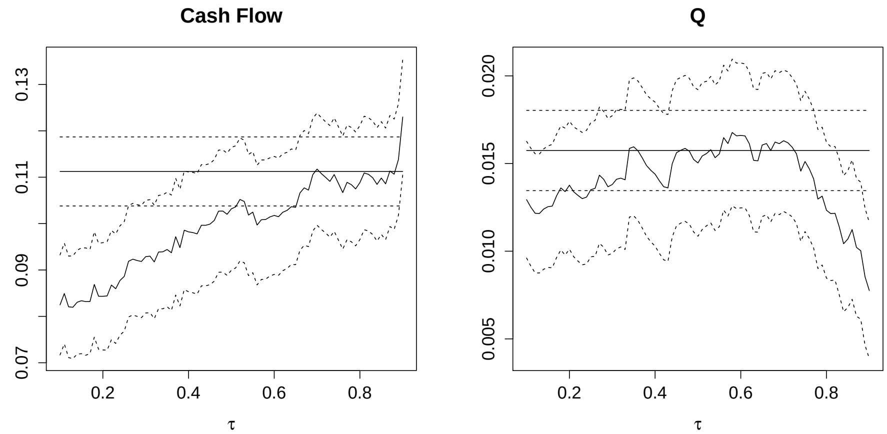
We estimate the above investment equation using FE quantile regression MD, and FE mean regression MD.[2] The QR technique explicitly allows for heterogeneity across the conditional quantile functions. The results for the estimated effects of Tobin's q and cash flow are presented in Figures 2.1a–2.1c, for different samples. The figures contain point estimates as well as 90% pointwise confidence bands for both QR and mean estimates.

The results regarding the sensitivity of investment to cash flow are presented in the left panels of Figures 2.1a–2.1c. The left panel of Figure 2.1a presents the results for the sample with $35 \leq T \leq 40$ and shows positive point estimates for both mean and quantile effects of cash flow on investment. The mean regression estimate is represented by the horizontal straight line, which shows a positive effect close to 0.11, and is statistically different from zero at usual levels of significance. Figure 2.1a also shows that the QR effects are positive and increasing along the quantile index τ. The coefficients are also statistically different from zero at usual levels of significance. This finding uncovers several important features. Firstly, this documents important heterogeneity on the response of investment spending to cash flow along the conditional quantile function. Firms in different quantiles of the conditional distribution of investment respond differently to marginal changes in the cash flow. Secondly, this heterogeneous increasing effect across quantiles also indicates that, for a fixed level of q, the variability of the investment spending across the conditional distribution increases as the level of cash flow increases. Intuitively, firms with larger cash flow are entitled to invest in a larger range in contrast to the firms with smaller cash flow. Thirdly, Figure 2.1a shows evidence that the investment spending is more sensitive to cash flow (large magnitude of the coefficients) for firms at high quantiles. The large coefficients for higher quantiles is an intuitive

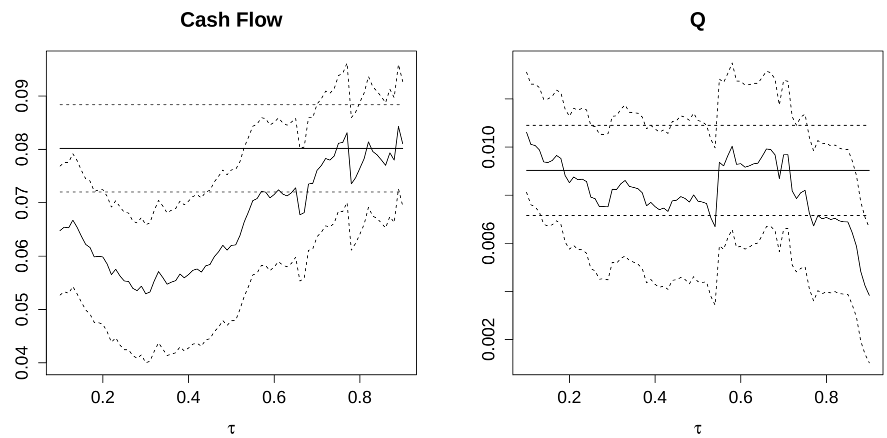
[2] Mean regression MD is defined as the weighted average of the OLS slope estimators with the corresponding inverse matrices as the weights.

Figure 2.1: MD Estimators of the Quantile and Mean Regression

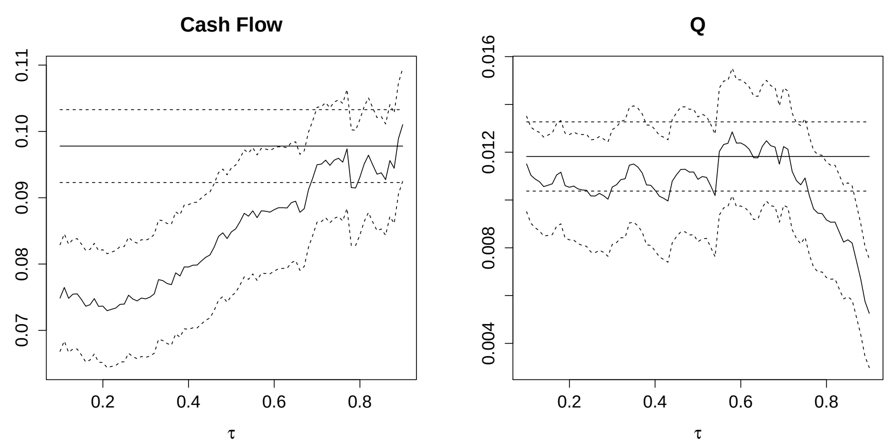
(a) Firms with $35 \leq T < 40$



(b) Firms with $T \geq 40$



(c) Firms with $T \geq 35$



The solid curve indicates the MDQR panel data estimates, the dotted curves are 90 % confidence bands for the panel data estimates, and the horizontal solid and dashed lines represent minimum distance mean regression estimates and the 90 % confidence bands.

result. The cash flow coefficient captures the potential sensitivity of investment to fluctuations in available internal finance–after investment opportunities and individual fixed effects are controlled for. Thus, the results show evidence that firms with high levels of investment spending are in fact more exposed to and dependent on fluctuations in internal finance. The left panels in Figures 2.1b and 2.1c collect the analogous results for samples with $T \geq 40$ and $T \geq 35$, respectively. The results are qualitatively similar to those in Figure 2.1a.

The results for both mean and quantile estimates of the sensitivity of investment to Tobin's q are presented in the right panels of Figures 2.1a–2.1c for different samples. According to the theory of investment equation, e.g., Fazzari et al. (1988), and previous empirical studies, e.g., Kaplan and Zingales (1997), on average, the investment demand has positive effect on the investment spending. This result is verified in our estimations. The right panels of Figures 2.1a–2.1c show the significant positive levels of the horizontal lines. The quantile estimates also show evidence that there is positive effect of the investment demand on the investment spending across the conditional quantile distribution of investment. Moreover, for all the three samples, there is evidence that estimated QR effects are relatively stable and do not vary much across quantiles. The QR point estimates are also relatively close to the mean effect. Thus, we find evidence that firms have similar responses to changes of investment demand across the conditional distribution of investment.

2.7 Summary

In this chapter, we study the asymptotic properties of the minimum distance estimator (MD-QR) for the fixed effects quantile regression panel model. We establish sufficient conditions for the consistency and asymptotic normality of the MD-QR estimator under sequential and joint asymptotics. In addition, we derive new asymptotical results for the standard quantile regression fixed effects (FE-QR) estimator under the sequential asymptotics. We introduce the use of sequential limits to study the asymptotic theory of quantile regression (QR) panel models with fixed effects. This is an important innovation because it facilitates the derivation of the limiting results and can be utilized in future works for other QR panel data models. One important insight from the analysis of the two different asymptotics is that the sequential limits substantially simplify the formal derivation of the asymptotic properties of both MD-QR and FE-QR. In addition, the MD-QR has the same limiting distribution under both sequential and joint limits. This result also holds for the FE-QR estimator.

The MD-QR has significant properties. Firstly, it is efficient in the class of minimum distance estimators. Secondly, it is computationally attractive. Monte Carlo results show that both the bias and the standard deviation are relatively small. Moreover, the great computing efficiency of the MD-QR estimator, especially when n is large, provides strong incentive for practitioners to adopt it as opposed to the FE-QR estimator.

There are many variants of the model that would extend the presented structure for the MD-QR that we leave for future research. These include analysis and extension of the methods for other models as censoring and duration are also a critical direction for future research. Applications to treatment effects models would be an interesting environment for further development of these methods.

Chapter 3

Testing Individual Slope Homogeneity in Quantile Regression Panel Data Models with an Application to Firm Capital Structure

3.1 Introduction

It is usual in applications of panel data models to impose a concomitant assumption of heterogeneous individual specific intercepts and homogeneous slope coefficients across individuals. The former condition has become standard in panel data models. However, the latter constraint might be seen as excessively strong and has become controversial as the availability of data increases. Heckman (2001) states "the most important discovery [from the widespread use of micro-data is] the evidence on the pervasiveness of heterogeneity and diversity in economic life." Motivated by the question on the benefits of pooling estimators *vis-à-vis* heterogeneous estimators, Baltagi et al. (2000) reinvestigate the advantages of pooling, and compare the performance of the homogeneous and heterogeneous estimators in an empirical study of cigarette demand. They conclude that pooled models outperform their heterogeneous counterparts. On the other hand, another branch of the literature uses shrinkage methodology to investigate the same question, whether to pool the data (see, e.g., Maddala and Hu (1996), Maddala et al. (1997), and Maddala et al. (2001)). These models do not assume homogeneity of the slope coefficient, and thus allow for heterogeneity across individuals.[1] Hsiao and Sun (2000) argue that if the individu-

[1]Another related literature includes random coefficient models. Swamy and Tavlas (2007) and Hsiao and Pesaran (2008) are good surveys for these models. For a general discussion on the

als do not share homogeneous coefficients, fixed effects estimation may not estimate any parameters of interest; hence, in empirical work, it is important to use formal tests to evaluate the conjecture of homogeneous coefficients across individuals.

There are several tests available in the literature for the hypothesis of slope homogeneity across individuals for mean regression models. Pesaran et al. (1996) propose an application of the Hausman (1978) testing procedure where the FE estimator is compared with the mean group estimator. Phillips and Sul (2003) suggest a "Hausman-type" test for slope homogeneity in the context of stationary first-order autoregressive panel data models, where the cross-section, n, is fixed as the time-series, T, goes to infinity. Hsiao (2003) describes a variation of the Breusch and Pagan (1979) test for the slope homogeneity, which is valid when both n and T dimensions tend to infinity. More recently, Pesaran and Yamagata (2008) (PY hereafter) propose a dispersion type test based on Swamy (1970) type test. PY standardize the Swamy type test so that this dispersion test can be applied when both n and T are large.

Motivated by the fact that formal tests for homogeneity of the slopes across cross-sectional units in panel data models are an indispensable tool for practitioners and also by the recent strong influence of QR panel data methods, this chapter contributes to the literature by developing testing procedures for homogeneity of the slope coefficients across individuals for FE QR models and a fixed quantile. A panel data QR model with different coefficients across quantiles and individuals is a flexible method since it is able to capture these two different sources of heterogeneity. In addition, when the individuals have heterogeneous slope coefficients, FE QR estimation has the potential to be very misleading since it is attempting to combine parameters in a fashion that may render the estimator inconsistent for any population parameters. Thus, we propose two tests, a Swamy ($\hat{S}$) and a standardized Swamy ($\hat{\Delta}$) type tests, with the null hypothesis of slope homogeneity across

modeling of heterogeneity, see Browning and Carro (2007).

individuals for a given quantile of interest. We derive the limiting distributions of the tests under different asymptotic sample size conditions. In particular, we show that under regularity conditions and the null hypothesis of the homogeneous slope coefficients, $\hat{S}$ converges to a χ^2 distribution as $T \to \infty$ and n is fixed; and also $\hat{\Delta}$ converges to the standard normal distribution as both $(T, n) \to \infty$, sequentially and jointly. Given these results, the critical values for a given level of significance are tabled and widely available, and a prominent advantage of these proposed tests is that they are very easy to implement in applications. When the null hypothesis of homogeneous slope coefficients is rejected for some selection of τ's, there is evidence of heterogeneous covariate effects across individuals, and as a result using fixed effect quantile regression assuming homogeneous slope coefficients is inappropriate. In addition, when the null is rejected by our formal tests, one could consider estimation of a set of parameters for each cross-sectional unit, or perhaps a shrinkage estimator could be entertained.

We conduct Monte Carlo simulations to evaluate the performance of the tests in finite samples. The simulation results show evidence that the proposed tests present empirical size that is very close the nominal size, and has good power performance. The numerical experiments also confirm that the finite sample performance improves with the sample size.

Finally, we illustrate the implementation of the proposed tests with a "target leverage" model. Many empirical studies assume a homogeneous speed of capital adjustment across firms (see, i.e., Flannery and Rangan (2006)). To investigate the potential different speed of adjustment, we use the tests developed in this chapter. As stated in Galvao and Montes-Rojas (2010), "QR panel data is a suitable tool for analyzing the behavior of 'target leverage' models since it allows controlling for individual specific intercepts, and most importantly, it allows exploring a range of conditional quantile functions exposing a variety of forms of conditional heterogeneity." Thus, we contribute to the discussion by uncovering two important sources of heterogeneity. Firstly, we test the common assumption in the literature of homogeneity of speed of convergence across firms. More specifically we test whether, for a fixed τth quantile, there is homogeneity across firms in the coefficient describing the speed of convergence. Secondly, we estimate the model for several different quantiles and examine the heterogeneity across the distribution. The results show evidence that, using our tests for selected quantiles and all the firms in the sample, one is able to reject the null hypothesis of homogeneous slope coefficient across the firms at usual levels of significance. In addition, we consider different subsets of the firms to examine robustness of the results. When considering firms with similar features, in particular a balanced panel of 31 years, one is not able to reject the null of same speed of adjustment across firms for some quantiles. We also document large heterogeneity in the speed of convergence across different parts of the conditional quantile function.

The rest of the chapter is organized as follows. Section 3.2 describes and discusses the null hypothesis and the tests proposed. Sections 3.3 and 3.4 study the asymptotic properties of the test statistics for static and dynamic models, respectively. The Monte Carlo simulation results are reported in Section 3.5. In Section 3.6 we illustrate the new approach with an application to firm capital structure, and Section 3.7 concludes the chapter.

3.2 The Null Hypothesis and the Proposed Tests

We consider a linear quantile regression panel model with n individuals and T time periods for each individual as

$$y_{it} = \boldsymbol{X}_{it}^{\top}\boldsymbol{\theta}_{i0} + \epsilon_{it}(\tau) := \alpha_{i0} + \boldsymbol{x}_{it}^{\top}\boldsymbol{\beta}_{i0} + \epsilon_{it}(\tau) \;\; t = 1, \ldots, T; \; i = 1, \cdots, n, \tag{3.1}$$

where y_{it} is the response variable for the ith individual at time t, $\boldsymbol{X}_{it}$ is its corresponding covariates with the first element being one, $\epsilon_{it}(\tau)$ is the error term whose τth quantile is zero conditional on $\boldsymbol{X}_{it}$, and $\boldsymbol{\theta}_{i0} := (\alpha_{i0}, \boldsymbol{\beta}_{i0}^\top)^\top$ is a $k+1$ vector of coefficients. In general, the coefficients depend on the quantile index τ. Since τ is fixed throughout the chapter, we will suppress this dependence for notational simplicity. The analogous version to equation (3.1) for the conditional quantile function of the response variable y_{it} can be represented as

$$Q_{y_{it}}(\tau|\boldsymbol{x}_{it}) = \boldsymbol{X}_{it}^\top \boldsymbol{\theta}_{i0} = \alpha_{i0} + \boldsymbol{x}_{it}^\top \boldsymbol{\beta}_{i0}.$$

For any given $\tau \in (0,1)$, we wish to test the following hypothesis of slope homogeneity across individuals

$$H_0 : \boldsymbol{\beta}_{i0} = \boldsymbol{\beta}_0$$

for some fixed vector $\boldsymbol{\beta}_0$ for all i, against the alternatives

$$H_1 : \boldsymbol{\beta}_{i0} \neq \boldsymbol{\beta}_{j0} \quad \exists \quad i, j.$$

To implement the tests, the strategy is to estimate the quantile regression coefficients using the time series for each individual, and then compare them with $\boldsymbol{\beta}_0$. Under the null, the estimates for all individuals should be close to $\boldsymbol{\beta}_0$. Therefore, a large value of the differences of these estimates and $\boldsymbol{\beta}_0$ indicates that the null should be rejected. However, in general, we do not observe the true coefficients $\boldsymbol{\beta}_0$, and thus we replace it with a weighted average of the estimates, $\hat{\boldsymbol{\beta}}_i$, from each individual. Under the null, all the $\hat{\boldsymbol{\beta}}_i$ should be close to each other, and to any weighted average of those estimates.

More specifically, put $\psi_\tau(u) := \tau - 1\{u \leq 0\}$. Denote the *slope* regression quantiles for each individual i by $\hat{\boldsymbol{\beta}}_i := \Xi\hat{\boldsymbol{\theta}}_i$, with

$$\hat{\boldsymbol{\theta}}_i := \underset{\boldsymbol{\theta}\in\mathbb{R}^{k+1}}{\arg\min} \frac{1}{T}\sum_{t=1}^{T}\rho_\tau(y_{it} - \boldsymbol{X}_{it}^\top\boldsymbol{\theta}),$$

$\Xi := [0_{k\times 1}|I_{k\times k}]$, where $\rho_\tau(u) := u\psi_\tau(u)$ as in Koenker and Bassett (1978). Define the following minimum distance (MD) estimator which is a weighted average of the slope regression quantiles

$$\hat{\boldsymbol{\beta}}_{MD} = \left(\sum_{i=1}^{n}\widehat{V}_i^{-1}\right)^{-1}\sum_{i=1}^{n}\widehat{V}_i^{-1}\hat{\boldsymbol{\beta}}_i, \tag{3.2}$$

with $\widehat{V}_i := \Xi\widehat{\widetilde{V}}_i\Xi^\top$, where $\widehat{\widetilde{V}}_i$ is a consistent estimator of the asymptotic variance-covariance matrix of the regression quantiles $\widetilde{V}_i := \tau(1-\tau)\widetilde{\Gamma}_i^{-1}\widetilde{\Omega}_i\widetilde{\Gamma}_i^{-1}$ with $\widetilde{\Gamma}_i := \mathrm{E}[f_i(0|\boldsymbol{X}_{it})\boldsymbol{X}_{it}\boldsymbol{X}_{it}^\top]$ and $\widetilde{\Omega}_i := \mathrm{E}[\boldsymbol{X}_{it}\boldsymbol{X}_{it}^\top]$ if the data are i.i.d. within each individual. The weighted mean of the $\hat{\boldsymbol{\beta}}_i$ defined above is a minimum distance estimator with weights being the inverse of the asymptotic variance-covariance matrices of the slope regression quantiles. For a thorough discussion of the minimum distance estimators, see Kodde et al. (1990), Newey and McFadden (1994), and Hsiao (2003). This weighted average is the benchmark for the comparison of the $\hat{\boldsymbol{\beta}}_i$'s in our tests. Therefore, we propose a quantile regression version of the Swamy type test as

$$\hat{S} := \sum_{i=1}^{n}\left(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_{MD}\right)^\top\left(\frac{\widehat{V}_i}{T}\right)^{-1}\left(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_{MD}\right),$$

and also a standardized Swamy test as

$$\hat{\Delta} = \sqrt{n}\frac{\frac{1}{n}\hat{S} - k}{\sqrt{2k}}.$$

As it will be clear later, when T is large and n is fixed, $\hat{S}$ is asymptotically $\chi^2_{(n-1)k}$ distributed. Therefore, $E\left[\frac{1}{n}\hat{S}\right]$ is approximately k and $Var\left[\frac{1}{n}\hat{S}\right]$ is approximately $\frac{2k}{n}$. In fact, under some mild conditions, the previous statement holds, and the standardized Swamy test statistic can be approximated by a standard normal random variable.

The intuition behind the tests is that under the null hypothesis $\hat{S}$ and $\hat{\Delta}$ should be close to zero. Under the null of homogeneity, all $\boldsymbol{\beta}_{i0}$ are the same. Thus, all estimates $\hat{\boldsymbol{\beta}}_i$ are close to each other and consequently to $\boldsymbol{\beta}_{i0}$, which is estimated by the weighted average $\hat{\boldsymbol{\beta}}_{MD}$. Therefore, both $\hat{S}$ and $\hat{\Delta}$ should be small. If there is evidence that either $\hat{S}$ or $\hat{\Delta}$ is large, we cast doubt that the null hypothesis of homogeneity of the slope coefficients holds.

Remark 3. *We use the minimum distance estimator $\hat{\beta}_{MD}$ as the benchmark to construct the test statistics in this chapter. The MD estimator simplifies the derivation of the limiting distributions of the test statistics as shown in the proofs. However, one could replace the MD estimator with the standard consistent FE-QR estimator, $\hat{\beta}_{FE}$, introduced by Koenker (2004). The main results we present here still hold, at the cost of lengthier proofs.*

3.3 Asymptotic Properties of the Tests

In this section, we investigate the asymptotic properties of the proposed tests. For the Swamy type test, we require n to be fixed and T to tend to infinity. For the standardized Swamy test, we let both T and n go to infinity. We analyze both sequential and simultaneous asymptotics in this case.

The use of large panel data asymptotics is common for testing slope homogeneity in the mean regression literature, as in PY. The Swamy type test was originally devised for panel data with small n and large T, while the standardized Swamy test was devised for panel data with both n and T large.

In this chapter, there are two main reasons why we require large panels. First, we need to consistently estimate $\boldsymbol{\beta}_{i0}$ for each individual, and the data employed in its estimation relies only on the time dimension. Second, the benchmark parameter, $\boldsymbol{\beta}_0$, is estimated by $\hat{\boldsymbol{\beta}}_{MD}$, which is an FE QR estimator. To achieve consistency of $\hat{\boldsymbol{\beta}}_{MD}$, it is also required that T tend to infinity. It is important to note that using large panels is common in panel data FE QR literature (see Kato et al. (2012)). In general, the FE QR estimator suffers from the incidental parameters problem (Neyman and Scott (1948)), it is therefore essential to allow T to increase to infinity to achieve for FE QR estimators .

We follow both the testing and the panel QR literatures and present results for the limiting distribution of the Swamy test ($\hat{S}$) as $T \to \infty$, and the limiting distributions of the standardized Swamy test ($\hat{\Delta}$) under two different asymptotics, T and n tend to infinity sequentially and simultaneously. The sequential asymptotics is defined as T diverging to infinity first, and then n. In the definition of the simultaneous asymptotics, T and n tend to infinity at the same time. We do not specify the exact relationship between n and T, although we maintain that T depends on n. For notational simplicity, we suppress this dependence. For a detailed discussion on sequential and simultaneous asymptotics for panel data, see Phillips and Moon (1999, 2000). In what follows, we adopt the following notation: $(T, n)_{seq} \to \infty$ means that first $T \to \infty$ and then $n \to \infty$, while $(T, n) \to \infty$ means T and n tend to infinity simultaneously. To obtain the desired results we use the assumptions of Chapter 2 and present the asymptotic distributions of the test statistics.

Theorem 5.

1. *Under Conditions A1–A6, we have* $\hat{S} \xrightarrow{d} \chi^2_{(n-1)k}$ *as* $T \to \infty$ *and* n *is fixed;*

2. *Under Conditions A1–A6, we have* $\hat{\Delta} \xrightarrow{d} N(0,1)$ *as* $(T, n)_{seq} \to \infty$*;*

3. *Under Conditions A1–A5, and A6', we have* $\hat{\Delta} \xrightarrow{d} N(0,1)$ *as* $(T, n) \to \infty$*, provided* $\frac{n^2 \log n}{T} |\log \delta_n|^2 \to 0$*, where* $\delta_n = \sqrt{\frac{\log n}{T}}$*.*

Proof. See Appendix C1. □

Theorem 1 provides the asymptotic distributions of $\hat{S}$ and $\hat{\Delta}$, and shows that it is very easy to find the critical values for the two tests. Since $\hat{S}$ is asymptotically $\chi^2_{(n-1)k}$ and $\hat{\Delta}$ is asymptotically normally distributed, the critical values for a given level of significance are tabled and widely available. Therefore, implementation of the proposed tests is simple. One computes the test statistics, $\hat{S}$ or $\hat{\Delta}$, sets the level of significance, and finds the critical values from the corresponding tables. The null hypothesis of homogeneity across individuals is rejected if the values of the test statistics are larger than the corresponding critical values.

3.4 Extensions to Dependent Data

In this section, we study the test statistics for independent data across individuals but within each individual, the data are relaxed to stationary and β-mixing. The mixing condition and the requirement on the joint conditional densities are in Chapter 2.

The following two theorems are asymptotic results for stationary β-mixing data, and are extensions of the theorems in Section 3.3.

Theorem 6. *1. Under Conditions A2–A6 and B1–B2, we have* $\hat{S} \xrightarrow{d} \chi^2_{(n-1)k}$ *as* $T \to \infty$ *and* n *is fixed;*

2. Under Conditions A2–A6 and B1–B2, we have $\hat{\Delta} \xrightarrow{d} N(0,1)$ *as* $(T,n)_{seq} \to \infty$*;*

3. Under Conditions A2–A5, A6', and B1–B2, we have $\hat{\Delta} \xrightarrow{d} N(0,1)$ *as* $(T,n) \to \infty$*, provided* $\frac{n^2 \log n}{T} |\log \delta_n|^2 \to 0$*, where* $\delta_n = \sqrt{\frac{\log n}{T}}$.

Proof. See Appendix C1. □

3.5 Finite Sample Simulations

In this section, we investigate the finite sample properties of the tests proposed above. We report empirical size and empirical power at nominal 5% level for various pairs of n and T. In particular, the simulations consider sample sizes $n \in \{10, 25, 50, 100\}$ and $T \in \{50, 100, 200\}$ and quantiles $\tau \in \{0.25, 0.50, 0.75\}$. The number of replications is 2,000.

3.5.1 Static Model

The static model under study is a location or a location-scale model as

$$y_{it} = \alpha_i + \beta_i x_{it} + (1 + \gamma x_{it}) u_{it}$$

for $t = 1, \cdots, T$ and $i = 1, \cdots, n$. For the location model $\gamma = 0$, and the location-scale model $\gamma = 0.5$. The innovations u_{it} are independent and identically distributed (i.i.d.) with distribution function F_u, and $x_{it} = 0.3\alpha_i + z_{it}$, where $z_{it} \overset{\text{i.i.d.}}{\sim} \chi_3^2$. For the simulations, we use two distributions for F_u, a standard normal ($N(0, 1)$) and a beta with shape parameters 2 and 6 ($B(2, 6)$).

We would like to test the null hypothesis $H_0 : \beta_i = \beta_j$ for all i, $j \in \{1, \cdots, n\}$ against $H_1 : \beta_i \neq \beta_j$ for any $i \neq j$. In the simulations, we set $\alpha_i = \frac{i-1}{n-1}$ (which means that α_i are uniformly distributed in the unit interval, and are fixed for different replications as in Pesaran and Yamagata (2008)). Under the null, we set $\beta_i = 1$ for all i; under the alternative $\beta_i = 0.25 + \frac{i-1}{n/2-1}$ for $i \leq \frac{n}{2}$ and 1 otherwise.

In the following tables, for any sample size combination we present results for the empirical size (left panels) and empirical power (right panels). We collect results for Swamy type test and $\hat{\Delta}$ test for each case. In addition, for both location and location-scale models, we report the results that use the true sparsity (for the three quartiles), the estimated sparsity with adjusted Hall-Sheather bandwidth rule (for the median), and the estimated sparsity with adjusted Bofinger bandwidth rule (for the median).

Testing with the Nonsandwich Form

We first investigate the location model ($\gamma = 0$). When data are generated by a location model, the sandwich form of the asymptotic covariance described previously in $\widetilde{V}_i$ simplifies to $\tau(1-\tau)E(\boldsymbol{X}_{i1}\boldsymbol{X}_{i1}^\top)^{-1}/f_U^2(x_p)$, where $\boldsymbol{X}_{i1} = (1, x_{i1})^\top$ and f_U is the density of F_U.

Table 3.1 displays the results for normal $N(0,1)$ and beta $B(2,6)$ innovations with the true sparsity functions. From Table 3.1, we see that for all the three quartiles, the empirical sizes are close to 5% and most of the empirical powers are close to 100% if not 100%.

Tables 3.2 presents the empirical size and power for the $N(0,1)$ and $B(2,6)$ innovations with the adjusted Hall-Sheather and Bofinger bandwidths when estimating the sparsity functions. We present results for models with two different bandwidths, namely, the Hall-Sheather, $h = 1.7h_{HS}$, and the Bofinger, $h = 1.8h_B$. The results indicate that the tests have good sizes, with empirical sizes close to the nominal 5%, and large power independent of the bandwidth choice. In addition, the results show similar performance for both Hall-Sheather and Bofinger bandwidths for the different sample sizes and innovations. Thus, the results are robust to the distributions of innovations and sample sizes.

Testing with the Sandwich Form

Next, we investigate the performance of our tests under the location and location-scale models. In particular, we focus on the sandwich form variance which is robust to the location-scale model. Again, assuming that the true sparsity function is known, we first study the empirical size and power. The results are listed in Table

Table 3.1: Empirical Size and Power for Static Location Model with $N(0,1)$ and $B(2,6)$ Innovations across Quartiles. Estimation with Nonsandwich Form and True Sparsity Function.

	τ		Empirical Size								Empirical Power							
			Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$
		T,n	10		25		50		100		10		25		50		100	
$N(0,1)$	0.25	50	0.058	0.050	0.049	0.047	0.052	0.051	0.049	0.049	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		100	0.052	0.047	0.048	0.045	0.055	0.054	0.057	0.056	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.047	0.041	0.054	0.051	0.051	0.050	0.052	0.052	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.5	50	0.053	0.044	0.043	0.041	0.050	0.050	0.052	0.052	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		100	0.051	0.044	0.050	0.049	0.052	0.051	0.054	0.054	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.046	0.043	0.052	0.050	0.055	0.054	0.050	0.050	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.75	50	0.058	0.053	0.046	0.043	0.048	0.046	0.039	0.039	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		100	0.049	0.042	0.059	0.058	0.054	0.053	0.057	0.056	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.057	0.049	0.053	0.051	0.052	0.051	0.053	0.053	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$B(2,6)$	0.25	50	0.050	0.043	0.034	0.034	0.054	0.053	0.044	0.044	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		100	0.056	0.049	0.056	0.055	0.046	0.046	0.049	0.048	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.042	0.038	0.048	0.044	0.042	0.042	0.046	0.046	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.5	50	0.042	0.038	0.034	0.032	0.034	0.033	0.027	0.027	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		100	0.049	0.045	0.041	0.039	0.038	0.038	0.040	0.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.043	0.039	0.042	0.039	0.040	0.040	0.034	0.034	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.75	50	0.030	0.025	0.024	0.023	0.021	0.020	0.017	0.017	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		100	0.040	0.031	0.038	0.036	0.029	0.029	0.036	0.035	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.048	0.042	0.048	0.047	0.041	0.039	0.041	0.041	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3.2: Empirical Size and Power for Static Location Model with $N(0,1)$ and $B(2,6)$ Innovations for Median Regressions. Estimation with Nonsandwich Form and Estimated Sparsity Function.

h			Empirical Size								Empirical Power							
			Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$
		T,n	10		25		50		100		10		25		50		100	
$1.7h_{HS}$	$N(0,1)$	50	0.056	0.050	0.056	0.053	0.060	0.059	0.060	0.060	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		100	0.048	0.044	0.062	0.059	0.057	0.056	0.048	0.048	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.048	0.040	0.051	0.049	0.047	0.045	0.052	0.052	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$B(2,6)$	50	0.071	0.066	0.097	0.093	0.112	0.111	0.147	0.147	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		100	0.064	0.058	0.090	0.087	0.094	0.094	0.114	0.113	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.065	0.057	0.077	0.073	0.073	0.072	0.087	0.087	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$1.8h_B$	$N(0,1)$	50	0.054	0.050	0.070	0.067	0.074	0.072	0.085	0.085	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		100	0.063	0.057	0.071	0.066	0.068	0.068	0.072	0.071	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.059	0.051	0.054	0.052	0.061	0.058	0.056	0.056	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$B(2,6)$	50	0.062	0.060	0.095	0.092	0.131	0.128	0.161	0.160	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		100	0.078	0.069	0.087	0.084	0.101	0.100	0.115	0.114	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.072	0.063	0.068	0.064	0.084	0.083	0.093	0.093	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

3.3.

Table 3.3 shows good empirical size and high power for the location and location-scale models for $N(0,1)$ and $B(2,6)$ innovations and for the three quartiles. The results for the sandwich form presented in Table 3.3 are analogous to those in Table 3.1. Thus, once again one can observe that most of the empirical sizes are close to 5% and most of the empirical powers are close to 100%, if not 100%. In addition, one can observe that the results for the location model ($\gamma = 0$) in Table 3.3 are similar to those in Table 3.1. This shows evidence that estimation of the sandwich variance in location models does not affect the size and power of the proposed tests.

Now, we report the simulation results with the sparsity being estimated. As in the previous case we present results for two different choices of bandwidth. For this exercise, for the location model, we set the bandwidths as $h = 0.5h_{HS}$ and $h = 0.4h_B$, and for the location-scale model, the bandwidths are given by $h = 0.8h_{HS}$ and $h = 0.6h_B$. The results are summarized in Table 3.4. The results suggest that, in general, the empirical sizes are close to the nominal 5%. However, one can observe that for small samples and HS bandwidth there is a slight size distortion in both Swamy and standardized Swamy tests. But the size distortion disappears as the sample size increases in both cases. The empirical power is again large for both tests regardless of the sample size and distributions of innovations. Finally, as in the previous case, the results given in Table 3.4 for the estimated sandwich variance are comparable to those in Table 3.2. Therefore, there is evidence that the estimation of the sandwich variance does not affect the properties of the tests.

3.5.2 Dynamic Model

In this subsection, we investigate the finite sample properties of the tests for a dynamic model. We report empirical size and power at nominal 5% level for various

Table 3.3: Empirical Size and Power for Static Location and Location-Scale Models with $N(0,1)$ and $B(2,6)$ Innovations across Quartiles. Estimation with Sandwich Form and True Sparsity Function.

γ		τ		Empirical Size								Empirical Power							
				Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$
			T, n	10		25		50		100		10		25		50		100	
0	$N(0,1)$	0.25	50	0.059	0.049	0.047	0.045	0.045	0.044	0.039	0.039	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.049	0.043	0.056	0.054	0.053	0.052	0.060	0.060	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.054	0.047	0.062	0.058	0.052	0.051	0.050	0.050	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.5	50	0.041	0.035	0.056	0.053	0.052	0.050	0.051	0.051	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.048	0.041	0.045	0.044	0.054	0.054	0.051	0.050	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.052	0.042	0.047	0.044	0.048	0.047	0.056	0.056	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.75	50	0.056	0.049	0.046	0.044	0.048	0.048	0.037	0.037	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.065	0.056	0.051	0.048	0.053	0.052	0.052	0.052	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.057	0.049	0.054	0.050	0.045	0.045	0.041	0.041	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$B(2,6)$	0.25	50	0.056	0.049	0.050	0.048	0.044	0.042	0.039	0.039	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.049	0.043	0.056	0.053	0.051	0.050	0.062	0.062	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.045	0.039	0.050	0.048	0.047	0.046	0.048	0.048	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.5	50	0.053	0.042	0.039	0.035	0.033	0.032	0.026	0.026	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.051	0.043	0.035	0.034	0.043	0.043	0.026	0.026	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.043	0.039	0.041	0.039	0.046	0.045	0.034	0.033	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.75	50	0.036	0.033	0.023	0.025	0.022	0.022	0.014	0.014	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.053	0.046	0.038	0.036	0.033	0.033	0.034	0.034	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.038	0.033	0.044	0.040	0.036	0.035	0.036	0.036	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

γ		τ		Empirical Size								Empirical Power							
				Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$
0.5	$N(0,1)$	0.25	50	0.055	0.047	0.059	0.058	0.054	0.052	0.057	0.057	0.518	0.498	0.699	0.687	0.886	0.884	0.986	0.986
			100	0.051	0.045	0.057	0.056	0.061	0.060	0.065	0.064	0.884	0.873	0.985	0.982	1.000	1.000	1.000	1.000
			200	0.055	0.048	0.052	0.049	0.047	0.046	0.053	0.053	0.997	0.997	1.000	1.000	1.000	1.000	1.000	1.000
		0.5	50	0.057	0.050	0.065	0.061	0.057	0.056	0.070	0.069	0.605	0.578	0.801	0.794	0.948	0.948	0.999	0.999
			100	0.048	0.044	0.050	0.047	0.056	0.055	0.074	0.073	0.931	0.924	0.990	0.990	1.000	1.000	1.000	1.000
			200	0.053	0.043	0.059	0.056	0.052	0.051	0.050	0.050	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.75	50	0.046	0.041	0.051	0.049	0.060	0.059	0.062	0.062	0.530	0.510	0.694	0.687	0.897	0.897	0.987	0.987
			100	0.056	0.049	0.053	0.051	0.063	0.062	0.064	0.064	0.857	0.847	0.978	0.977	0.999	0.999	1.000	1.000
			200	0.051	0.042	0.049	0.048	0.054	0.053	0.057	0.057	0.998	0.998	1.000	1.000	1.000	1.000	1.000	1.000
	$B(2,6)$	0.25	50	0.053	0.043	0.047	0.046	0.053	0.053	0.049	0.049	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.050	0.044	0.054	0.051	0.051	0.050	0.045	0.045	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.055	0.049	0.046	0.045	0.047	0.046	0.039	0.039	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.5	50	0.050	0.042	0.051	0.049	0.047	0.046	0.040	0.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.053	0.045	0.048	0.046	0.045	0.043	0.047	0.047	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.056	0.049	0.052	0.050	0.042	0.041	0.050	0.050	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		0.75	50	0.041	0.034	0.036	0.035	0.035	0.035	0.028	0.028	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.057	0.053	0.043	0.041	0.040	0.040	0.042	0.042	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.047	0.038	0.049	0.047	0.039	0.039	0.045	0.044	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3.4: Empirical Size and Power for Static Location and Location-Scale Models with $N(0,1)$ and $B(2,6)$ Innovations for Median Regressions. Estimation with Sandwich Form and Estimated Sparsity Function.

γ	h			Empirical Size								Empirical Power							
				Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$
			T,n	10		25		50		100		10		25		50		100	
0	$0.5h_{HS}$	$N(0,1)$	50	0.036	0.031	0.023	0.022	0.025	0.024	0.011	0.011	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.045	0.038	0.036	0.035	0.028	0.028	0.032	0.032	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.054	0.047	0.042	0.038	0.051	0.050	0.047	0.047	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$B(2,6)$	50	0.043	0.035	0.035	0.034	0.030	0.029	0.030	0.030	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.064	0.058	0.066	0.064	0.062	0.061	0.077	0.076	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.067	0.061	0.070	0.069	0.091	0.090	0.096	0.095	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$0.4h_B$	$N(0,1)$	50	0.053	0.049	0.064	0.061	0.063	0.061	0.063	0.063	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.044	0.039	0.043	0.042	0.034	0.033	0.029	0.029	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.044	0.041	0.036	0.032	0.028	0.027	0.024	0.024	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		$B(2,6)$	50	0.068	0.062	0.088	0.085	0.087	0.087	0.100	0.099	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.064	0.056	0.056	0.053	0.063	0.061	0.067	0.067	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.049	0.043	0.051	0.050	0.060	0.060	0.062	0.062	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.5	$0.8h_{HS}$	$N(0,1)$	50	0.046	0.041	0.044	0.043	0.044	0.044	0.038	0.038	0.409	0.388	0.551	0.546	0.776	0.775	0.933	0.933
			100	0.042	0.036	0.041	0.040	0.039	0.039	0.024	0.024	0.813	0.798	0.944	0.943	0.997	0.997	1.000	1.000
			200	0.046	0.043	0.043	0.043	0.042	0.041	0.017	0.017	0.997	0.996	1.000	1.000	1.000	1.000	1.000	1.000
		$B(2,6)$	50	0.058	0.054	0.070	0.069	0.064	0.064	0.073	0.073	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.055	0.052	0.066	0.063	0.069	0.069	0.064	0.064	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.047	0.044	0.061	0.059	0.070	0.070	0.065	0.065	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$0.6h_B$	$N(0,1)$	50	0.106	0.095	0.132	0.130	0.173	0.172	0.225	0.225	0.543	0.526	0.746	0.743	0.920	0.916	0.995	0.995
			100	0.054	0.046	0.066	0.065	0.065	0.063	0.059	0.059	0.838	0.823	0.968	0.965	0.999	0.999	1.000	1.000
			200	0.038	0.032	0.036	0.035	0.025	0.024	0.023	0.023	0.995	0.994	1.000	1.000	1.000	1.000	1.000	1.000
		$B(2,6)$	50	0.120	0.106	0.178	0.173	0.241	0.238	0.345	0.345	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			100	0.071	0.065	0.101	0.099	0.120	0.118	0.144	0.143	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			200	0.052	0.043	0.059	0.057	0.051	0.050	0.049	0.049	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

pairs of n and T. The model under study is a dynamic model as

$$y_{it} = \alpha_i + \beta_i y_{it-1} + u_{it}$$

for $t = 1, ..., T$ and $i = 1, ..., n$. Innovations u_{it} are i.i.d. with distribution function F_u. We set F_u as a standard normal $(N(0,1))$ and a beta with shape parameters 2 and 6 $(B(2,6))$. In generating y_{it} we also set $y_{i,-49} = 0$, drop the first 50 observations, and use the last T observations for estimation.

We would like to test the null hypothesis $H_0 : \beta_i = \beta_j$ for all i, j against $H_1 : \beta_i \neq \beta_j$ for any $i \neq j$. In the simulations, we set $\alpha_i = \frac{i-1}{n-1}$. Under the null, we set $\beta_i = 0.5$ for all i; under the alternative $\beta_i = 0.05 + \frac{0.8(i-1)}{n/2-1}$ for $i \leq \frac{n}{2}$ and 0.5 otherwise.

Tables 3.5 and 3.6 summarize the empirical size and power respectively. In Table 3.5, we use the true sparsity at the quantile index of interest. The empirical sizes are around 5% for all the quartiles and both the errors, and the empirical sizes are getting closer to 5% as T increases or when both n and T increase. The empirical powers for the normal errors are in general higher than 70%; and the empirical powers are higher than 50% in most cases for $B(2,6)$ errors.

Table 3.6 reports the empirical size and power using the estimated sparsity with adjusted Hall-Sheather and Bofinger bandwidths. In particular, we use bandwidths $h = 1.7h_{HS}$ and $h = 1.8h_B$, as in the static case. The results in Table 6 show evidence that the empirical size is, in general, better when using $1.7h_{HS}$. But both cases present reasonably good empirical sizes. The empirical powers are similar for both bandwidths, and they are large, mostly higher than 90%.

Table 3.5: Empirical Size and Power for Dynamic Location Model with $N(0,1)$ and $B(2,6)$ Innovations across Quartiles. Estimation with Nonsandwich Form and True Sparsity Function.

	τ		Empirical Size								Empirical Power							
			Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$
		T, n	10		25		50		100		10		25		50		100	
$N(0,1)$	0.25	50	0.048	0.041	0.043	0.042	0.034	0.034	0.030	0.030	0.732	0.714	0.869	0.863	0.974	0.974	0.998	0.998
		100	0.057	0.048	0.049	0.045	0.058	0.058	0.046	0.046	0.984	0.984	0.999	0.999	1.000	1.000	1.000	1.000
		200	0.054	0.048	0.051	0.048	0.048	0.047	0.054	0.054	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.5	50	0.043	0.038	0.041	0.037	0.037	0.036	0.028	0.028	0.819	0.804	0.930	0.926	0.994	0.992	1.000	1.000
		100	0.049	0.044	0.055	0.053	0.052	0.051	0.038	0.038	0.994	0.991	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.050	0.044	0.043	0.039	0.042	0.041	0.046	0.046	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.75	50	0.040	0.038	0.026	0.026	0.032	0.032	0.033	0.033	0.749	0.730	0.858	0.851	0.974	0.974	0.999	0.999
		100	0.048	0.042	0.056	0.052	0.054	0.054	0.042	0.042	0.984	0.982	0.999	0.999	1.000	1.000	1.000	1.000
		200	0.046	0.042	0.044	0.043	0.060	0.058	0.052	0.050	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$B(2,6)$	0.25	50	0.036	0.034	0.028	0.027	0.023	0.022	0.016	0.016	0.914	0.904	0.978	0.978	1.000	1.000	1.000	1.000
		100	0.046	0.044	0.044	0.042	0.040	0.040	0.034	0.034	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.056	0.052	0.037	0.036	0.042	0.041	0.039	0.039	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.5	50	0.028	0.022	0.019	0.019	0.020	0.020	0.013	0.013	0.770	0.748	0.862	0.858	0.979	0.978	1.000	1.000
		100	0.040	0.036	0.030	0.028	0.030	0.030	0.028	0.028	0.988	0.988	0.998	0.998	1.000	1.000	1.000	1.000
		200	0.042	0.036	0.044	0.042	0.041	0.040	0.034	0.034	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.75	50	0.027	0.024	0.024	0.024	0.012	0.012	0.012	0.012	0.494	0.464	0.582	0.578	0.778	0.778	0.946	0.946
		100	0.042	0.036	0.030	0.030	0.042	0.041	0.026	0.026	0.904	0.896	0.977	0.976	1.000	1.000	1.000	1.000
		200	0.044	0.038	0.040	0.038	0.038	0.038	0.041	0.041	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3.6: Empirical Size and Power for Dynamic Location Model with $N(0,1)$ and $B(2,6)$ Innovations for Median Regressions. Estimation with Nonsandwich Form and Estimated Sparsity Function.

h			Empirical Size								Empirical Power							
			Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$	Swamy	$\hat{\Delta}$
		T,n	10		25		50		100		10		25		50		100	
$1.7h_{HS}$	$N(0,1)$	50	0.046	0.040	0.047	0.046	0.044	0.044	0.042	0.042	0.756	0.741	0.894	0.892	0.984	0.984	1.000	1.000
		100	0.049	0.043	0.046	0.045	0.050	0.048	0.047	0.047	0.991	0.990	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.042	0.038	0.048	0.046	0.062	0.062	0.042	0.042	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$B(2,6)$	50	0.060	0.054	0.076	0.074	0.084	0.084	0.080	0.079	0.770	0.753	0.903	0.900	0.991	0.990	1.000	1.000
		100	0.063	0.057	0.077	0.072	0.081	0.079	0.093	0.093	0.990	0.989	0.998	0.998	1.000	1.000	1.000	1.000
		200	0.060	0.052	0.060	0.058	0.068	0.068	0.086	0.086	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$1.8h_B$	$N(0,1)$	50	0.074	0.064	0.076	0.073	0.088	0.086	0.100	0.100	0.804	0.794	0.920	0.918	0.990	0.990	1.000	1.000
		100	0.073	0.065	0.078	0.076	0.106	0.106	0.102	0.102	0.990	0.989	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.076	0.070	0.073	0.070	0.084	0.084	0.092	0.092	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$B(2,6)$	50	0.069	0.062	0.069	0.066	0.074	0.073	0.094	0.094	0.796	0.780	0.906	0.902	0.992	0.992	1.000	1.000
		100	0.074	0.068	0.076	0.074	0.083	0.082	0.105	0.104	0.993	0.992	1.000	1.000	1.000	1.000	1.000	1.000
		200	0.062	0.055	0.074	0.072	0.079	0.078	0.092	0.090	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

3.6 An Application: Target Capital Structure Adjustment

In this section, we illustrate the testing procedure with an application to a study of speed of capital structure adjustment. From the usual model in the literature, one can compute this speed by subtracting the estimated autoregressive coefficient in the dynamic model from one. With panel datasets, many studies assume that the speed of capital adjustments are the same across firms. In a recent paper, Oztekin and Flannery (2012) explore the potential heterogeneity in the coefficients of interest by studing the relationship between the capital structure adjustment speeds and the institutional determinants. To determine whether the institutional environment significantly affects firms' adjustment speeds, they estimate the partial adjustment model of leverage for 37 countries allowing heterogeneity across countries.

We build on this literature by estimating a panel data QR model to study the homogeneity of the speed of capital structure adjustment. The proposed model allows us to make two important contributions. First, we apply the tests proposed in this chapter to test the standard assumption in the literature that imposes homogeneity of speed of convergence across firms. More specifically, we test the null hypothesis that, for a fixed τ, the coefficient describing the speed of convergence across firms is homogeneous. Second, we estimate the model for several quantiles and examine heterogeneity across the conditional distribution. Using the model of Flannery and Rangan (2006), we have

$$MDR_{it+1} = \delta\alpha_i + X_{it}(\delta\beta) + MDR_{it}(1-\delta) + u_{it}, \tag{3.3}$$

where the coefficient δ is the adjustment speed toward the target, MDR_{it} is a firm's market debt ratio as a measure of leverage, and X_{it} is a set of covariate.

In this study, our data are from the Compustat Industrial Annual dataset between the year 1971 and 2005. Following existing literature, we drop financial firms (SIC codes 6000–6999), regulated utilities (SIC codes 4900–4999), and non-profit organizations (SIC codes greater than or equal to 9000). We exclude firm-years with missing or negative value for fixed assets and sales, with missing or less-than-ten-million 1983 dollar book value of total assets, and with growth rates of fixed assets, sales, and book value of total assets greater than 100%. We also omit firms that do not have enough variation in MDR: the ones with MDR mostly zero. The log of total assets is adjusted to the 1983 dollar with the consumer price index from the Bureau of Labor Statistics, as is in Galvao and Montes-Rojas (2010). The final sample includes 255 firms with 25–31 years of data.

To study a range of covariate effects in equation (3.3), Galvao and Montes-Rojas (2010) estimate the following model

$$Q_{MDR_{it}}(\tau|\alpha_i(\tau), MDR_{it-1}, X_{it-1}) = \alpha_i(\tau) + \rho(\tau)MDR_{it-1} + X_{it-1}^{\top}\beta(\tau) \qquad (3.4)$$

where $\rho(\tau) = (1-\delta(\tau))$, and X_{it} is a vector of regressors containing several covariates such that $X =$(*EBITTA, MB, DEPTA, LnTA, FATA, RDDum, RDTA*) with the following definitions.[2] "MDR: market debt ratio = book value of (short-term plus long-term) debt/market value of assets. $EBITTA$: earnings before interest and taxes, as a proportion of total assets; it is a measure of profitability. MB: market to book ratio of assets–book liabilities plus market value of equity divided by book value of total assets. $DEPTA$: depreciation as a proportion of total assets. $LnTA$: log of asset size, measured in 1983 dollars ×1,000,000, deflated by the consumer price index. $FATA$: fixed asset proportion–property, plant, and equipment/total assets. $RDDum$: dummy variable equal to one if red the firm did not report R&D expenses. $RDTA$: R&D expenses as a proportion of total assets."

It is standard in the literature to implicitly assume that in Model (3.4) the

[2] The computation with the Compustat reference is relegated to Appendix C2.

autoregressive coefficients are all the same across firms for a given τ. However, assuming that all the firms share the same $\rho(\tau)$ implies that all the firms have the same rate of adjustment of the market debt ratio. Intuitively, this assumption seems unrealistic and might not hold for vastly different firms. For example, small firms might have more difficulties in adjusting the market debt ratio in contrast to large firms. So before we pool the data of the firms together, we should formally test the assumption that all the firms share the same rate of adjustment.

Formally, the null hypothesis is

$$H_0 : \rho_i(\tau) = \rho(\tau)$$

for all i and some τ. We apply the suggested Swamy $\hat{S}$ and $\hat{\Delta}$ tests to the data, and thus supply formal tests for the assumption that all the firms share the same rate of adjustment.

We first report estimates of a quantile regression model for each firm separately. We estimate the model for several quantiles; in particular, $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$. The descriptive statistics and the histogram of $\hat{\rho}_i(\tau)$ are shown in Table 3.7 and Figure 3.1, respectively. This procedure could indicate whether there is evidence of heterogeneity across the firms. From the results, one can see that, for a given quantile, there is evidence of heterogeneity across individuals. This pattern is common for all selected quantiles. Moreover, the histograms show that the mass is relatively dispersed. Thus, the evidence provided in the table and figure casts doubts on the assumption that all firms share the same rate of adjustment.

The next step is to apply formal tests to the data. We report results for Swamy ($\hat{S}$) and standardized Swamy ($\hat{\Delta}$) tests for the selected quantiles. We also provide the minimum distance estimates defined in equation (3.2) ($\hat{\rho}_{MD}$) for selected quantiles. When estimating the variances, we use both HS and Bofinger bandwidths, with $h = 0.5h_{HS}$, $h = 0.8h_{HS}$, $h = 0.4h_B$, and $h = 0.6h_B$. We implement our tests using

Figure 3.1: The Histograms of $\hat{\rho}_i(\tau)$ of All the Firms

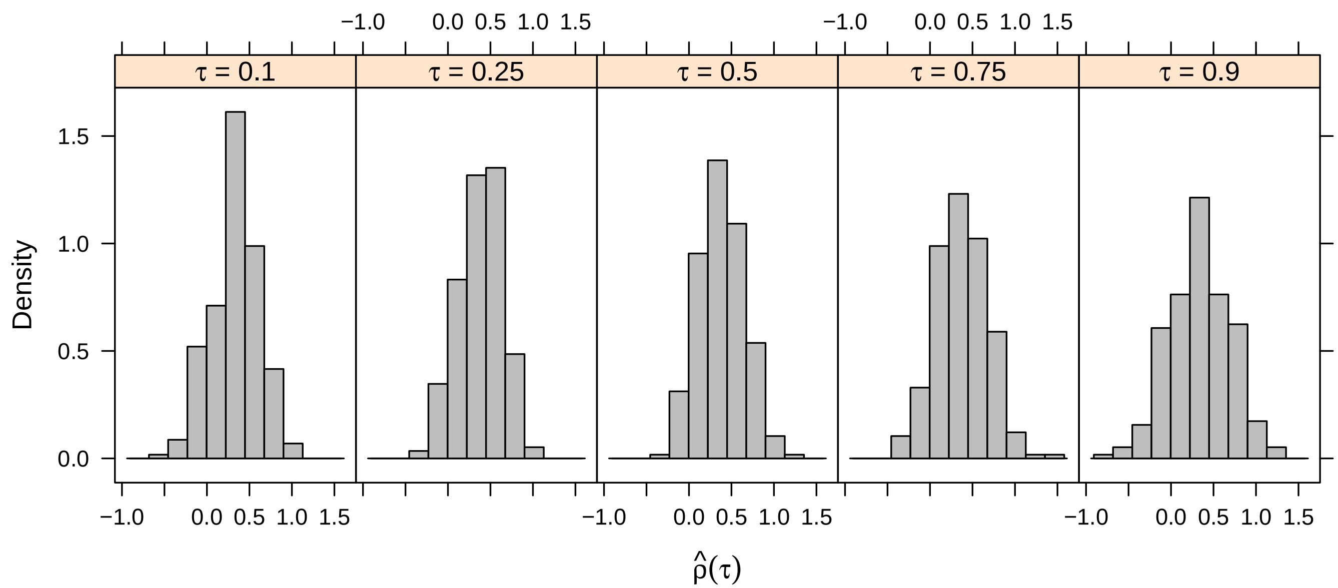


Table 3.7: The Summary Statistics of $\hat{\rho}_i(\tau)$ for All Firms

	min	1st quartile	Median	Mean	3rd quartile	max
$\tau = 0.1$	-0.5107	0.1621	0.3531	0.3413	0.5100	1.0570
$\tau = 0.25$	-0.3297	0.2035	0.3834	0.3721	0.5508	1.0850
$\tau = 0.5$	-0.2421	0.2027	0.3811	0.3914	0.5972	1.3320
$\tau = 0.75$	-0.3854	0.1674	0.3827	0.3765	0.5873	1.4870
$\tau = 0.9$	-0.8164	0.1022	0.3254	0.3374	0.6071	1.1690

only the sandwich form to estimate the variances, since the Monte Carlo simulations show evidence that the use of the sandwich variance does not affect the properties of the tests. Table 3.8 reports the resulting test statistics. One can see that the null hypothesis, namely, all firms share the same rate of adjustment, is strongly rejected for both $\hat{S}$ and $\hat{\Delta}$ tests for each τ. Thus, the formal tests reject the hypothesis that all the firms share the same rate of adjustment. From this, although firms can consider target debt ratio when making their decisions, the results from formal tests provide evidence that the heterogeneity is too large for the researchers to assume that firms share the same speed of adjustment. Hence, one should be cautious when analyzing the results provided in the literature for polling estimates. Another interesting feature described in Table 3.8 is the MD coefficients. The results show that the speed of adjustment also varies substantially across quantiles. The highest convergence speed is associated with high quantiles of debt ratio, about 70%. For the median the speed is about 50%, and for the first decile around 60%. Therefore, the exercise we conduct uncovers two new insights. First, we provide evidence of strong heterogeneity in the rate of adjustment of the market debt ratio. Second, there is also large heterogeneity and asymmetry in the speed of adjustment across the quantiles of the conditional of debt ratio.

Reviewing Table 3.7, we find some unusual phenomena. The minima of $\hat{\rho}_i(\tau)$ are negative for all quantiles. In addition, some of the estimates are greater than 1. This indicates that some of the firms do not converge to their target market debt ratios

Table 3.8: The Test Statistics for All the Firms

h		$0.5h_{HS}$	$0.8h_{HS}$	$0.4h_B$	$0.6h_B$
$\tau = 0.1$	$\hat{\rho}_{MD}$	0.378	0.381	0.370	0.379
	$\hat{S}$	611***	368***	789***	602***
	$\hat{\Delta}$	15.5***	4.85***	23.7***	15.4***
$\tau = 0.25$	$\hat{\rho}_{MD}$	0.430	0.417	0.442	0.426
	$\hat{S}$	853***	487***	1080***	737***
	$\hat{\Delta}$	26.2***	10.1***	36.5***	21.4***
$\tau = 0.5$	$\hat{\rho}_{MD}$	0.463	0.482	0.461	0.473
	$\hat{S}$	1020***	660***	1186***	852***
	$\hat{\Delta}$	33.5***	17.7***	41.2***	26.4***
$\tau = 0.75$	$\hat{\rho}_{MD}$	0.399	0.395	0.401	0.398
	$\hat{S}$	793***	471***	999***	687***
	$\hat{\Delta}$	23.5***	9.38***	32.9***	19.1***
$\tau = 0.9$	$\hat{\rho}_{MD}$	0.308	0.333	0.315	0.308
	$\hat{S}$	686***	460***	893***	681***
	$\hat{\Delta}$	18.8***	8.88***	28.3***	18.9***

$\hat{\rho}_{MD}$ denotes the minimum distance estimator in equation(3.2). $\hat{S}$ is the Swamy test, and $\hat{\Delta}$ the standardized Swamy test. $^{\star}p < 10\%$, $^{\star\star}p < 5\%$, $^{\star\star\star}p < 1\%$. The 10%, 5%, and 1% critical values for $\hat{S}$ are $\chi^2_{254,0.1} = 283.3$, $\chi^2_{254,0.05} = 292.2$, and $\chi^2_{254,0.01} = 309.4$, respectively, and for $\hat{\Delta}$ those values are $z_{0.1} = 1.28$, $z_{0.05} = 1.64$, and $z_{0.01} = 2.32$, respectively.

Table 3.9: The Summary Statistics of $\hat{\rho}(\tau)$ for All Firms with $0 < \hat{\rho}_i(\tau) < 0.97$

	min	1st quartile	Median	Mean	3rd quartile	max
$\tau = 0.1$	0.0409	0.2790	0.4026	0.4229	0.5602	0.8669
$\tau = 0.25$	0.0123	0.2904	0.4582	0.4427	0.5881	0.9362
$\tau = 0.5$	0.0092	0.2924	0.4441	0.4567	0.6280	0.9638
$\tau = 0.75$	0.0146	0.3080	0.4458	0.4628	0.6255	0.9564
$\tau = 0.9$	0.0043	0.2983	0.4399	0.4601	0.6367	0.9688

for various reasons. To avoid the unit root problem and lack of convergence, we drop the firms with $\hat{\rho}_i(\tau) > 0.97$ and $\hat{\rho}_i(\tau) < 0$ Thus, as a robustness check, to study rates of convergence to the target market debt ratios, we drop those unusual firms. As a result, we obtain a sample with 164 firms. Table 3.9 shows the descriptive statistics, and Figure 3.2 presents the histogram of the estimated $\hat{\rho}_i(\tau)$.

After dropping the unusual firms, the mean and median are slightly larger than those in the previous case. The histograms also show great dispersion of the estimated ρ for all quantiles. Formal tests for the null of homogeneity across individuals are reported in Table 3.10. The results for Bofinger bandwidth strongly reject the null hypothesis that the firms share the same rates of adjustment for all quantiles. The results regarding $0.5h_{HS}$ also reject the null for all quantiles. However, the results for the $0.8h_{HS}$ are mixed. The tests only reject the null hypothesis for $\tau = 0.25$ and $\tau = 0.5$. These results somewhat show that after we selected a more homogeneous sample by excluding unusual firms, it is possible to observe homogeneity for restricted parts of the distribution. Further evidence of an increase in homogeneity is given by the MD estimates. One can observe from Table 3.10 that the estimates of speeds of convergence do not vary as much as in the previous case across the quantiles.

Lastly, we attempt to test the null hypothesis for a set of firms that are in principle even more homogeneous. To this end, we choose firms that survive for an extended period of time. We work with a balanced panel containing only firms

Figure 3.2: The Histograms of $\hat{\rho}_i(\tau)$ of All the Firms with $0 < \hat{\rho}_i(\tau) < 0.97$

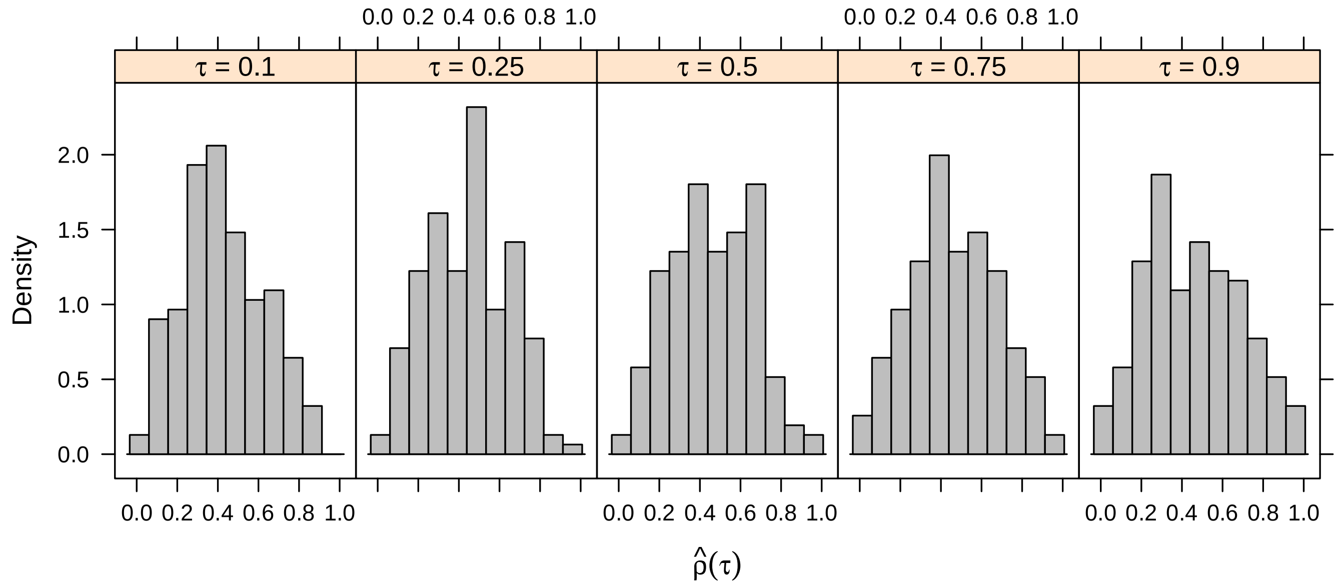


Table 3.10: The Test Statistics for All the Firms with $0 < \hat{\rho}_i(\tau) < 0.97$.

h		$0.5h_{HS}$	$0.8h_{HS}$	$0.4h_B$	$0.6h_B$
$\tau = 0.1$	$\hat{\rho}_{MD}$	0.416	0.423	0.410	0.417
	$\hat{S}$	259***	145	335***	254***
	$\hat{\Delta}$	5.23***	-1.04	9.46***	4.99***
$\tau = 0.25$	$\hat{\rho}_{MD}$	0.435	0.439	0.434	0.437
	$\hat{S}$	354***	201**	443***	304***
	$\hat{\Delta}$	10.5***	2.02**	15.4***	7.73***
$\tau = 0.5$	$\hat{\rho}_{MD}$	0.455	0.455	0.456	0.455
	$\hat{S}$	373***	214***	447***	290***
	$\hat{\Delta}$	11.5***	2.77***	15.6***	6.98***
$\tau = 0.75$	$\hat{\rho}_{MD}$	0.465	0.463	0.463	0.466
	$\hat{S}$	263***	164	321***	230***
	$\hat{\Delta}$	5.46***	-0.01	8.69***	3.64***
$\tau = 0.9$	$\hat{\rho}_{MD}$	0.442	0.450	0.443	0.443
	$\hat{S}$	221***	149	302***	218***
	$\hat{\Delta}$	3.15***	-0.82	7.60***	3.01***

$\hat{\rho}_{MD}$ denotes the minimum distance estimator in equation(3.2). $\hat{S}$ is the Swamy test, and $\hat{\Delta}$ the standardized Swamy test. $^{\star}p < 10\%$, $^{\star\star}p < 5\%$, $^{\star\star\star}p < 1\%$. The 10%, 5%, and 1% critical values for $\hat{S}$ are $\chi^2_{163,0.1} = 186.5$, $\chi^2_{163,0.05} = 193.8$, and $\chi^2_{163,0.01} = 207.9$, respectively, and for $\hat{\Delta}$ those values are $z_{0.1} = 1.28$, $z_{0.05} = 1.64$, and $z_{0.01} = 2.32$, respectively.

Table 3.11: The Summary Statistics of $\hat{\rho}_i(\tau)$ for All Firms with $0 < \hat{\rho}_i(\tau) < 0.97$ and $T = 31$.

	min	1st quartile	Median	Mean	3rd quartile	max
$\tau = 0.1$	0.0622	0.2544	0.3751	0.4051	0.5429	0.8284
$\tau = 0.25$	0.0748	0.2503	0.3946	0.4090	0.5118	0.9362
$\tau = 0.5$	0.0092	0.2408	0.4209	0.4241	0.6139	0.9638
$\tau = 0.75$	0.0913	0.2814	0.4034	0.4265	0.5756	0.8459
$\tau = 0.9$	0.0125	0.2569	0.3637	0.4001	0.5195	0.9455

that are in the sample for thirty-one years, $T = 31$. The rationale is that the firms with the same life-span (assuming that the data availability reflects the life-spans of the firms) might share similar features. So, we restrict the sample to firms with $0 < \hat{\rho}_i(\tau) < 0.97$ and $T = 31$. The new sample contains 73 firms. Table 3.11 and Figure 3.3 report the descriptive statistics and the histogram, respectively. The mean and median are similar to the last case, and the histograms show that the mass is not concentrated at any value.

Formal tests are conducted using the new sample, and the test statistics are reported in Table 3.12. The results are mixed. For the top of the conditional distribution of debt ratio ($\tau = 0.9$) all tests do not reject the null hypothesis of common slope, but $\hat{S}$ which is significant only at 10%. Thus, there is evidence that for high quantiles the firms behave similarly and adjust to the same target. For the remaining quantiles, all the tests agree in rejecting the null but when $h = 0.8H_{HS}$. When performing the test with HS bandwidth and $h = 0.8H_{HS}$ the tests do not reject the null of slope homogeneity with the exception of the median, which rejects at 5% level of significance. This result indicates that when using a more homogeneous sample, as expected, it is possible not to reject the hypothesis that firms have same rate of adjustment of the market debt ratio. Finally, one can see from Table 3.12 that the MD estimates are very homogeneous across quantiles. Thus, for more homogeneous firms, we have evidence of homogeneity in both individual and quantile dimensions.

Figure 3.3: The Histograms of $\hat{\rho}_i(\tau)$ of All the Firms with $0 < \hat{\rho}_i(\tau) < 0.97$ and $T = 31$

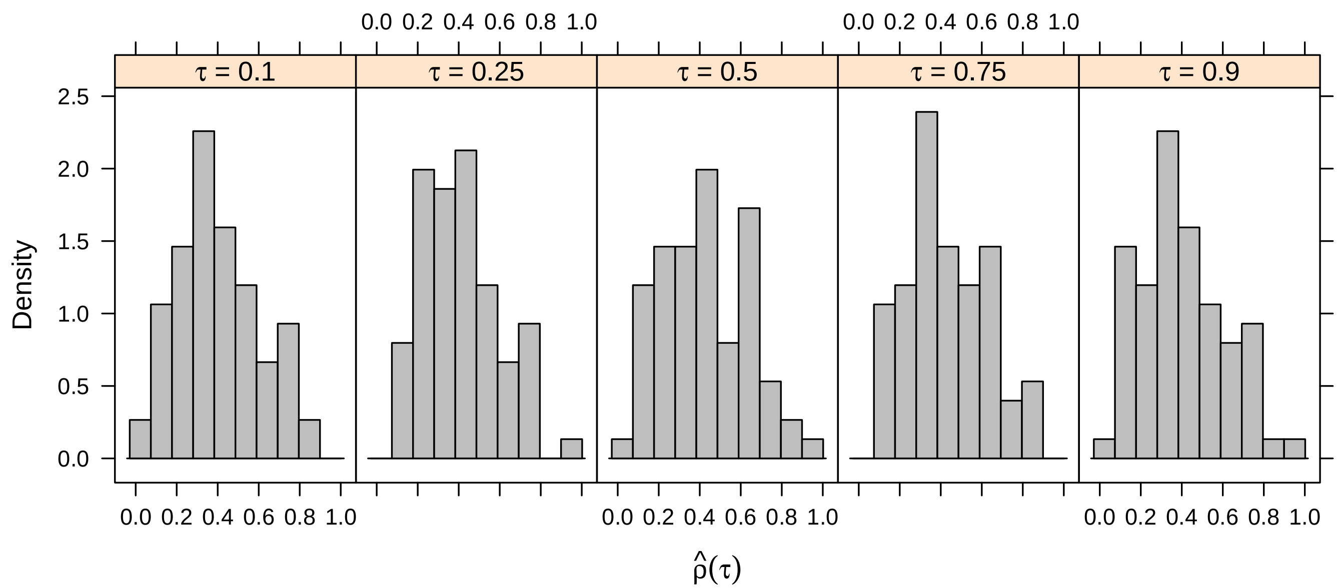


Table 3.12: The Test Statistics for All the Firms with $0 < \hat{\rho}_i(\tau) < 0.97$ and $T = 31$.

h		$0.5h_{HS}$	$0.8h_{HS}$	$0.4h_B$	$0.6h_B$
$\tau = 0.1$	$\hat{\rho}_{MD}$	0.402	0.412	0.389	0.403
	$\hat{S}$	151***	84	196***	147***
	$\hat{\Delta}$	6.46***	0.94	10.2***	6.16***
$\tau = 0.25$	$\hat{\rho}_{MD}$	0.414	0.414	0.417	0.414
	$\hat{S}$	140***	78	183***	117***
	$\hat{\Delta}$	5.58***	0.44	9.07***	3.65***
$\tau = 0.5$	$\hat{\rho}_{MD}$	0.418	0.420	0.418	0.418
	$\hat{S}$	173***	100**	203***	133***
	$\hat{\Delta}$	8.25***	2.20**	10.8***	5.00***
$\tau = 0.75$	$\hat{\rho}_{MD}$	0.426	0.422	0.427	0.424
	$\hat{S}$	112***	67	138***	96**
	$\hat{\Delta}$	3.19***	-0.48	5.41***	1.88**
$\tau = 0.9$	$\hat{\rho}_{MD}$	0.374	0.392	0.361	0.376
	$\hat{S}$	76	61	90*	75
	$\hat{\Delta}$	0.25	-1.00	1.44	0.19

$\hat{\rho}_{MD}$ denotes the minimum distance estimator in equation(3.2). $\hat{S}$ is the Swamy test, and $\hat{\Delta}$ the standardized Swamy test. $^{\star}p < 10\%$, $^{\star\star}p < 5\%$, $^{\star\star\star}p < 1\%$. The 10%, 5%, and 1% critical values for $\hat{S}$ are $\chi^2_{72,0.1} = 87.7$, $\chi^2_{72,0.05} = 92.8$, and $\chi^2_{72,0.01} = 102.8$, respectively, and for $\hat{\Delta}$ those values are $z_{0.1} = 1.28$, $z_{0.05} = 1.64$, and $z_{0.01} = 2.32$, respectively.

3.7 Conclusion

We have proposed Swamy and standardized Swamy (Δ) tests for the null hypothesis of slope homogeneity for the panel data fixed effects quantile regression model. These tests are important tools for practitioners, since they allow researchers to investigate the poolability of individual slopes in FE QR framework, for selected quantiles. We have derived the limiting distribution of the tests for large panels under sequential and joint asymptotics. The interpretation of the test results is simple. If the tests do not reject the null hypothesis of slope homogeneity across individuals, one could estimate the standard fixed effects panel data model, e.g., implementing FE QR estimation (Koenker (2004)). On the other hand, if our tests reject the null hypothesis, one could estimate the parameters of interest using data for the individuals; for instance, estimating individual quantile regressions using the time series data. We have conducted a finite sample study, and suggested potential bandwidths for estimating the nuisance parameters when using our proposed tests. Finally, we have illustrated the tests using a panel dataset of the firms' financial variables. We reject the null hypothesis of the equality of the rates of capital structure adjustment across firms.

BIBLIOGRAPHY

Abadie, A., J. Angrist, and G. Imbens, 2002: Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, **70(1)**, 91–117.

Abadie, A. and G. Imbens, 2006: Large sample properties of matching estimators for average treatment effects. *Econometrica*, **74**, 235–267.

Abrevaya, J., 2001: The effects of demographics and maternal behavior on the distribution of birth outcomes. *Empirical Economics*, **26(1)**, 247–257.

Abrevaya, J. and C. M. Dahl, 2008: The effects of birth inputs on birthweight: evidence from quantile estimation on panel data. *Journal of Business and Economic Statistics*, **26**, 379–397.

Ahn, S. C. and P. Schmidt, 1995: Efficient estimation of models for dynamic panel data. *Journal of Econometrics*, **68**, 5–27.

Almeida, H. and M. Campello, 2007: Financial constraints, asset tangibility and corporate investment. *Review of Financial Studies*, **20**, 1429–1460.

Almeida, H., M. Campello, and A. Galvao, 2010: Measurement errors in investment equations. *Review of Financial Studies*, **23**, 3279–3328.

Almond, D., K. Y. Chay, and D. S. Lee, 2005: The costs of low birth weight. *Quarterly Journal of Economics*, **120(3)**, 1031–1083.

Amemiya, T., 1974: Bivariate probit analysis: Minimum chi-square methods. *Journal of the American Statistical Association*, **69**, 940–944.

Amemiya, T., 1976: The maximum likelihood, the minimum chi-square and the nonlinear weighted least-squares estimator in the general qualitative response model. *Journal of the American Statistical Association*, **71**, 347–351.

Amemiya, T., 1978: The estimation of simultaneous equation generalized probit model. *Econometrica*, **46**, 1193–1205.

Angrist, J., V. Chernozhukov, and I. Fernandez-Val, 2006: Quantile regression under misspecification, with an application to the u.s. wage structure. *Econometrica*, **74**, 539–563.

Angrist, J. and J. S. Pischke, 2009: *Mostly Harmless Econometrics*. Princeton University Press, Princeton, New Jersey.

Baltagi, B., J. Griffin, and W. Xiong, 2000: To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. *The Review of Economics and Statistics*, **82**, 117–126.

Barnow, B. S., G. G. Cain, and A. S. Goldberger, 1980: Issues in the analysis of selectivity bias. *Evaluation Studies*, E. Stromsdorfer and G. Farkas, Eds., Sage, San Francisco, Vol. 5.

Behrman, R. E., A. S. Butler, and Committee on Understanding Premature Birth and Assuring Healthy Outcomes, 2007: *Preterm Birth: Causes, Consequences, and Prevention*. The National Academies Press, Washington, D.C.

Berkson, J., 1944: Application of the logistic function to bio-assay. *Journal of the American Statistical Association*, **39**, 357–365.

Bertrand, M. and S. Mullainathan, 2005: Bidding for oil and gas leases in the gulf of mexico: A test of the free cash flow model?, university of Chicago and MIT, Mimeo.

Bertrand, M. and A. Schoar, 2003: Managing with style: The effect of managers on firm policies. *Quarterly Journal of Economics*, **118**, 1169–1208.

Bickel, P., C. Klaassen, Y. Ritov, and J. Wellner, 1993: *Efficient and Addaptive Estimation for Semiparametric Models*. The Johns Hopkins University Press, Baltimore, Maryland.

Blundell, R. and J. Powell, 2003: Endogeneity in nonparametric and semiparametric regression models. *Advances in Economics and Econometrics, Vol. II*, M. Dewatripont, L. Hansen, and S. Turnovsky, Eds., Cambridge University Press.

Breusch, T. and A. Pagan, 1979: A simple test for heteroskedasticity and random coefficient variation. *Econometrica*, **47**, 1287–1294.

Browning, M. and J. Carro, 2007: Heterogeneity and microeconometrics modeling. *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress, Volume III*, R. Blundell, W. Newey, and T. Persson, Eds., Cambridge Press, chap. 3, 47–74.

Canay, I., 2011: A simple approach to quantile regression for panel data. *Econometrics Journal*, **14**, 368–386.

Cattaneo, M., 2010: Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics*, **155**, 138–154.

Cesur, R. and I. Kelly, 2010: From cradle to classroom: High birth weight and cognitive outcomes. *Health Economics*, **13(2)**, 2.

Chamberlain, G., 1982: Multivariate regression models for panel data. *Journal of Econometrics*, **18**, 5–46.

Chamberlain, G., 1984: Panel data. *Handbook of Econometrics, Vol. 2*, Z. Griliches and M. Intrilligator, Eds., North Holland, Elsevier, Amsterdam.

Chen, X., O. Linton, and I. Van Keilegom, 2003: Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, **71**, 1591–1608.

Chernozhukov, V. and I. Fernandez-Val, 2011: Inference for extremal conditional quantile models, with an application to market and birthweight risks. *Review of Economic Studies*, **78(2)**.

Chernozhukov, V. and C. Hansen, 2005: An iv model of quantile treatment effects. *Econometrica*, **73**, 245–261.

Chernozhukov, V. and C. Hansen, 2006: Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, **132**, 491–525.

Chernozhukov, V. and C. Hansen, 2008: Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, **142**, 379–398.

Chernozhukov, V. and H. Hong, 2002: Three-step censored quantile regression and extramarital affairs. *Journal of the American Statistical Association*, **97**, 872–882.

Dehejia, R. and S. Wahba, 1999: Causal effects in nonexperimental studies: Revaluating the evaluation of training programs. *Journal of the American Statistical Association*, **94**, 1053–1062.

Fazzari, S., R. G. Hubbard, and B. Petersen, 1988: Financing constraints and corporate investment. *Brooking Papers on Economic Activity*, **1**, 141–195.

Ferguson, T. S., 1958: A model of generating best asymptotically normal estimates with application to the estimation of bacterial densities. *Annals of Mathematical Statistics*, **29**, 1046–1062.

Fernandez-Val, I., 2005: Bias correction in panel data models with individual specific parameters, mimeo.

Firpo, S., 2007: Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, **75**, 259–276.

Flannery, M. J. and K. P. Rangan, 2006: Partial adjustment toward capital structures. *Journal of Financial Economics*, **79**, 469–506.

Florens, J. P., J. J. Heckman, C. Meghir, and E. Vytlacil, 2008: Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica*, **76**, 1191–1206.

Flores, C. A., 2007: Estimation of dose-response functions and optimal doses with a continuous treatment, mimeo.

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

Graham, B., J. Hahn, and J. Powell, 2009: The incidental parameter problem in a non-differentiable panel data model. *Economics Letters*, **105**, 181–182.

Hadlock, C., 1998: Ownership, liquidity, and investment. *RAND Journal of Economics*, **29**, 487–508.

Hahn, J., 1998: On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, **66**, 315–331.

Hahn, J. and G. Kuersteiner, 2011: Bias reduction for dynamic nonlinear panel model with fixed effects. *Econometric Theory*, dOI: 10.1017/S0266466611000028.

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

Hall, P., J. Racine, and Q. Li, 2004: Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, **99**, 1015–1026.

Hausman, J., 1978: Specification tests in econometrics. *Econometrica*, **46**, 1251–1271.

He, X. and Q.-M. Shao, 1996: A general bahadur representation of m-estimators and its applications to linear regressions with nonstochastic designs. *The Annals of Statistics*, **24**, 2608–2630.

He, X. and Q.-M. Shao, 2000: On parameters of increasing dimensions. *Journal of Multivariate Analysis*, **73**, 120–135.

Heckman, J., 2001: Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy*, **109**, 673–748.

Heckman, J., H. Ichimura, J. Smith, and P. Todd, 1998: Characterizing selection bias using experimental data. *Econometrica*, **66**, 1017–1098.

Heckman, J. and H. E. Vytlacil, 2005: Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica*, **73**, 669–738.

Hendricks, W. and R. Koenker, 1991: Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association*, **87**, 58–68.

Hirano, K., G. Imbens, and G. Ridder, 2003: Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, **71**, 1161–1189.

Hirano, K. and G. W. Imbens, 2004: The propensity score with continuous treatment. *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, A. Gelman and X.-L. Meng, Eds., Wiley.

Hoshi, T., A. Kashyap, and D. Scharfstein, 1991: Corporate structure, liquidity, and investment: Evidence from japanese industrial groups. *Quarterly Journal of Economics*, **106**, 33–60.

Hsiao, C., 2003: *Analysis of Panel Data*. Cambridge University Press, New York, New York.

Hsiao, C. and M. H. Pesaran, 2008: Random coefficient models. *Advanced Studies in Theoretical and Applied Econometrics, Vol. 46, The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, L. Matyas and P. Sevestre, Eds., Springer-Verlag, Berlin.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu, 2002: Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics*, **109**, 107–150.

Hsiao, C. and B. Sun, 2000: To pool or not to pool panel data. *Penal Data Econometrics: Future Directions: Papers in Honour of Professor Pietro Balestra*, J. Krishnakumar and E. Ronchetti, Eds., Elsevier Science, chap. 9, 181–198.

Imai, G. and D. van Dyk, 2004: Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, **99**, 854–866.

Imbens, G., 2000: The role of the propensity score in estimating doseresponse function. *Biometrika*, **87**, 706–710.

Imbens, G. and W. Newey, 2009: Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, **77**, 1481–1512.

Imbens, G., W. Newey, and W. Ridder, 2006: Mean-squared-error calculations for average treatment effects, mimeo.

Kaplan, S. and L. Zingales, 1997: Do financing constraints explain why investment is correlated with cash flow? *Quarterly Journal of Economics*, **112**, 169–215.

Kato, K., 2012: Asymptotic normality of powells kernel estimator. *Annals of the Institute of Statistical Mathematics*, **64**, 255–273.

Kato, K. and A. Galvao, 2010: Smoothed quantile regression for panel data, mimeo.

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

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

Kodde, D. A., F. C. Palm, and G. A. Pfann, 1990: Asymptotic least-squares estimation efficiency considerations and applications. *Journal of Applied Econometrics*, **5**, 229–243.

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

Koenker, R., 2005: *Quantile regression*. Cambridge University Press, New York, New York.

Koenker, R. and G. W. Bassett, 1978: Regression quantiles. *Econometrica*, **46**, 33–49.

Koenker, R. and K. Hallock, 2001: Quantile regression. *Journal of Economic Perspectives*, **15**, 143–156.

Koenker, R., J. A. F. Machado, C. L. Skeels, and A. H. Welsh, 1994: Momentary lapses: Moment expansions and the robustness of minimum distance estimation. *Econometric Theory*, **10**, 172–197.

Kosorok, M., 2008: *Introductioin to Empirical Processes and Semiparametric Inference*. Springer.

Lamont, O., 1997: Cash flow and investment: Evidence from internal capital markets. *The Journal of Finance*, **52**, 83–110.

Lechner, M., 2001: Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. *Econometric Evaluation of Labour Market Policies*, M. Lechner and F. Pfeiffer, Eds., Physical Springer, Heidelberg.

Lee, L. F., 1992: Amemiya's generalized least squares and tests of overidentification in simultaneous equation models with qualitative or limited dependent variables. *Econometric Reviews*, **11**, 319–328.

Lee, L. F., 2010: Pooling estimates with different rates of convergence: a minimum χ^2 approach with emphasis on a social interactions model. *Econometric Theory*, **26**, 260–299.

Lee, N., H. R. Moon, and M. Weidner, 2012: Analysis of interactive fixed effects dynamic linear panel regression with measurement error. *Economics Letters*, **117**, 239–242.

Lee, Y., 2012: Partial mean processes with generated regressors: Continuous treatment effects and nonseparable models, mimeo.

Lehmann, E. L. and G. Casella, 1998: *Theory of Point Estimation*. 2nd ed. Springer-Verlag, New York, New York.

Li, Q., J. Racine, and J. Wooldridge, 2009: Efficient estimation of average treatment effects with mixed categorical and continuous data. *Journal of Business & Economic Statistics*, **27**, 206–223.

Maddala, G. S. and W. Hu, 1996: The pooling problem. *Econometrics of Panel Data*, L. Matyas and P. Sevestre, Eds., Kluwer Academic Publishers, 307–322.

Maddala, G. S., H. Li, and V. K. Srivastava, 2001: A comparative study of different shrinkage estimators for panel data models. *Annals of Economics and Finance*, **2**, 1–30.

Maddala, G. S., R. P. Trost, H. Li, and F. Joutz, 1997: Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business and Economic Statistics*, **15**, 90–100.

Malinvaud, E., 1970: *Statistical Methods of Econometrics*. North Halland.

Malmendier, U. and G. Tate, 2005: Ceo overconfidence and corporate investment. *The Journal of Finance*, **60**, 2661–2700.

Moon, H. R. and F. Schorfheide, 2002: Minimum distance estimation of nonstationary time series models. *Econometric Theory*, **18**, 1385–1407.

Moon, H. R., M. Shum, and M. Weidner, 2012: Estimation of random coefficients logit demand models with interactive fixed effects, cemmap working paper CWP08/12.

Murphy, K. and R. Topel, 1985: Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics*, **3**, 88–97.

Nagaraj, N. and W. Fuller, 1991: Estimation of the parameters of linear time series models subject to nonlinear restrictions. *Annals of Statistics*, **19**, 1143–1154.

Newey, W. K., 1984: A method of moments interpretation of sequential estimators. *Economics Letters*, **14**, 201–206.

Newey, W. K., 1991: Uniform convergence in probability and stochastic equicontinuity. *Econometrica*, **59**, 1161–1167.

Newey, W. K. and D. L. McFadden, 1994: Large sample estimation and hypothesis testing. *Handbook of Econometrics, Vol. 4*, R. F. Engle and D. L. McFadden, Eds., North Holland, Elsevier, Amsterdam.

Newey, W. K., J. Powell, and F. Vella, 1999: Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, **67**, 565–603.

Neyman, J., 1949: Contribution to the theory of χ^2 test. *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman, Ed., University of California Press.

Neyman, J. and E. L. Scott, 1948: Consistent estimates based on partially consistent observations. *Econometrica*, **16**, 1–32.

Oztekin, O. and M. Flannery, 2012: Institutional determinant of capital structure adjustment speeds. *Journal of Financial Economics*, **103**, 88–112.

Pakes, A. and D. Pollard, 1989: Simulations and the asymptotics of optimization estimators. *Econometrica*, **57**, 1027–1057.

Pesaran, H., R. Smith, and K. S. Im, 1996: Dynamic linear models for heterogeneous panels. *The Econometrics of Panel Data: A Handbook of the Theory with Applications*, L. Matyas and P. Sevestre, Eds., Kluwer Academic Publishers, Dordrecht, 145–195.

Pesaran, M. H. and T. Yamagata, 2008: Testing slope homogeneity in large panels. *Journal of Econometrics*, **142**, 50–93.

Phillips, P. and H. Moon, 1999: Linear regression limit theory for nonstationary panel data. *Econometrica*, **67**, 1057–1111.

Phillips, P. and H. Moon, 2000: Nonstationary panel data analysis: an overview of some recent developments. *Econometric Reviews*, **19**, 263–286.

Phillips, P. and D. Sul, 2003: Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal*, **6**, 217–259.

Pinkse, J., 2000: Nonparametric two-step regression functions when regressors and errors are dependent. *Canadian Journal of Statistics*, **28**, 289–300.

Ponomareva, M., 2011: Quantile regression for panel data models with fixed effects and small t : Identification and estimation. mimeo.

Powell, D., 2010: Unconditional quantile regression for panel data with exogenous or endogenous regressors, mimeo.

Powell, J. L., 1991: Estimation of monotonic regression models under quantile regressions. *Nonparametric and semiparametric models in econometrics*, W. Barnett, J. Powell, and G. Tauchen, Eds., Cambridge University Press, Cambridge.

Rao, C. R., 1965: *Linear Statistical Inference and Its Applications*. Wiley, New York.

Rosen, A. M., 2012: Set identification via quantile restrictions in short panels. *Journal of Econometrics*, **166**, 127–137.

Rosenbaum, P. and D. Rubin, 1983: The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41–55.

Rothenberg, T. J., 1973: *Efficient Estimation with A Priori Information*. Yale University Press.

Rubin, D., 1977: Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, **2**, 1–28.

Serfling, R. J., 1980: *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Shin, H. and R. Stulz, 1998: Are internal capital markets efficient? *Quarterly Journal of Economics*, **113**, 531–552.

Stein, J., 2003: Agency information and corporate investment. *Handbook of the Economics of Finance*, M. Harris and R. Stulz, Eds., Elsevier/North-Holland, Amsterdam.

Swamy, P. and G. S. Tavlas, 2007: Random coefficient models. *A Companion to Theoretical Econometrics*, B. Baltagi, Ed., Malden: Blackwell Publishing Ltd, 410–428.

Swamy, P. A. V. B., 1970: Efficient inference in a random coefficient regression model. *Econometrica*, **38**, 311–323.

Taylor, W. F., 1953: Distance functions and regular best asymptotically normal estimates. *Annals of Mathematical Statistics*, **24**, 85–92.

van der Vaart, A., 1998: *Asymptotic Statistics*. Cambridge University Press, New York, New York.

van der Vaart, A. and J. A. Wellner, 1996: *Weak convergence and empirical processes*. Springer-Verlag Press, New York, New York.

Webb, R., 2011: High birth weight and socio-economic status, working paper.

Wei, Y. and R. J. Carroll, 2009: Quantile regression with measurement error. *Journal of the American Statistical Association*, **104**, 1129–1143.

Wooldridge, J., 2010: *Econometric Analysis of Cross Section and Panel Data, 2nd Edition*. The MIT Press, Cambridge, Maryland.

Appendix A

In Appendix A1, we provide asymptotic properties of a generic Z-estimator. More specifically, we describe the model, provide conditions, and state results. The proofs are collected in Appendix A2.

Appendix A1: Asymptotic Results for Generic Z-Estimators

Let Θ and $\mathcal{L}$ denote Banach spaces, and $\mathcal{H}$ a norm space. Let $\mathbb{Z}_n : \Theta \times \mathcal{H} \mapsto \mathcal{L}$, $\mathrm{Z} : \Theta \times \mathcal{H} \mapsto \mathcal{L}$ be random maps and a deterministic map, respectively. We suppress the dependence of $\mathbb{Z}$ on n for simplicity. The Z-estimator $\hat{\theta}$ is defined as the root of

$$\mathbb{Z}(\theta, \hat{h}) = 0,$$

where $\hat{h}$ is a first step estimator of a possibly infinite dimensional nuisance parameter. This general theory is an extension of Theorem 1 of Chen et al. (2003) in that the parameter of interest is a Banach valued quantity instead of a Euclidean vector, and of Theorem 3.3.1 of van der Vaart and Wellner (1996) to the model with a nuisance parameter.

Consistency

We first derive a general consistency result for a Z-estimator in a Banach space. To obtain the consistency of the generic Z-estimator, we impose the following conditions.

C.1 $||\mathbb{Z}(\hat{\theta}, \hat{h})||_{\mathcal{L}} = o_{p^*}(1)$.

C.2 $||\mathrm{Z}(\theta_n, h_0)||_{\mathcal{L}} \to 0$ implies $\theta_n \to \theta_0$ for any sequences $\theta_n \in \Theta$.

C.3 Uniformly in $\theta \in \Theta$, $\mathrm{Z}(\theta, h)$ is continuous at h_0.

C.4 $||\hat{h} - h_0||_{\mathcal{H}} = o_{p^*}(1)$.

C.5 For all sequences $\delta_n \downarrow 0$,

$$\sup_{\theta\in\Theta, ||h-h_0||_{\mathcal{H}}\leq\delta_n} \frac{||\mathbb{Z}(\theta, h) - \mathrm{Z}(\theta, h)||_{\mathcal{L}}}{1 + ||\mathbb{Z}(\theta, h)||_{\mathcal{L}} + ||\mathrm{Z}(\theta, h)||_{\mathcal{L}}} = o_{p^*}(1).$$

Condition **C.1** requires that $\hat{\theta}$ solves the estimating equation $||\mathbb{Z}(\theta, \hat{h})||_{\mathcal{L}} = 0$ only asymptotically. Condition **C.2** is an identification of the parameter. Condition **C.3** is a smooth assumption of Z in h only at h_0. Condition **C.4** requires that the nuisance parameter is consistently estimated. Condition **C.5** is a high level assumption and can be stated in more primitive conditions for specific cases. Further, condition **C.5** is implied by the following uniform convergence condition of $\mathbb{Z}$ to Z.

C5S For any sequences $\delta_n \downarrow 0$,

$$\sup_{\theta\in\Theta, ||h-h_0||_{\mathcal{H}}\leq\delta_n} ||\mathbb{Z}(\theta, h) - \mathrm{Z}(\theta, h)||_{\mathcal{L}} = o_{p^*}(1).$$

This set of conditions are similar to conditions of Theorem 1 of Chen et al. (2003).

The following lemma summarizes the consistency of a generic Z-estimator.

Lemma 1. *Suppose that* $\theta_0 \in \Theta$ *satisfies* $\mathrm{Z}(\theta_0, h_0) = 0$ *with* $h_0 \in \mathcal{H}$ *and that conditions* **C.1**–**C.5** *hold. Then* $||\hat{\theta} - \theta_0||_{\Theta} = o_{p^*}(1)$.

Proof. See Appendix A2. □

Weak Convergence

Now we provide a general result for the Z-estimator. For the proof of weak convergence of the Z-estimator, consistency is assumed without loss of generality. Therefore the parameter space is replaced by $\Theta_\delta \times \mathcal{H}_\delta$ where $\Theta_\delta := \{\theta \in \Theta : ||\theta - \theta_0||_\Theta < \delta\}$ as in Chen et al. (2003) and $\mathcal{H}_\delta := \{h \in \mathcal{H} : ||h - h_0||_\mathcal{H} < \delta\}$.

Because the parameter spaces are a Banach and a normed space, we need a notion of derivatives for maps from a Banach or a normed space to a Banach space. Let $\varTheta$ and $\mathbb{L}$ denote Banach spaces, and $\mathbb{H}$ a normed space. Fréchet differentiability of a map $\phi : \varTheta \mapsto \mathbb{L}$ at $\theta \in \varTheta$ means that there exists a continuous, linear map $\phi'_\theta : \varTheta \mapsto \mathbb{L}$ with

$$\frac{||\phi(\theta + h_n) - \phi(\theta) - \phi'_\theta(h_n)||}{||h_n||} \to 0$$

for all sequences $\{h_n\} \subset \varTheta$ with $||h_n|| \to 0$ and $\theta + h_n \in \varTheta$ for all $n \geq 1$; see p. 26 of Kosorok (2008). Pathwise derivative of a map $\varphi : \mathbb{H} \mapsto \mathbb{L}$ at $h \in \mathbb{H}$ in the direction $[\bar{h} - h]$ is

$$\varphi'_h[\bar{h} - h] = \lim_{\varrho \to 0} \frac{\varphi(h + \varrho(\bar{h} - h)) - \varphi(h)}{\varrho}$$

with $\{h + \varrho(\bar{h} - h) : \varrho \in [0, 1]\} \subset \mathbb{H}$, provided that the limit exists; see Chen et al. (2003). To obtain the weak limit, we impose the following sufficient conditions.

G.1 $||\mathbb{Z}(\hat{\theta}, \hat{h})||_\mathcal{L} = o_{p^*}(n^{-1/2})$.

G.2 The map $\theta \mapsto \mathrm{Z}(\theta, h_0)$ is Fréchet differentiable at θ_0 with a continuously invertible derivative $\mathrm{Z}_1(\theta_0, h_0)$.

G.3 For all $\theta \in \Theta_\delta$ the pathwise derivative $\mathrm{Z}_2(\theta, h_0)[h - h_0]$ of $\mathrm{Z}(\theta, h_0)$ exists in all directions $[h - h_0] \in \mathcal{H}$. Moreover, for all $(\theta, h) \in \Theta_{\delta_n} \times \mathcal{H}_{\delta_n}$ with a positive sequence $\delta_n = o(1)$:

G31 $||\mathrm{Z}(\theta, h_0) - \mathrm{Z}(\theta, h) - \mathrm{Z}_2(\theta, h_0)[h - h_0]||_\mathcal{L} = 0$ uniformly in θ.

G32 $||\mathrm{Z}_2(\theta, h_0)[h - h_0] - \mathrm{Z}_2(\theta_0, h_0)[h - h_0]||_\mathcal{L} \leq o(1)\delta_n$.

G.4 The estimator $\hat{h} \in \mathcal{H}$ with probability tending to one; and $||\hat{h} - h_0||_{\mathcal{H}} = o_{p^*}(n^{-1/4})$.

G.5 For any $\delta_n \downarrow 0$,

$$\sup_{||\theta-\theta_0||\leq\delta_n, ||h-h_0||_{\mathcal{H}}\leq\delta_n} \frac{||\sqrt{n}(\mathbb{Z}-\mathrm{Z})(\theta,h) - \sqrt{n}(\mathbb{Z}-\mathrm{Z})(\theta_0,h_0)||_{\mathcal{L}}}{1+\sqrt{n}||\mathbb{Z}(\theta,h)||_{\mathcal{L}} + \sqrt{n}||\mathrm{Z}(\theta,h)||_{\mathcal{L}}} = o_{p^*}(1).$$

G.6 $\sqrt{n}(\mathrm{Z}_2(\theta_0,h_0)[\hat{h}-h_0] + (\mathbb{Z}-\mathrm{Z})(\theta_0,h_0))$ converges weakly to a tight random element $\mathbb{G}$ in $\mathcal{L}$.

Condition **G.1** requires $\hat{\theta}$ to solve the estimating equation only asymptotically. Conditions **G.2** is a smoothness condition for Z. Conditions **G.3** and **G.4** are the same as conditions (2.3) and (2.4) of Chen et al. (2003). Conditions **G.5** and **G.6** are high level assumptions, and correspond to conditions (2.5) and (2.6) of Chen et al. (2003). More primitive conditions are provided for more specific cases. Moreover, condition **G.5** is implied by

G.5' For any $\delta_n \downarrow 0$,

$$\sup_{||\theta-\theta_0||\leq\delta_n, ||h-h_0||_{\mathcal{H}}\leq\delta_n} ||\sqrt{n}(\mathbb{Z}-\mathrm{Z})(\theta,h) - \sqrt{n}(\mathbb{Z}-\mathrm{Z})(\theta_0,h_0)||_{\mathcal{L}} = o_{p^*}(1).$$

Now we provide a general result for Z-estimators.

Lemma 2. *Suppose that* $\theta_0 \in \Theta_\delta$ *satisfies* $\mathrm{Z}(\theta_0,h_0) = 0$*, that* $\hat{\theta} = \theta_0 + o_{p^*}(1)$*, and that conditions* **G.1**–**G.6** *hold. Then* $\sqrt{n}(\hat{\theta}-\theta_0) \rightsquigarrow \mathrm{Z}_1^{-1}(\theta_0,h_0)\mathbb{G}$.

Proof. See Appendix A2. □

The Validity of the Bootstrap

There are two potential difficulties when constructing the confidence bands for the DRF. First, closed-form expressions of the covariance kernel are hard to calculate.

This mainly is due to the estimation of the nuisance parameters. Second, even if closed-form expressions of the covariance kernel are available, they are useful only when the set $\mathcal{T}$ is finite. Thus, we use the ordinary nonparametric bootstrap method for inferences. We show that the bootstrap estimator of the asymptotic distribution of $\sqrt{n}(\hat{\beta}(t) - \beta_0(t))$ is consistent. Let $\{(\boldsymbol{X}_i^*, Y_i^*, T_i^*)\}_{i=1}^n$ be randomly drawn with replacement from $\{(\boldsymbol{X}_i, Y_i, T_i)\}_{i=1}^n$. Let $\hat{\pi}^*$ be the estimator of π_0 using $\{(\boldsymbol{X}_i^*, Y_i^*, T_i^*)\}_{i=1}^n$. Let $\mathbb{Z}^*(\beta, \pi)$ denote the resampled average. The bootstrap estimator $\hat{\beta}^*$ satisfies

$$||\mathbb{Z}^*(\hat{\beta}^*, \hat{\pi}^*)|| = o_{p^*}(n^{-1/2}).$$

Following Chen et al. (2003), an asterisk denotes a probability or moment computed using bootstrapped data conditional on the original data set. Now consider the following conditions, which, respectively, correspond to conditions (2.4B), (2.5'B), and (2.6B) of Chen et al. (2003):

G4B With P*-probability tending to one, $\hat{\pi}^* \in \Pi$ and $||\hat{\pi}^* - \hat{\pi}||_\Pi = o_{p^*}(n^{-1/4})$.

G5B For any $\delta_n \downarrow 0$,

$$\sup_{||\beta-\beta_0||\le\delta_n, ||\pi-\pi_0||_\Pi\le\delta_n} ||\sqrt{n}(\mathbb{Z}^* - \mathbb{Z})(\beta, \pi) - \sqrt{n}(\mathbb{Z}^* - \mathbb{Z})(\beta_0, \pi_0)||_{\mathcal{L}} = o_{p^*}(1).$$

G6B $\sqrt{n}(\mathrm{Z}_2(\hat{\beta}, \hat{\pi})[\hat{\pi}^* - \hat{\pi}] + (\mathbb{Z}^* - \mathbb{Z})(\hat{\beta}, \hat{\pi}))$ converges weakly to a tight random element $\mathbb{G}$ in $\mathcal{L}$ in P*-probability.

The following lemma is the version of Theorem B of Chen et al. (2003) with a infinite dimensional parameter of interest.

Lemma 3. *Replace "θ" and "h" by "β" and "π", respectively, in conditions* **G.1**–**G.6**. *Suppose $\beta_0 \in \mathrm{int}(\mathcal{B})$ and $\hat{\beta} \overset{a.s.}{\to} \beta_0$. Assume that conditions* **G.1**, **G.4**, *and* **G.5** *are satisfied with "in probability" replaced by "almost surely". Let conditions* **G.2**

and **G.3** *hold with* π_0 *replaced by* $\pi \in \Pi_{\delta_n}$. *Also, assume that* $\mathrm{Z}_1(\beta;\pi)$ *is continuous in* π *at* $\beta = \beta_0$ *and* $\pi = \pi_0$. *Then* $\sqrt{n}(\hat{\beta}^* - \hat{\beta}) \rightsquigarrow \mathrm{Z}_1^{-1}(\beta_0, \pi_0)\mathbb{G}$ *in* P^**-probability.*

Proof. See Appendix A2. □

Appendix A2: Proofs of the Results in Appendix A1

This appendix collects the proofs for the asymptotic properties of a generic Z-estimator described in Lemmas 1–3 above.

Proof of **Lemma 1**. By condition **C.2**, it suffices to show that $||\mathrm{Z}(\hat{\theta}, h_0)||_{\mathcal{L}} = o_{p^*}(1)$. Using triangle inequality,

$$||\mathrm{Z}(\hat{\theta}, h_0)||_{\mathcal{L}} \leq ||\mathrm{Z}(\hat{\theta}, h_0) - \mathrm{Z}(\hat{\theta}, \hat{h})||_{\mathcal{L}} + ||\mathrm{Z}(\hat{\theta}, \hat{h}) - \mathbb{Z}(\hat{\theta}, \hat{h})||_{\mathcal{L}} + ||\mathbb{Z}(\hat{\theta}, \hat{h})||_{\mathcal{L}}.$$

By conditions **C.3** and **C.4**, $||\mathrm{Z}(\hat{\theta}, h_0) - \mathrm{Z}(\hat{\theta}, \hat{h})||_{\mathcal{L}} = o_{p^*}(1)$. By condition **C.1**, $||\mathbb{Z}(\hat{\theta}, \hat{h})||_{\mathcal{L}} = o_{p^*}(1)$. Also,

$$\begin{aligned} ||\mathrm{Z}(\hat{\theta}, \hat{h}) - \mathbb{Z}(\hat{\theta}, \hat{h})||_{\mathcal{L}} &= o_{p^*}(1) + o_{p^*}(||\mathbb{Z}(\hat{\theta}, \hat{h})||_{\mathcal{L}}) + o_{p^*}(||\mathrm{Z}(\hat{\theta}, \hat{h})||_{\mathcal{L}}) \\ &= o_{p^*}(1) + o_{p^*}(1) + o_{p^*}(||\mathrm{Z}(\hat{\theta}, h_0)||_{\mathcal{L}}) + o_{p^*}(1) \end{aligned}$$

where the first equality is by condition **C.5** and the second equality is by conditions **C.1** and **C.3**. Therefore, we have $||\mathrm{Z}(\hat{\theta}, h_0)||_{\mathcal{L}} \leq o_{p^*}(1)$ and hence the result. □

Proof of **Lemma 2**. **Step 1:** $\sqrt{n}$**-consistency**

We start the proof by showing that $\hat{\theta}$ is $\sqrt{n}$-consistent for θ_0 in Θ. By definition, the Fréchet differentiability of $\mathrm{Z}(\theta, h_0)$ implies the existence of a continuous linear map $\mathrm{Z}_1(\theta_0, f_0)$ such that

$$\frac{||\mathrm{Z}(\theta, f_0) - \mathrm{Z}(\theta_0, f_0) - \mathrm{Z}_1(\theta_0, f_0)(\theta - \theta_0)||_{\mathcal{L}}}{||\theta - \theta_0||_\Theta} = o(1).$$

By triangle inequality, it follows

$$||\mathrm{Z}_1(\theta_0, h_0)(\theta - \theta_0)||_{\mathcal{L}} \leq ||\mathrm{Z}(\theta, h_0) - \mathrm{Z}(\theta_0, h_0)||_{\mathcal{L}} + o(||\theta - \theta_0||_\Theta).$$

Since the derivative $\mathrm{Z}_1(\theta_0, h_0)$ is continuously invertible by condition **G.2**, there exists a positive constant c such that $||\mathrm{Z}_1(\theta_0, h_0)(\theta_1 - \theta_2)||_{\mathcal{L}} \geq c||\theta_1 - \theta_2||_\Theta$ for every θ_1 and $\theta_2 \in \Theta_\delta$. Therefore, it follows

$$(c - o(1))||\theta - \theta_0||_\Theta \leq ||\mathrm{Z}(\theta, h_0) - \mathrm{Z}(\theta_0, h_0)||_{\mathcal{L}}, \tag{A1}$$

and

$$(c - o_{p^*}(1))||\hat{\theta} - \theta_0||_\Theta \leq ||\mathrm{Z}(\hat{\theta}, h_0) - \mathrm{Z}(\theta_0, h_0)||_{\mathcal{L}} = ||\mathrm{Z}(\hat{\theta}, h_0)||_{\mathcal{L}} \tag{A2}$$

with probability tending to one. By triangle inequality and conditions **G.1** and **G.6**, the right hand side of the previous inequality is bounded by

$$||\mathrm{Z}(\hat{\theta}, h_0) - \mathrm{Z}(\hat{\theta}, \hat{h})||_{\mathcal{L}} + ||\mathrm{Z}(\hat{\theta}, \hat{h}) - \mathbb{Z}(\hat{\theta}, \hat{h}) + \mathrm{Z}(\theta_0, h_0) - \mathbb{Z}(\theta_0, h_0)||_{\mathcal{L}} + O_p(n^{-1/2}). \tag{A3}$$

For the first term, we have

$$\begin{aligned}||\mathrm{Z}(\hat{\theta},h_0)-\mathrm{Z}(\hat{\theta},\hat{h})||_{\mathcal{L}} \leq & ||\mathrm{Z}(\hat{\theta},h_0)-\mathrm{Z}(\hat{\theta},\hat{h})-\mathrm{Z}_2(\hat{\theta},h_0)[\hat{h}-h_0]||_{\mathcal{L}} \\ & + ||\mathrm{Z}_2(\hat{\theta},h_0)[\hat{h}-h_0]-\mathrm{Z}_2(\theta_0,h_0)[\hat{h}-h_0]||_{\mathcal{L}} \\ & + ||\mathrm{Z}_2(\theta_0,h_0)[\hat{h}-h_0]||_{\mathcal{L}} \\ \leq & o_{p^*}(1)||\theta-\theta_0||_{\Theta}+O_{p^*}(n^{-1/2}) \\ \leq & ||\mathrm{Z}(\hat{\theta},h_0)||_{\mathcal{L}}\times o_{p^*}(1)+O_{p^*}(n^{-1/2}),\end{aligned}$$

where the first inequality is by triangle inequality, the second one by conditions **G.3** and **G.6**, and the third by inequality (A1).

As for the second term in (A3), by condition **G.5**,

$$\begin{aligned}||\mathrm{Z}(\hat{\theta},\hat{h})-\mathbb{Z}(\hat{\theta},\hat{h})+\mathrm{Z}(\theta_0,h_0)-\mathbb{Z}(\theta_0,h_0)||_{\mathcal{L}} = & o_{p^*}(1/\sqrt{n}+||\mathbb{Z}(\hat{\theta},\hat{h})||_{\mathcal{L}}+||\mathrm{Z}(\hat{\theta},\hat{h})||_{\mathcal{L}}) \\ = & o_{p^*}(1/\sqrt{n})+o_{p^*}(||\mathrm{Z}(\hat{\theta},\hat{h})||_{\mathcal{L}})\end{aligned}$$

The second equality is by condition **G.1**, $||\mathbb{Z}(\hat{\theta},\hat{h})||_{\mathcal{L}} = o_{p^*}(1/\sqrt{n})$. By triangle inequality,

$$||\mathrm{Z}(\hat{\theta},\hat{h})||_{\mathcal{L}} \leq ||\mathrm{Z}(\hat{\theta},\hat{h})-\mathbb{Z}(\hat{\theta},\hat{h})+\mathrm{Z}(\theta_0,h_0)-\mathbb{Z}(\theta_0,h_0)||_{\mathcal{L}}+O_{p^*}(1/\sqrt{n}).$$

It then follows

$$(1-o_{p^*}(1))||\mathrm{Z}(\hat{\theta},\hat{h})-\mathbb{Z}(\hat{\theta},\hat{h})+\mathrm{Z}(\theta_0,h_0)-\mathbb{Z}(\theta_0,h_0)||_{\mathcal{L}} \leq o_{p^*}(1/\sqrt{n})$$

Thus, formula (A3) is bounded by

$$||\mathrm{Z}(\hat{\theta},h_0)||_{\mathcal{L}}\times o_{p^*}(1)+O_{p^*}(n^{-1/2})$$

and the right side of the equality in (A2) satisfies

$$(1 - o_{p^*}(1))||\mathrm{Z}(\hat{\theta}, h_0)||_{\mathcal{L}} \leq O_{p^*}(n^{-1/2}). \tag{A4}$$

Therefore, $(c - o_p(1))\sqrt{n}||\hat{\theta} - \theta_0||_\Theta \leq O_{p^*}(1)$ and $\hat{\theta}$ is $\sqrt{n}$-consistent for θ_0 in Θ.

Step 2: Weak Convergence

Now we show the weak convergence. By conditions **G.2** and **G.3**,

$$\begin{aligned}
&||\mathrm{Z}(\hat{\theta}, \hat{h}) - \mathrm{Z}(\theta_0, h_0) - \mathrm{Z}_1(\theta_0, h_0)(\hat{\theta} - \theta_0) - \mathrm{Z}_2(\theta_0, h_0)[\hat{h} - h_0]||_{\mathcal{L}} \\
=&||\mathrm{Z}(\hat{\theta}, \hat{h}) - \mathrm{Z}(\hat{\theta}, h_0) - \mathrm{Z}_2(\hat{\theta}, h_0)[\hat{h} - h_0] + \mathrm{Z}(\hat{\theta}, h_0) - \mathrm{Z}(\theta_0, h_0) - \mathrm{Z}_1(\theta_0, h_0)(\hat{\theta} - \theta_0) \\
&+ \mathrm{Z}_2(\hat{\theta}, h_0)[\hat{h} - h_0] - \mathrm{Z}_2(\theta_0, h_0)[\hat{h} - h_0]||_{\mathcal{L}} \\
\leq&||\mathrm{Z}(\hat{\theta}, \hat{h}) - \mathrm{Z}(\hat{\theta}, h_0) - \mathrm{Z}_2(\hat{\theta}, h_0)[\hat{h} - h_0]||_{\mathcal{L}} \\
&+ ||\mathrm{Z}(\hat{\theta}, h_0) - \mathrm{Z}(\theta_0, h_0) - \mathrm{Z}_1(\theta_0, h_0)(\hat{\theta} - \theta_0)||_{\mathcal{L}} \\
&+ ||\mathrm{Z}_2(\hat{\theta}, h_0)[\hat{h} - h_0] - \mathrm{Z}_2(\theta_0, h_0)[\hat{h} - h_0]||_{\mathcal{L}} \\
=&o_{p^*}(n^{-1/2}) + o_{p^*}(n^{-1/2}) = o_{p^*}(n^{-1/2}).
\end{aligned}$$

Therefore, it follows

$$\mathrm{Z}_1(\theta_0, h_0)\sqrt{n}(\hat{\theta} - \theta_0) + \sqrt{n}\mathrm{Z}_2(\theta_0, h_0)[\hat{h} - h_0] = \sqrt{n}(\mathbb{Z}(\theta_0, h_0) - \mathrm{Z}(\theta_0, h_0)) + o_{p^*}(1)$$

and

$$\mathrm{Z}_1(\theta_0, h_0)\sqrt{n}(\hat{\theta} - \theta_0) = -\sqrt{n}(\mathrm{Z}_2(\theta_0, h_0)[\hat{h} - h_0] + (\mathbb{Z} - \mathrm{Z})(\theta_0, h_0)) + o_{p^*}(1) \rightsquigarrow \mathbb{G}$$

by condition **G.6**.

Now by condition **G.2** and the continuous mapping theorem, we have $\sqrt{n}(\hat{\theta} - \theta_0) \rightsquigarrow \mathrm{Z}_1^{-1}(\theta_0, h_0)\mathbb{G}$. □

Proof of **Lemma 3**. The assertion that $||\hat{\beta}^* - \hat{\beta}|| = O_{p^*}(n^{-1/2})$ a.s. [P] can be shown in a similar way as the proof of the $\sqrt{n}$-consistency of $\hat{\beta}$. Therefore we omit the proof and only show the weak convergence in probability of the bootstrap estimator.

Note that

$$
\begin{aligned}
&||\mathbb{Z}^*(\hat{\beta}^*, \hat{\pi}^*) - \mathbb{Z}^*(\hat{\beta}, \hat{\pi}) - \mathrm{Z}_1(\hat{\beta}, \hat{\pi})(\hat{\beta}^* - \hat{\beta}) - \mathrm{Z}_2(\hat{\beta}, \hat{\pi})[\hat{\pi}^* - \hat{\pi}]|| \\
=&||\mathrm{Z}(\hat{\beta}^*, \hat{\pi}^*) - \mathrm{Z}(\hat{\beta}^*, \hat{\pi}) - \mathrm{Z}_2(\hat{\beta}, \hat{\pi})[\hat{\pi}^* - \hat{\pi}] + \mathrm{Z}(\hat{\beta}^*, \hat{\pi}) - \mathrm{Z}(\hat{\beta}, \hat{\pi}) - \mathrm{Z}_1(\hat{\beta}, \hat{\pi})(\hat{\beta}^* - \hat{\beta}) \\
&+ [(\mathbb{Z}^*(\hat{\beta}^*, \hat{\pi}^*) - \mathbb{Z}(\hat{\beta}^*, \hat{\pi}^*)) - (\mathbb{Z}^*(\hat{\beta}, \hat{\pi}) - \mathbb{Z}(\hat{\beta}, \hat{\pi}))] \\
&+ [(\mathbb{Z}(\hat{\beta}^*, \hat{\pi}^*) - \mathrm{Z}(\hat{\beta}^*, \hat{\pi}^*)) - (\mathbb{Z}(\hat{\beta}, \hat{\pi}) - \mathrm{Z}(\hat{\beta}, \hat{\pi}))] \\
&+ \mathrm{Z}_2(\hat{\beta}, \hat{\pi})[\hat{\pi}^* - \hat{\pi}] - \mathrm{Z}_2(\hat{\beta}^*, \hat{\pi})[\hat{\pi}^* - \hat{\pi}]|| \\
\leq&||\mathrm{Z}(\hat{\beta}^*, \hat{\pi}^*) - \mathrm{Z}(\hat{\beta}^*, \hat{\pi}) - \mathrm{Z}_2(\hat{\beta}, \hat{\pi})[\hat{\pi}^* - \hat{\pi}]|| \\
&+ ||\mathrm{Z}(\hat{\beta}^*, \hat{\pi}) - \mathrm{Z}(\hat{\beta}, \hat{\pi}) - \mathrm{Z}_1(\hat{\beta}, \hat{\pi})(\hat{\beta}^* - \hat{\beta})|| \\
&+ ||(\mathbb{Z}^*(\hat{\beta}^*, \hat{\pi}^*) - \mathbb{Z}(\hat{\beta}^*, \hat{\pi}^*)) - (\mathbb{Z}^*(\hat{\beta}, \hat{\pi}) - \mathbb{Z}(\hat{\beta}, \hat{\pi}))|| \\
&+ ||(\mathbb{Z}(\hat{\beta}^*, \hat{\pi}^*) - \mathrm{Z}(\hat{\beta}^*, \hat{\pi}^*)) - (\mathbb{Z}(\hat{\beta}, \hat{\pi}) - \mathrm{Z}(\hat{\beta}, \hat{\pi}))|| \\
&+ ||\mathrm{Z}_2(\hat{\beta}, \hat{\pi})[\hat{\pi}^* - \hat{\pi}] - \mathrm{Z}_2(\hat{\beta}^*, \hat{\pi})[\hat{\pi}^* - \hat{\pi}]|| \\
=&o_{p^*}(n^{-1/2}).
\end{aligned}
$$

The first term is $o_{p^*}(n^{-1/2})$ by condition **G3** (version of **Lemma 3**) and **G4B**. The second term is $o_{p^*}(n^{-1/2})$ by condition **G2** (version of **Lemma 3**) and $\sqrt{n}$-consistency of $\hat{\theta}^*$. The third and fourth terms are $o_{p^*}(n^{-1/2})$ by the triangular inequality and conditions **G5'** (almost sure version) and **G5B**. And the fifth term is $o_{p^*}(n^{-1/2})$ by condition **G3** (version of **Lemma 3**) and $\sqrt{n}$-consistency of $\hat{\theta}^*$.

Therefore, it follows

$$\begin{aligned} Z_1(\hat{\beta},\hat{\pi})\sqrt{n}(\hat{\beta}^* - \hat{\beta}) + \sqrt{n}Z_2(\hat{\beta},\hat{\pi})[\hat{\pi}^* - \hat{\pi}] =& \sqrt{n}(\mathbb{Z}^*(\hat{\beta}^*,\hat{\pi}^*) - \mathbb{Z}^*(\hat{\beta},\hat{\pi})) + o_{p^*}(1) \\ =& -\sqrt{n}(\mathbb{Z}^*(\hat{\beta},\hat{\pi}) - \mathbb{Z}(\hat{\beta},\hat{\pi})) + o_{p^*}(1) \end{aligned}$$

and

$$\begin{aligned} & Z_1(\hat{\beta},\hat{\pi})\sqrt{n}(\hat{\beta}^* - \hat{\beta}) \\ =& -\sqrt{n}Z_2(\hat{\beta},\hat{\pi})[\hat{\pi}^* - \hat{\pi}] - \sqrt{n}(\mathbb{Z}^*(\hat{\beta},\hat{\pi}) - \mathbb{Z}(\hat{\beta},\hat{\pi})) + o_{p^*}(1) \rightsquigarrow \mathbb{G} \end{aligned}$$

in $\mathcal{L}$ in P^*-probability by condition **G.6**. We can replace $Z_1(\hat{\beta},\hat{\pi})$ by $Z_1(\beta_0,\pi_0)$ with probability one. Now by condition **G.2** (version of **Lemma 3**) and the continuous mapping theorem, we have $\sqrt{n}(\hat{\beta} - \beta_0) \rightsquigarrow Z_1^{-1}(\beta_0,\pi_0)\mathbb{G}$. □

Appendix A3: Long Proofs of the Results in Chapter 1

This appendix collects the proofs for the results given in the text.

Proof of **Theorem 1**. Fixing $t = t_0$, by law of iterated expectations, $\mathrm{E}[m(Y(t_0);\beta(t_0))] = \mathrm{E}\{\mathrm{E}[m(Y(t_0);\beta(t_0))|\boldsymbol{X}]\}$. For the conditional expectation,

$$\begin{aligned} \mathrm{E}[m(Y(t_0);\beta(t_0))|\boldsymbol{X}] =& \mathrm{E}[m(Y(t_0);\beta(t_0))|\boldsymbol{X}, T = t_0] = \mathrm{E}[m(Y;\beta(t_0))|\boldsymbol{X}, T = t_0] \\ =& \lim_{\Delta t \downarrow 0} \mathrm{E}[m(Y;\beta(t_0))|\boldsymbol{X}, T \in [t_0, t_0 + \Delta t]], \end{aligned}$$

where the first equality is by condition **I.II.1**, the second equation is by the fact that if $T = t_0$, then $Y = Y(t_0)$, and the third equality is by condition **I.III.2**. Moreover, we have

$$\begin{aligned} &\mathrm{E}[m(Y;\beta(t_0))|\boldsymbol{X}, T \in [t_0, t_0+\Delta t]] \\ &\qquad = \mathrm{E}[\mathbf{1}(T \in [t_0, t_0+\Delta t]) m(Y;\beta(t_0))|\boldsymbol{X}, T \in [t_0, t_0+\Delta t]]. \end{aligned} \tag{A5}$$

By law of total expectation

$$\begin{aligned} &\mathrm{E}[\mathbf{1}\{T \in [t_0, t_0+\Delta t]\} m(Y;\beta(t_0))|\boldsymbol{X}] \\ =&\mathrm{E}[\mathbf{1}\{T \in [t_0, t_0+\Delta t]\} m(Y;\beta(t_0))|\boldsymbol{X}, T \in [t_0, t_0+\Delta t]]\mathrm{P}(T \in [t_0, t_0+\Delta t]|\boldsymbol{X}) \\ +&\mathrm{E}[\mathbf{1}\{T \in [t_0, t_0+\Delta t]\} m(Y;\beta(t_0))|\boldsymbol{X}, T \notin [t_0, t_0+\Delta t]]\mathrm{P}(T \notin [t_0, t_0+\Delta t]|\boldsymbol{X}), \end{aligned}$$

the right hand side of equation (A5) equals

$$\begin{aligned} &\frac{\mathrm{E}[\mathbf{1}\{T \in [t_0, t_0+\Delta t]\} m(Y;\beta(t_0))|\boldsymbol{X}]}{\mathrm{P}(T \in [t_0, t_0+\Delta t]|\boldsymbol{X})} \\ =&\frac{\mathrm{E}[(\mathbf{1}\{T \le t_0+\Delta t\} - \mathbf{1}\{T < t_0\}) m(Y;\beta(t_0))|\boldsymbol{X}]}{F_{T|\boldsymbol{X}}(t_0+\Delta t|\boldsymbol{X}) - F_{T|\boldsymbol{X}}(t_0|\boldsymbol{X})}, \end{aligned}$$

where $F_{T|\boldsymbol{X}}$ denotes the conditional distribution function of T given $\boldsymbol{X}$. Noting that

$$\begin{aligned} &\mathrm{E}[(\mathbf{1}\{T \le t_0+\Delta t\} - \mathbf{1}\{T \le t_0\}) m(Y;\beta(t_0))|\boldsymbol{X} = \boldsymbol{x}] \\ =&\iint (\mathbf{1}\{t \le t_0+\Delta t\} - \mathbf{1}\{t \le t_0\}) m(y;\beta(t_0)) f_{T,Y|\boldsymbol{X}}(t, y|\boldsymbol{x})\, dtdy \\ =&\int \int_{t_0}^{t_0+\Delta t} m(y;\beta(t_0)) f_{T,Y|\boldsymbol{X}}(t, y|\boldsymbol{x})\, dtdy, \end{aligned}$$

it follows

$$
\begin{aligned}
&\lim_{\Delta t \downarrow 0} \mathrm{E}[m(Y;\beta(t_0))|\boldsymbol{X}=\boldsymbol{x}, T\in[t_0,t_0+\Delta t]] \\
=&\lim_{\Delta t \downarrow 0} \frac{\int\int_{t_0}^{t_0+\Delta t} m(y;\beta(t_0)) f_{T,Y|\boldsymbol{X}}(t,y|\boldsymbol{x})\, dtdy}{F_{T|\boldsymbol{X}}(t_0+\Delta t|\boldsymbol{x}) - F_{T|\boldsymbol{X}}(t_0|\boldsymbol{x})} \\
=&\lim_{\Delta t \downarrow 0} \frac{\int m(y;\beta(t_0)) f_{T,Y|\boldsymbol{X}}(t_0+\epsilon_1\Delta t, y|\boldsymbol{x})\, dy}{f_{T|\boldsymbol{x}}(t_0+\epsilon_2\Delta t|\boldsymbol{x})} \\
=&\frac{\int m(y;\beta(t_0)) f_{T,Y|\boldsymbol{X}}(t_0,y|\boldsymbol{x}) \frac{f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})}{f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})}\, dy}{f_{T|\boldsymbol{X}}(t_0|\boldsymbol{x})} \\
=&\frac{\int m(y;\beta(t_0)) f_{T|\boldsymbol{X},Y}(t_0|\boldsymbol{x},y) f_{Y|\boldsymbol{X}}(y|\boldsymbol{x})\, dy}{f_{T|\boldsymbol{X}}(t_0|\boldsymbol{x})} \\
=&\frac{\mathrm{E}[m(Y;\beta(t_0)) f_{T|\boldsymbol{X},Y}(t_0|\boldsymbol{x},Y)|\boldsymbol{X}=\boldsymbol{x}]}{f_{T|\boldsymbol{X}}(t_0|\boldsymbol{x})}
\end{aligned}
$$

where ϵ_1 and ϵ_2 are fixed numbers in [0,1]. The second equality is by mean value theorems for differentiation and integration. And the third equality is by condition **I.III.1** and dominated convergence theorem. Hence $\mathrm{E}[m(Y(t_0);\beta(t_0))] = \mathrm{E}\left[m(Y;\beta(t_0))\frac{f_{T|Y,\boldsymbol{X}}(t_0|Y,\boldsymbol{X})}{f_{T|\boldsymbol{X}}(t_0|\boldsymbol{X})}\right]$. □

Proof of **Theorem 2**. The general result in the previous lemma for consistency of the Z-estimator can be applied to our continuous treatment model as stated in the following theorem with $\theta_0 = \beta_0(t)$, $h_0 = \pi(\cdot;t)$, $\mathbb{Z}(\theta,h)(t) = \mathbb{E}\psi_{\beta,\pi,t}$, and $\mathrm{Z}(\theta,h)(t) = \mathrm{E}\psi_{\beta,\pi,t}$, where $\psi_{\beta,\pi,t} = m(y;\beta(t))\pi(\boldsymbol{u};t)$. In this case, $\Theta = \mathcal{L} = \ell^\infty(\mathcal{T})$ and $||\cdot||_\Theta = ||\cdot||_{\mathcal{L}} = |\cdot|_\infty$, while $\mathcal{H} = \Pi$, a function class with domain $\mathcal{U}\times\mathcal{T}$, and $||\cdot||_{\mathcal{H}} = ||\cdot||_\Pi = \sup_{t\in\mathcal{T}}\sup_{\boldsymbol{u}\in\mathcal{U}}|\cdot| = |\cdot|_\infty$. For any $\delta>0$, $\Pi_\delta = \{\pi\in\Pi : |\pi-\pi_0|_\infty < \delta\}$.

First we show that condition **C.3**, the continuity of $\mathrm{E}m(Y;\beta(t))\pi(\boldsymbol{U};t)$ at π_0 uniformly over $\beta(t)\in\ell^\infty(\mathcal{T})$, is satisfied. For any $|\pi-\pi_0|_\infty \le \delta$, which is equivalent to $\sup_{t\in\mathcal{T}}\sup_{\boldsymbol{u}\in\mathcal{U}}|\pi(\boldsymbol{u};t)-\pi_0(\boldsymbol{u};t)|\leq\delta$, we have

$$\begin{aligned}&|\mathrm{E}[m(Y;\beta(t))\pi(\boldsymbol{U};t)]-\mathrm{E}[m(Y;\beta(t))\pi_0(\boldsymbol{U};t)]|_\infty\\=&|\mathrm{E}[m(Y;\beta(t))(\pi(\boldsymbol{U};t)-\pi_0(\boldsymbol{U};t))]|_\infty\leq|\mathrm{E}[m(Y;\beta(t))]|_\infty\delta.\end{aligned}$$

Therefore, condition **C.3** is implied by condition **C.III**.

Condition **C.V'** and Corollary 9.27 of Kosorok (2008) imply condition **C5S** which in turn implies **C.5**. □

Proof of **Corollary 1**, *Consistency of* $\hat{\mu}(t)$. To show consistency, we verify the conditions of **Theorem 2**. Note that in this case we have a closed form solution $\mu_0(t)=\frac{\mathrm{E}Y\pi_0(\boldsymbol{U};t)}{\mathrm{E}\pi_0(\boldsymbol{U};t)}$. Condition **C.I** is verified by the fact that $\hat{\mu}=\frac{\mathbb{E}Y\hat{\pi}_0(\boldsymbol{U})}{\mathbb{E}\hat{\pi}_0(\boldsymbol{U})}$ is the exact zero of the estimating equation.

For condition **C.II**, if there is a sequence $\mu_n(t)$ such that $\mathrm{E}[(Y-\mu_n(t))\pi_0(\boldsymbol{U};t)]\to 0$ uniformly, then $\mu_n(t)\to\frac{\mathrm{E}Y\pi_0(\boldsymbol{U};t)}{\mathrm{E}\pi_0(\boldsymbol{U};t)}=\mu_0(t)$ uniformly. To see this, we note that $\mathrm{E}[(Y-\mu_n(t))\pi_0(\boldsymbol{U};t)]=o(1)$ implies $\mu_n(t)=\frac{\mathrm{E}Y\pi_0(\boldsymbol{U};t)}{\mathrm{E}\pi_0(\boldsymbol{U};t)}+\frac{o(1)}{\mathrm{E}\pi_0(\boldsymbol{U};t)}=\frac{\mathrm{E}Y\pi_0(\boldsymbol{U};t)}{\mathrm{E}\pi_0(\boldsymbol{U};t)}+o(1)$.

Condition **C.III** is verified by the direct calculation

$$|\mathrm{E}[m(Y;\mu(t))]|_\infty=|\mathrm{E}[Y-\mu(t)]|_\infty=\mathrm{E}|Y|+|\mu(t)|_\infty,$$

both of which are finite by condition **AC.I**.

As for condition **C.V'**, noting that $\psi_{1\mu,t}=y-\mu(t)$, automatically $\{\psi_{1\beta,t}:\beta\in\ell^\infty(\mathcal{T}),t\in\mathcal{T}\}$ is Glivenko-Cantelli. To see this, note

$$\sup_{\mu(t)\in\ell^\infty(\mathcal{T})}|\mathbb{E}[Y-\mu(t)]-\mathrm{E}[Y-\mu(t)]|_\infty=|\mathbb{E}Y-\mathrm{E}Y|\xrightarrow{p}0$$

by Khintchine's weak law of large numbers. It has envelope $F_1(y)=y+|\mu(t)|_\infty$ by condition **AC.I**. Condition **AC.II** completes the verification of condition **C.V'**.

Hence, all the conditions of **Theorem 2** are satisfied. □

Proof of **Corollary 2**, *Consistency of* $\hat{q}_\tau(t)$. Condition **C.I** is satisfied by the computational properties of quantile regression estimator of Theorem 3.3 of Koenker and Bassett (1978) and conditions **C.4** and **QC.II**

$$|\mathbb{E}[(\tau - \mathbf{1}\{Y < \hat{q}_\tau(t)\})\hat{\pi}(\boldsymbol{U};t)]| \leq const \cdot \sup_{i \leq n} \frac{\hat{\pi}(\boldsymbol{U}_i;t)}{n}$$
$$\leq const \cdot \frac{||\hat{\pi}(\boldsymbol{u};t)||_\Pi}{n} = const \cdot \frac{||\pi_0(\boldsymbol{u};t)||_\Pi + o_p(1)}{n} = O_{p^*}(1/n).$$

Condition **C.II** holds by condition **QC.I**. Condition **C.III** is satisfied because $\tau - \mathbf{1}\{y < q_\tau(t)\}$ is a bounded function. Condition **CV'** is implied by the fact that the function class $\{\psi_{1q,t} : q \in \ell^\infty(\mathcal{T}), t \in \mathcal{T}\}$ is Glivenko-Cantelli because it is a Vapnik-Červonenkis class and by condition **QC.III**. □

Proof of **Theorem 3**. We first verify condition **G.3**. To find the pathwise derivative of $\mathrm{Z}(\beta, \pi_0)$ with respect to π, we conduct the following calculations. For any $\bar{\pi}$ such that $\{\pi_0 + \alpha(\bar{\pi} - \pi_0) : \alpha \in [0,1]\} \subset \Pi$,

$$\frac{\mathrm{E}[m(Y;\beta)(\pi_0 + \alpha(\bar{\pi} - \pi_0))] - \mathrm{E}[m(Y;\beta)\pi_0]}{\alpha} = \mathrm{E}[m(Y;\beta)(\bar{\pi} - \pi_0)]$$

and has the limit $\mathrm{E}[m(Y;\beta)(\bar{\pi} - \pi_0)]$ as $\alpha \to 0$. Therefore $\mathrm{Z}_2(\beta, \pi_0)[\pi - \pi_0] = \mathrm{E}[m(Y;\beta)(\pi - \pi_0)]$ in all directions $[\pi - \pi_0] \in \Pi$.

Condition **G31** is satisfied by noting

$$|\mathrm{E}[m(Y;\beta(t))\pi_0(\boldsymbol{U};t)] - \mathrm{E}[m(Y;\beta(t))\pi(\boldsymbol{U};t)] - \mathrm{E}[m(Y;\beta(t))(\pi - \pi_0)(\boldsymbol{U};t)]|_\infty = 0.$$

And condition **G32** is verified by

$$
\begin{aligned}
&|\mathrm{E}[m(Y;\beta(t))(\pi-\pi_0)(\boldsymbol{U};t)]-\mathrm{E}[m(Y;\beta_0(t))(\pi-\pi_0)(\boldsymbol{U};t)]|_\infty\\
=&|\mathrm{E}[m(Y;\beta(t))-m(Y;\beta_0(t))(\pi-\pi_0)(\boldsymbol{U};t)]|_\infty\\
\le&|\mathrm{E}[m(Y;\beta(t))]-\mathrm{E}[m(Y;\beta_0(t))]|_\infty o(1)=\delta_n o(1),
\end{aligned}
$$

where the last equality is by condition **G.III**.

As for condition **G.5**, by Corollary 9.32 (iii) of Kosorok (2008), condition **G.V** implies that $\{\psi_{\beta,\pi,t}:\beta\in\ell_\delta^\infty(\mathcal{T}),\pi\in\Pi_\delta,t\in\mathcal{T}\}$ is Donsker, which in turn implies **G.5'** by Lemma 3.3.5 of van der Vaart and Wellner (1996). Therefore, we obtain condition **G.5** by condition **G.1** and inequality (A4).

Finally, **G.VI** is a representation of **G.6**. □

Proof of **Corollary 3**, *Weak Convergence of* $\hat{\mu}(t)$. Condition **G.1** is satisfied because the estimator is an exact Z-estimator.

The map $\mu\mapsto\mathrm{E}(Y-\mu)f_0(\boldsymbol{U})$ is Fréchet differentiable and is verified by the following calculation

$$
\frac{|\mathrm{E}[(Y-\mu(t))\pi_0(\boldsymbol{U};t)]-\mathrm{E}[(Y-\mu_0(t))\pi_0(\boldsymbol{U};t)]-\mathrm{E}[\pi_0(\boldsymbol{U};t)(\mu(t)-\mu_0(t))]|_\infty}{|\mu(t)-\mu_0(t)|_\infty}=0.
$$

Thus the Fréchet derivative is $\mathrm{E}\pi_0(\boldsymbol{U};t)$. For any μ_1 and μ_2,

$$
|\mathrm{E}[\pi_0(\boldsymbol{U};t)\mu_1(t)]-\mathrm{E}[\pi_0(\boldsymbol{U};t)\mu_2(t)]|_\infty\ge c|\mu_1(t)-\mu_2(t)|_\infty,
$$

and therefore is continuously invertible.

Condition **G.III** is verified by

$$
|\mathrm{E}[Y-\mu(t)]-\mathrm{E}[Y-\mu_0(t)]|_\infty=|\mu(t)-\mu_0(t)|_\infty.
$$

Condition **G.V** is implied by Conditions **AG.I** and **AG.II** and Corollary 9.32 (i) of Kosorok (2008) completes the verification.

Finally condition **G.VI** is implied by condition **AG.III**. □

Proof of **Corollary 4**, *Weak Convergence of* $\hat{q}_\tau(t)$. Condition **G.I** was verified in the proof of **Corollary 2**. For condition **G.II**, note that

$$\begin{aligned}
&|\mathrm{E}[(\tau - \mathbf{1}\{Y \leq q_\tau(t)\})\pi_0(\boldsymbol{U};t)] - \mathrm{E}[(\tau - \mathbf{1}\{Y \leq q_{\tau 0}(t)\})\pi_0(\boldsymbol{U};t)] \\
&+ \mathrm{E}[\pi_0(\boldsymbol{U};t) f_Y(q_{\tau 0})(q_\tau(t) - q_{\tau 0}(t))]|_\infty \\
=&|\mathrm{E}[\{\mathbf{1}\{Y \leq q_{\tau 0}(t)\} - \mathbf{1}\{Y \leq q_\tau(t)\} + f_Y(q_{\tau 0})(q_\tau(t) - q_{\tau 0}(t))\}\pi_0(\boldsymbol{U};t)]|_\infty \\
\asymp&|\mathrm{E}[\{\mathbf{1}\{Y \leq q_{\tau 0}(t)\} - \mathbf{1}\{Y \leq q_\tau(t)\} + f_Y(q_{\tau 0})(q_\tau(t) - q_{\tau 0}(t))\}]|_\infty M_\pi \\
=&|F_Y(q_{\tau 0}(t)) - F_Y(q_\tau(t)) + f_Y(q_{\tau 0})(q_\tau(t) - q_{\tau 0}(t))|_\infty M_\pi = o(|q_\tau(t) - q_{\tau 0}(t)|_\infty).
\end{aligned}$$

Condition **G.III** is satisfied because the distribution function of Y is continuous. And condition **G.V** was verified in the proof of **Corollary 2**. Condition **G.VI** holds by condition **QG.II**. □

Proof of **Claim 1**. A regular parametric submodel of the joint distribution of $(Y, T, \boldsymbol{X})$ with distribution function $F(y, t, \boldsymbol{x}; \gamma)$ has the log-likelihood

$$\begin{aligned}
\log f(y,t,\boldsymbol{x};\gamma) &= \int_{\varsigma \in \mathcal{T}} \left[\log f_{Y(\varsigma)|\boldsymbol{X}}(y|\boldsymbol{x};\gamma) + \log f_{T|\boldsymbol{X}}(\varsigma|\boldsymbol{x};\gamma)\right] \, d\mathbf{1}\{\varsigma \geq t\} + \log f_{\boldsymbol{X}}(\boldsymbol{x};\gamma) \\
&= \left[\log f_{Y(t)|\boldsymbol{X}}(y|\boldsymbol{x};\gamma) + \log f_{T|\boldsymbol{X}}(t|\boldsymbol{x};\gamma)\right] + \log f_{\boldsymbol{X}}(\boldsymbol{x};\gamma)
\end{aligned}$$

with $F(y,t,\boldsymbol{x};\gamma_0) = F(y,t,\boldsymbol{x})$. The score of this model is

$$\begin{aligned} S(y,t,\boldsymbol{x};\gamma_0) &= \frac{d}{d\gamma}\log f(y,t,\boldsymbol{x};\gamma_0) \\ &= S_y(y,t,\boldsymbol{x}) + S_T(t,\boldsymbol{x}) + S_{\boldsymbol{X}}(\boldsymbol{x}), \end{aligned}$$

where

$$\begin{aligned} S_y(y,t,\boldsymbol{x}) =& \frac{d}{d\gamma}\log f_{Y(t)|\boldsymbol{X}}(y|\boldsymbol{x};\gamma_0) \\ S_T(t,\boldsymbol{x}) =& \frac{d}{d\gamma}\log f_{T|\boldsymbol{X}}(t|\boldsymbol{x};\gamma_0) \\ S_{\boldsymbol{X}}(\boldsymbol{x}) =& \frac{d}{d\gamma}\log f_{\boldsymbol{X}}(\boldsymbol{x};\gamma_0), \end{aligned}$$

Therefore, the tangent set of this model is the collection of the score functions of the form above and the tangent space $\dot{\mathcal{P}}_F$ is the closed linear span of the tangent set.

Recall that $\mathrm{E}[m(Y(t);\beta(t))] = 0$ if and only if $\beta(t) = \beta_0(t)$ for each $t \in \mathcal{T}$. By implicit function theorem,

$$\frac{\partial \beta_0(t)}{\partial \gamma}(\gamma) = -\Gamma_0^{-1}(t)\Upsilon(\gamma_0)$$

and

$$\begin{aligned} \Upsilon(\gamma_0) =& \frac{\partial}{\partial\gamma}\int m(Y(t);\beta(t))\, dF(y,t,\boldsymbol{x};\gamma_0) \\ =& \mathrm{E}[m(Y(t);\beta_0(t))S_y(Y(t),t,\boldsymbol{X})] + \mathrm{E}[m(Y(t);\beta_0(t))S_{\boldsymbol{X}}(\boldsymbol{X})]. \end{aligned}$$

We need to find $\Psi_\beta(y,t,\boldsymbol{x})$ such that

$$\frac{\partial \beta_0(t)}{\partial \gamma}(\gamma) = \mathrm{E}[\Psi_\beta(Y,T,\boldsymbol{X})S(Y,T,\boldsymbol{X};\gamma_0)]$$

for all regular parametric submodels.

It can be verified that

$$\Psi_\beta(y,t,\boldsymbol{x}) = -\Gamma_0^{-1}(t)\psi(y,t,\boldsymbol{x},\beta_0,\pi_0,e_0)$$

where $\psi(y,t,\boldsymbol{x},\beta_0,\pi_0,e_0) = \pi(y,\boldsymbol{x},t)m(Y;\beta_0(t)) - (\pi(y,\boldsymbol{x},t)-1)e_0(\boldsymbol{X},\beta_0(t))$ with $e_0(\boldsymbol{X},\beta(t)) = \mathrm{E}[m(Y;\beta(t))|\boldsymbol{X}]$. □

Proof of **Theorem 4**. We first verify that $\sqrt{n}\mathbb{E}[\psi(Y_i,\boldsymbol{X}_i,t,\beta_0,\pi_0,e_0)]$ converges weakly in $\ell^\infty(\mathcal{T})$. By condition **G.V**, $\psi_t = \psi(y,\boldsymbol{x},t,\beta_0,\pi_0,e_0)$ is Donsker, implying the weak convergence.

The uniform semiparametric efficiency follows from the weak convergence above and the pointwise semiparametric efficiency of Lemma 3 by Theorem 18.9 of Kosorok (2008).

Now we verify that the formula in condition **G.VI** equals the left hand-side of condition **E.**, which implies that the influence function of the two-step estimator is efficient. To this end, we begin with the formula in condition **G.VI**,

$$\begin{aligned}
&\sqrt{n}(\mathrm{E}m(Y;\beta_0(t))(\pi(\boldsymbol{U};t)-\pi_0(\boldsymbol{U};t))|_{\pi=\hat{\pi}} + \mathbb{E}m(Y;\beta_0(t))\pi_0(\boldsymbol{U};t)) \\
=&\sqrt{n}(\mathbb{E}m(Y;\beta_0(t))(\hat{\pi}(\boldsymbol{U};t)-\pi_0(\boldsymbol{U};t)) + \mathbb{E}m(Y;\beta_0(t))\pi_0(\boldsymbol{U};t)) \\
=&\sqrt{n}(\mathbb{E}m(Y;\beta_0(t))\hat{\pi}(\boldsymbol{U};t),
\end{aligned}$$

where the first equality is by condition **G.5'** which in turn is implied by **G.5'**. □

Proof of **Claim 2**. By Mean Value Theorem,

$$\pi(\boldsymbol{u};t;\hat{\vartheta}) - \pi(\boldsymbol{u};t;\vartheta_0) = \pi'(\boldsymbol{u};t;\vartheta^*)(\hat{\vartheta}-\vartheta_0),$$

where ϑ^* is a convex combination of $\hat{\vartheta}$ and ϑ_0. Therefore,

$$\begin{aligned}&|\pi(\boldsymbol{u};t;\hat{\vartheta}) - \pi(\boldsymbol{u};t;\vartheta_0)|_\infty = |\pi'(\boldsymbol{u};t;\vartheta^*)(\hat{\vartheta} - \vartheta_0)|_\infty \\ \leq &|\pi'(\boldsymbol{u};t;\vartheta^*)|_\infty||\hat{\vartheta} - \vartheta_0|| = O_p(n^{-1/2})\end{aligned}$$

since $|\pi'(\boldsymbol{u};t;\vartheta^*)|_\infty$ is bounded and $||\hat{\vartheta} - \vartheta_0|| = O_p(n^{-1/2})$. There, conditions **C.IV** and **G.IV** are verified.

For condition **C.V**, by Theorem 19.7 of van der Vaart (1998), this parametric Lipschitz continuous functional class is Donsker.

To verify the weak convergence of condition **C.VI**, we need to use the functional delta method, which involves Hadamard differentiability of a map between norm spaces. A map $\phi : \mathbb{D}_\phi \mapsto \mathbb{E}$ is Hadamard differentiable at $\theta \in \mathbb{D}$, tangentially to a set $\mathbb{D}_0$, if there exists a continuous linear map $\phi'_\theta : \mathbb{D} \mapsto \mathbb{E}$ such that

$$\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} \to \phi'_\theta(h),$$

as $n \to \infty$, for all converging sequences $t_n \to 0$ and $h_n \to h \in \mathbb{D}_0$, with $h_n \in \mathbb{D}$ and $\theta + t_n h_n \in \mathbb{D}_\phi$ for all $n \geq 1$ sufficiently large; see p. 22 of Kosorok (2008).

We first verify that η is Hadamard differentiable at ϑ tangentially to $\mathbb{R}^{d_\vartheta}$. For any $l_n \to 0$ and $h_n \to h \in \mathbb{R}^{d_\vartheta}$,

$$\begin{aligned}\frac{\eta(\vartheta + l_n h_n) - \eta(\vartheta)}{l_n} &= \frac{\pi(\boldsymbol{u};t;\vartheta + l_n h_n) - \pi(\boldsymbol{u};t;\vartheta)}{l_n} \\ &= \frac{\pi'(\boldsymbol{u};t;\vartheta^*) l_n h_n}{l_n} \to \pi'(\boldsymbol{u};t;\vartheta)h.\end{aligned}$$

Using the functional delta method, since $\pi'(\boldsymbol{u};t;\vartheta)$ is uniformly bounded,

$$\sqrt{n}(\pi(\boldsymbol{u};t;\hat{\vartheta}) - \pi(\boldsymbol{u};t;\vartheta)) \rightsquigarrow \pi'(\boldsymbol{u};t;\vartheta)Z_\vartheta \text{ in } \ell^\infty(\mathcal{U} \times \mathcal{T})$$

where $Z_\vartheta \sim N(0, \Im_\vartheta^{-1})$. □

Appendix B

Appendix B1: Proofs of the Theorems in Chapter 2, Section 3

Consistency of $\hat{\boldsymbol{\beta}}_{MD}$ under Sequential Asymptotics

We first show the consistency of the QR estimators for each individual, from which the desired result follows. Because of the non-differentiability of the criterion function of the QR estimator, we use the general theory of the M-estimator of van der Vaart (1998).

Proof of Theorem 1.1. Recall that $\boldsymbol{\theta}_{i0} = (\alpha_{i0}, \boldsymbol{\beta}_0^\top)^\top$. First, we show that under Conditions A1–A3, $\hat{\boldsymbol{\theta}}_i \xrightarrow{p} \boldsymbol{\theta}_{i0}$ as $T \to \infty$ for **each** i. Since $\hat{\boldsymbol{\theta}}_i$ is the M-estimator that maximize the criterion function

$$\boldsymbol{\theta} \mapsto \mathbb{M}_{iT}(\boldsymbol{\theta}) = -\frac{1}{T}\sum_{t=1}^{T}[\rho_\tau(y_{it} - \boldsymbol{X}_{it}^\top\boldsymbol{\theta}) - \rho_\tau(y_{it} - \boldsymbol{X}_{it}^\top\boldsymbol{\theta}_{i0})],$$

we verify the two conditions in display (5.8) of Theorem 5.7 of van der Vaart (1998). Let $M_i(\boldsymbol{\theta})$ denote $\mathrm{E}\mathbb{M}_{iT}(\boldsymbol{\theta})$, and Θ_i be a compact set that contains $\boldsymbol{\theta}_{i0}$.

The first condition of Theorem 5.7 of van der Vaart (1998), the uniform convergence of the criterion functions

$$\sup_{\boldsymbol{\theta}\in\Theta_i} |\mathbb{M}_{iT}(\boldsymbol{\theta}) - M_i(\boldsymbol{\theta})| \xrightarrow{p} 0$$

as $T \to \infty$, holds if we verify the conditions of Corollary 3.1 of Newey (1991). Note that Θ_i is a compact set. Also, the pointwise convergence of the criterion function holds by weak law of large numbers (LLN) for i.i.d. data . Note that for any fixed

$\boldsymbol{\theta} \in \Theta_i$

$$\begin{aligned}
&\mathrm{E}|\rho_\tau(y_{it} - \boldsymbol{X}_{it}^\top \boldsymbol{\theta}) - \rho_\tau(y_{it} - \boldsymbol{X}_{it}^\top \boldsymbol{\theta}_{i0})| \\
=&\mathrm{E}\left|(\boldsymbol{X}_{it}^\top(\boldsymbol{\theta} - \boldsymbol{\theta}_{i0}))\psi_\tau(y_{it} - \boldsymbol{X}_{it}^\top \boldsymbol{\theta}_{i0})\right. \\
&\left. - \int_0^{\boldsymbol{X}_{it}^\top(\boldsymbol{\theta}-\boldsymbol{\theta}_{i0})} \left(\mathbf{1}\{y_{it} - \boldsymbol{X}_{it}^\top \boldsymbol{\theta}_{i0} \le s\} - \mathbf{1}\{y_{it} - \boldsymbol{X}_{it}^\top \boldsymbol{\theta}_{i0} \le 0\}\right) ds\right| \\
\le&\mathrm{E}\left|(\boldsymbol{X}_{it}^\top(\boldsymbol{\theta} - \boldsymbol{\theta}_{i0}))\right| \\
&+ \mathrm{E}\int_0^{|\boldsymbol{X}_{it}^\top(\boldsymbol{\theta}-\boldsymbol{\theta}_{i0})|} \left(\mathbf{1}\{y_{it} - \boldsymbol{X}_{it}^\top \boldsymbol{\theta}_{i0} \le s\} + \mathbf{1}\{y_{it} - \boldsymbol{X}_{it}^\top \boldsymbol{\theta}_{i0} \le 0\}\right) ds \\
\le&3\mathrm{E}||\boldsymbol{X}_{it}||||\boldsymbol{\theta} - \boldsymbol{\theta}_{i0}|| < \infty.
\end{aligned}$$

The second line in the previous display holds by the identity of Knight (1998)

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v (\mathbf{1}\{u \le s\} - \mathbf{1}\{u \le 0\})\, ds.$$

Noting that

$$|\rho_\tau(y_{it} - \boldsymbol{X}_{it}^\top \boldsymbol{\theta}) - \rho_\tau(y_{it} - \boldsymbol{X}_{it}^\top \tilde{\boldsymbol{\theta}})| \le 3||\boldsymbol{X}_{it}||||\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}||$$

and that $\frac{1}{T}\sum_{t=1}^T \mathrm{E}||\boldsymbol{X}_{it}|| = O(1)$ which is implied by Condition 2, we finished verifying conditions of Corollary 1 of Newey (1991).

For the second condition of Theorem 5.7 of van der Vaart (1998), note that by the identity of Knight,

$$\begin{aligned}
\mathbb{M}_{iT}(\boldsymbol{\theta}) =&\frac{1}{T}\sum_{t=1}^T \left\{(\boldsymbol{X}_{it}^\top(\boldsymbol{\theta} - \boldsymbol{\theta}_{i0}))\psi_\tau(y_{it} - \boldsymbol{X}_{it}^\top \boldsymbol{\theta}_{i0})\right. \\
&\left. - \int_0^{\boldsymbol{X}_{it}^\top(\boldsymbol{\theta}-\boldsymbol{\theta}_{i0})} \left(\mathbf{1}\{y_{it} - \boldsymbol{X}_{it}^\top \boldsymbol{\theta}_{i0} \le s\} - \mathbf{1}\{y_{it} - \boldsymbol{X}_{it}^\top \boldsymbol{\theta}_{i0} \le 0\}\right) ds\right\},
\end{aligned}$$

and

$$
\begin{aligned}
M_i(\boldsymbol{\theta}) =& \mathrm{E}\Bigg\{ (\boldsymbol{X}_{it}^{\top}(\boldsymbol{\theta}-\boldsymbol{\theta}_{i0}))\psi_{\tau}(y_{it}-\boldsymbol{X}_{it}^{\top}\boldsymbol{\theta}_{i0}) \\
&- \int_0^{\boldsymbol{X}_{it}^{\top}(\boldsymbol{\theta}-\boldsymbol{\theta}_{i0})} \left(\mathbf{1}\{y_{it}-\boldsymbol{X}_{it}^{\top}\boldsymbol{\theta}_{i0} \le s\} - \mathbf{1}\{y_{it}-\boldsymbol{X}_{it}^{\top}\boldsymbol{\theta}_{i0} \le 0\}\right) ds \Bigg\} \\
=& -\mathrm{E}\left[\int_0^{\boldsymbol{X}_{it}^{\top}(\boldsymbol{\theta}-\boldsymbol{\theta}_{i0})} \left(F_i(s|\boldsymbol{X}_{it}) - \tau\right) ds \right]. \qquad \text{(A6)}
\end{aligned}
$$

Thus, by Condition A3 and expression of $M_i(\boldsymbol{\theta})$ in display (A6), we see that for every $\delta > 0$,

$$\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_{i0}|| \ge \delta} M_i(\boldsymbol{\theta}) < M_i(\boldsymbol{\theta}_{i0}) = 0.$$

Since we verified that $\hat{\boldsymbol{\theta}}_i \xrightarrow{p} \boldsymbol{\theta}_{i0}$, we obtain that $\hat{\boldsymbol{\beta}}_i \xrightarrow{p} \boldsymbol{\beta}_0$ as $T \to \infty$ for each i. Moreover, by Condition 6, $\hat{V}_i \xrightarrow{p} V_i$ for each i, it follows that for *fixed* n, as $T \to \infty$

$$\hat{\boldsymbol{\beta}}_{MD} = \left(\sum_{i=1}^{n} \hat{V}_i^{-1}\right)^{-1} \sum_{i=1}^{n} \hat{V}_i^{-1}\hat{\boldsymbol{\beta}}_i \xrightarrow{p} \left(\sum_{i=1}^{n} V_i^{-1}\right)^{-1} \sum_{i=1}^{n} V_i^{-1}\boldsymbol{\beta}_0 = \boldsymbol{\beta}_0.$$

Hence, it follows that $\hat{\boldsymbol{\beta}}_{MD} \xrightarrow{p} \boldsymbol{\beta}_0$ as $(T, n)_{seq} \to \infty$. □

Remark 4. *Strictly speaking, Condition 6 is not really necessary;* $\hat{V}_i$ *can converge to anything because the rightmost equality would hold as long as* $\hat{\boldsymbol{\beta}}_i$ *is consistent as* $T \to \infty$.

Consistency of $\hat{\boldsymbol{\beta}}_{MD}$ under Joint Asymptotics

To show Theorems 1.2 we first prove the following lemma, which provides the consistency of the QR estimators uniformly across individuals.

Lemma 4. *Under Conditions A1–A3, we have* $\max_{1\le i\le n} ||\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}|| = o_p(1)$ *as* $(n, T) \to \infty$ *and* $\frac{\log n}{T} \to 0$.

Proof. The proof of the uniform consistency of the QR estimators over i is similar to that of Theorem 1 of Kato et al. (2012).

Let $\mathbb{Q}_{iT}(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T}\rho_\tau(y_{it} - \boldsymbol{X}_{it}^\top\boldsymbol{\theta}) - T^{-1}\sum_{t=1}^{T}\rho_\tau(y_{it} - \boldsymbol{X}_{it}^\top\boldsymbol{\theta}_{i0})$, the centered objective function for each i, and $Q_i(\boldsymbol{\theta}) = \mathrm{E}\mathbb{Q}_{iT}(\boldsymbol{\theta})$. Fix any $\delta > 0$. Let $B_i(\delta) := \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_{i0}|| \leq \delta\}$, the ball with center $\boldsymbol{\theta}_{i0}$ and radius δ. For each $\boldsymbol{\theta}_i \notin B_i(\delta)$, define $\tilde{\boldsymbol{\theta}}_i = r_i\boldsymbol{\theta}_i + (1 - r_i)\boldsymbol{\theta}_{i0}$, where $r_i = \frac{\delta}{||\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i0}||}$. So $\tilde{\boldsymbol{\theta}}_i \in \partial B_i(\delta) := \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_{i0}|| = \delta\}$, the boundary of $B_i(\delta)$. Since $\mathbb{Q}_{iT}(\boldsymbol{\theta})$ is convex, and $\mathbb{Q}_{iT}(\boldsymbol{\theta}_{i0}) = 0$, we have

$$r_i\mathbb{Q}_{iT}(\boldsymbol{\theta}_i) \geq \mathbb{Q}_{iT}(\tilde{\boldsymbol{\theta}}_i) = Q_i(\tilde{\boldsymbol{\theta}}_i) + (\mathbb{Q}_{iT}(\tilde{\boldsymbol{\theta}}_i) - Q_i(\tilde{\boldsymbol{\theta}}_i)) > \epsilon_\delta + (\mathbb{Q}_{iT}(\tilde{\boldsymbol{\theta}}_i) - Q_i(\tilde{\boldsymbol{\theta}}_i)), \tag{A7}$$

the last inequality is by by the identity of Knight (1998) and Condition A3.

Thus, we have the following

$$\begin{aligned}\left\{\max_{1\leq i\leq n}||\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}|| > \delta\right\} &\overset{(a)}{\subseteq} \{\mathbb{Q}_{iT}(\boldsymbol{\theta}_i) \leq 0, \exists\, \boldsymbol{\theta}_i \notin B_i(\delta)\} \\ &\overset{(b)}{\subseteq} \cup_{i=1}^{n}\left\{\sup_{\boldsymbol{\theta}_i\in B_i(\delta)}|\mathbb{Q}_{iT}(\boldsymbol{\theta}_i) - Q_i(\boldsymbol{\theta}_i)| \geq \epsilon_\delta\right\}.\end{aligned}$$

Relation (a) holds because, by definition, $\hat{\boldsymbol{\theta}}_i$ minimizes $\mathbb{Q}_{iT}(\boldsymbol{\theta})$, and $\mathbb{Q}_{iT}(\boldsymbol{\theta}_{i0}) = 0$. Relation (b) holds by the rightmost inequality of line (A7). Then, it follows that

$$\begin{aligned}\mathrm{P}\left\{\max_{1\leq i\leq n}||\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}|| > \delta\right\} &\leq \mathrm{P}\left\{\cup_{i=1}^{n}\left\{\sup_{\boldsymbol{\theta}_i\in B_i(\delta)}|\mathbb{Q}_{iT}(\boldsymbol{\theta}_i) - Q_i(\boldsymbol{\theta}_i)| \geq \epsilon_\delta\right\}\right\} \\ &\leq \sum_{i=1}^{n}\mathrm{P}\left\{\sup_{\boldsymbol{\theta}_i\in B_i(\delta)}|\mathbb{Q}_{iT}(\boldsymbol{\theta}_i) - Q_i(\boldsymbol{\theta}_i)| \geq \epsilon_\delta\right\} \\ &\leq n\max_{1\leq i\leq n}\mathrm{P}\left\{\sup_{\boldsymbol{\theta}_i\in B_i(\delta)}|\mathbb{Q}_{iT}(\boldsymbol{\theta}_i) - Q_i(\boldsymbol{\theta}_i)| \geq \epsilon_\delta\right\}.\end{aligned}$$

Therefore, if we could show that

$$\max_{1\le i\le n} \mathrm{P}\left\{\sup_{\boldsymbol{\theta}_i\in B_i(\delta)} |\mathbb{Q}_{iT}(\boldsymbol{\theta}_i) - Q_i(\boldsymbol{\theta}_i)| \ge \epsilon_\delta\right\} = o(1/n), \tag{A8}$$

the proof of the lemma is completed.

Without loss of generality, we assume $\boldsymbol{\theta}_{i0} = \mathbf{0}$ for all i. Then the balls $B_i(\delta)$ for all i are identical, and thus we denote them by $B(\delta)$. In addition, the function $g(\boldsymbol{\theta}) := \rho_\tau(u - \boldsymbol{X}^\top\boldsymbol{\theta}) - \rho_\rho(u)$ has the following property $|g(\boldsymbol{\theta}_1) - g(\boldsymbol{\theta}_2)| \le CM||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||$ for some fixed constant $C > 0$ by the identity of Knight (1998). Let $\kappa := CM$.

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

$$\begin{aligned}
|\mathbb{Q}_{iT}(\boldsymbol{\theta}) - Q_i(\boldsymbol{\theta})| - |\mathbb{Q}_{iT}(\boldsymbol{\theta}^j) - Q_i(\boldsymbol{\theta}^j)| \le & |\mathbb{Q}_{iT}(\boldsymbol{\theta}) - Q_i(\boldsymbol{\theta}) - \mathbb{Q}_{iT}(\boldsymbol{\theta}^j) + Q_i(\boldsymbol{\theta}^j)| \\
\le & |\mathbb{Q}_{iT}(\boldsymbol{\theta}) - \mathbb{Q}_{iT}(\boldsymbol{\theta}^j)| + |Q_i(\boldsymbol{\theta}) - Q_i(\boldsymbol{\theta}^j)| \\
\le & CM\frac{\epsilon}{3\kappa} + CM\frac{\epsilon}{3\kappa} \le \frac{2\epsilon}{3}.
\end{aligned}$$

It then follows that for any $\epsilon > 0$, $\sup_{\boldsymbol{\theta}\in B(\delta)} |\mathbb{Q}_{iT}(\boldsymbol{\theta}) - Q_i(\boldsymbol{\theta})| \le \max_{1\le j\le K} |\mathbb{Q}_{iT}(\boldsymbol{\theta}^j) - Q_i(\boldsymbol{\theta}^j)| + \frac{2\epsilon}{3}$, and

$$\begin{aligned}
\mathrm{P}\left\{\sup_{\boldsymbol{\theta}\in B(\delta)} |\mathbb{Q}_{iT}(\boldsymbol{\theta}) - Q_i(\boldsymbol{\theta})| > \epsilon\right\} \le & \mathrm{P}\left\{\max_{1\le j\le K} |\mathbb{Q}_{iT}(\boldsymbol{\theta}^j) - Q_i(\boldsymbol{\theta}^j)| + \frac{2\epsilon}{3} > \epsilon\right\} \\
= & \mathrm{P}\left\{\max_{1\le j\le K} |\mathbb{Q}_{iT}(\boldsymbol{\theta}^j) - Q_i(\boldsymbol{\theta}^j)| > \frac{\epsilon}{3}\right\} \\
\le & \sum_{i=1}^{K} \mathrm{P}\left\{|\mathbb{Q}_{iT}(\boldsymbol{\theta}^j) - Q_i(\boldsymbol{\theta}^j)| > \frac{\epsilon}{3}\right\}.
\end{aligned}$$

Because data are i.i.d. within each individual, the rightmost of the inequalities is less or equal to

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

by Hoeffding's inequality. Because $\frac{\log n}{T} \to 0$, it follows that $O(\exp(-T)) = o(1/n)$.

□

With the lemma of the uniform consistency in hand, we prove Theorem 1.2.

Proof of Theorem 1.2. To show the consistency of $\hat{\boldsymbol{\beta}}_{MD}$ for joint asymptotics, we do the following computation.

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{MD} - \boldsymbol{\beta}_0 &= \left(\frac{1}{n}\sum_{i=1}^{n}\hat{V}_i^{-1}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\hat{V}_i^{-1}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0)\\ &= \left(\frac{1}{n}\sum_{i=1}^{n}\hat{V}_i^{-1}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\hat{V}_i^{-1}o_p(1) = o_p(1)\end{aligned}$$

The last equality holds because $\max_{1\le i\le n}||\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}|| = o_p(1)$ by the Lemma 4 for i.i.d. data . □

Consistency of $\hat{\boldsymbol{\beta}}_{FE}$ under Sequential Asymptotics

Note that $\hat{\boldsymbol{\beta}}_{FE}$ is essentially the estimator of the slope coefficient of the pooled panel data with individual dummy variables being added in the regression. More specifically, $\hat{\boldsymbol{\beta}}_{FE}$ can be obtained by estimating the quantile regression model

$$y = \boldsymbol{\iota}\boldsymbol{\alpha}_0 + \boldsymbol{x}\boldsymbol{\beta}_0 + u$$

where $y = (y_{11}, ..., y_{1T}, ..., y_{n1}, ..., y_{nT})^\top$, $\boldsymbol{\iota} = I \otimes (1, ..., 1)^\top$, $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)^\top$ with $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iT})$, $u = (u_{11}, ..., u_{1T}, ..., u_{n1}, ..., u_{nT})^\top$, $\boldsymbol{\alpha}_0 = (\alpha_{10}, ..., \alpha_{n0})^\top$, and $\boldsymbol{\beta}_0$ is the vector of slope coefficients.

Proof of Theorem 2.1. Following the strategy of the proof of Theorem 1.1, we first show that $(\hat{\alpha}_1, ..., \hat{\alpha}_n, \hat{\boldsymbol{\beta}}_{FE}^\top) \xrightarrow{p} (\boldsymbol{\alpha}_0^\top, \boldsymbol{\beta}_0^\top) := \boldsymbol{\theta}_0$ as $T \to \infty$. More specifically, the estimator maximizes the following criterion function

$$\boldsymbol{\theta} \mapsto \mathbb{M}_{nT}(\boldsymbol{\theta}) = -\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}(\rho_\tau(y_{it} - \alpha_i - \boldsymbol{x}_{it}^\top\boldsymbol{\beta}) - \rho_\tau(y_{it} - \alpha_{i0} - \boldsymbol{x}_{it}^\top\boldsymbol{\beta}_0)),$$

where $\boldsymbol{\theta} = (\alpha_1, ..., \alpha_n, \boldsymbol{\beta}^\top)^\top$, therefore, we verify the conditions of Theorem 5.7 of van der Vaart (1998). Let $M_n(\boldsymbol{\theta})$ denote $\mathrm{E}\mathbb{M}_{nT}(\boldsymbol{\theta})$.

Using the identity of Knight (1998), we obtain

$$\mathbb{M}_{nT}(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\Bigg\{\frac{1}{n}\sum_{i=1}^{n}(\alpha_i - \alpha_{i0} + \boldsymbol{x}_{it}^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0))\psi_\tau(y_{it} - \alpha_{i0} - \boldsymbol{x}_{it}^\top\boldsymbol{\beta}_0)$$
$$- \frac{1}{n}\sum_{i=1}^{n}\int_0^{\alpha_i - \alpha_{i0} + \boldsymbol{x}_{it}^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0)}\Big(\mathbf{1}\{y_{it} - \alpha_{i0} - \boldsymbol{x}_{it}^\top\boldsymbol{\beta}_0 \le s\} - \mathbf{1}\{y_{it} - \alpha_{i0} - \boldsymbol{x}_{it}^\top\boldsymbol{\beta}_0 \le 0\}\Big)\, ds\Bigg\},$$

and

$$M_n(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\int_0^{\alpha_i - \alpha_{i0} + \boldsymbol{x}_{it}^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0)}(F_i(s|\boldsymbol{X}_{it}) - \tau)\, ds\right]. \tag{A9}$$

First, we show the uniform convergence of the criterion functions

$$\sup_{\boldsymbol{\theta} \in \mathcal{A}^n \times \mathcal{B}} |\mathbb{M}_{nT}(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta})| \xrightarrow{p} 0$$

as $T \to \infty$. This can be done by verifying the conditions of Corollary 3.1 of Newey (1991) as in Theorem 1.1. Thus, we omit the details.

Second, using Condition A3 and expression of $M_n(\boldsymbol{\theta})$ in display (A9), we see that for every $\delta > 0$,

$$\sup_{||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \ge \delta} M_n(\boldsymbol{\theta}) < M_n(\boldsymbol{\theta}_0) = 0.$$

Since $(\hat{\alpha}_1, ..., \hat{\alpha}_n, \hat{\boldsymbol{\beta}}_{FE}^\top) \xrightarrow{p} (\boldsymbol{\alpha}_0^\top, \boldsymbol{\beta}_0^\top)$ as $T \to \infty$, it follows that $\hat{\boldsymbol{\beta}}_{FE} \xrightarrow{p} \boldsymbol{\beta}_0$ as

$T \to \infty$. Hence $\hat{\boldsymbol{\beta}}_{FE} \xrightarrow{p} \boldsymbol{\beta}_0$ as $(T, n)_{seq} \to \infty$. □

Consistency of $\hat{\boldsymbol{\beta}}_{FE}$ under Joint Asymptotics

The proof of Theorem 2.2 is given in Theorem 3.1 and Remark A1 of Kato et al. (2012).

Asymptotic Normality of $\hat{\boldsymbol{\beta}}_{MD}$ under Sequential Asymptotics

We first show that as $T \to \infty$, the quantile regression estimators converge to a random variable for each i. Then as $n \to \infty$, $\hat{\boldsymbol{\beta}}_{MD}$ converges to a normal distribution.

Proof of Theorem 3.1. First, we verify that for each individual i, Condition A1 on p. 120 of Koenker (2005) is implied by Condition A4; Conditions A2 (i)–(ii) of Koenker (2005) are implied by Conditions A4 and A5; and Condition A2 (iii) of Koenker (2005) is implied by A2. Therefore, for each individual

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, V_i).$$

We first fix n and let T tend to infinity. It then follows that

$$\begin{aligned}
\sqrt{nT}(\hat{\boldsymbol{\beta}}_{MD} - \boldsymbol{\beta}_0) &= \left(\frac{1}{n} \sum_{i=1}^{n} \hat{V}_i^{-1} \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{V}_i^{-1} \sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0) \\
&\xrightarrow{d} \left(\frac{1}{n} \sum_{i=1}^{n} V_i^{-1} \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i^{-1} N(0, V_i) \\
&= \left(\frac{1}{n} \sum_{i=1}^{n} V_i^{-1} \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} N(0, V_i^{-1})
\end{aligned}$$

The second line in the display above holds because $\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, V_i)$, $\hat{V}_i \xrightarrow{p} V_i$ for each i as $T \to \infty$ by Condition 6, and Slutsky's theorem.

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

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{MD} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, V).$$

□

Asymptotic Normality of $\hat{\boldsymbol{\beta}}_{MD}$ under Joint Asymptotics

To show this result, we provide the following auxiliary lemmas.

Lemma 5. *If A is a nonsingular matrix, and $\hat{A}_T = A + O_p(\vartheta_T)$ is an estimator of A, where $\vartheta_T \to 0$ as $T \to \infty$, then we have $\hat{A}_T^{-1} = A^{-1} + O_p(\vartheta_T)$.*

Proof. By Taylor theorem of the matrix form, we have

$$\hat{A}_T^{-1} = A^{-1} - (A^{-2})^{\top}(\hat{A}_T - A) + o_p(\hat{A}_T - A) = A^{-1} + O_p(\vartheta_T).$$

□

Lemma 6. *If $X_n = Y_n + \varpi_n$, where $\varpi_n = O_p(X_n^2)$ and $X_n \xrightarrow{p} 0$ as $n \to \infty$, then $|X_n|$ is bounded by* const. $\times |Y_n|$ *with probability approaching one.*

Proof. We need to show there exists a constant c_0, such that $\mathrm{P}(|X_n| > c_0|Y_n|) \to 0$, or $\mathrm{P}(|Y_n + \varpi_n| > c_0|Y_n|) \to 0$, since $X_n = Y_n + \varpi_n$. Noticing the events inclusion $\{|Y_n + \varpi_n| > c_0|Y_n|\} \subseteq \{|Y_n| + |\varpi_n| > c_0|Y_n|\}$, it follows that $\mathrm{P}(|Y_n + \varpi_n| > c_0|Y_n|) \leq \mathrm{P}(|Y_n| + |\varpi_n| > c_0|Y_n|)$. Therefore, it suffices to show that $\mathrm{P}(|\varpi_n| > (c_0 - 1)|Y_n|) \to 0$, or $\mathrm{P}\left(\frac{|\varpi_n|}{|X_n - \varpi_n|} > c_0 - 1\right) \to 0$. Since $\varpi_n = O_p(X_n^2)$ and $X_n \xrightarrow{p} 0$, we obtain the conclusion. □

Denote $\mathbb{S}_{iT}(\boldsymbol{\theta}) := \frac{1}{T}\sum_{t=1}^{T}\psi_\tau(y_{it} - \boldsymbol{X}_{it}^\top\boldsymbol{\theta})\boldsymbol{X}_{it}$, where $\psi_\tau(u) := \tau - \mathbf{1}\{u \leq 0\}$. Let $S_i(\boldsymbol{\theta}) := \mathrm{E}(\mathbb{S}_{iT}(\boldsymbol{\theta})) = \mathrm{E}[\tau - F_i(\boldsymbol{X}_{i1}(\boldsymbol{\theta} - \boldsymbol{\theta}_{i0})|\boldsymbol{X}_{i1})]$. It is well know that the term $\sqrt{T}([\mathbb{S}_{iT}(\hat{\boldsymbol{\theta}}_i) - \mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})] - [S_i(\hat{\boldsymbol{\theta}}_i) - S_i(\boldsymbol{\theta}_{i0})]) = o_p(1)$ for each individual since it is stochastic equicontinuous and $\hat{\boldsymbol{\theta}}_i \xrightarrow{p} \boldsymbol{\theta}_{i0}$. However, for the panel data, we need to consider the order of $\max_{1\leq i\leq n}\{[\mathbb{S}_{iT}(\hat{\boldsymbol{\theta}}_i) - \mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})] - [S_i(\hat{\boldsymbol{\theta}}_i) - S_i(\boldsymbol{\theta}_{i0})]\}$. The following two lemmas provide such an order.

Lemma 7. *If* $\max_{1\leq i\leq n}||\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}|| = O_p(\delta_n)$, *where* $\lim\limits_{n\to\infty}\delta_n = 0$, *then under Conditions A1–A5 and A6', we have* $\max_{1\leq i\leq n}\{[\mathbb{S}_{iT}(\hat{\boldsymbol{\theta}}_i) - \mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})] - [S_i(\hat{\boldsymbol{\theta}}_i) - S_i(\boldsymbol{\theta}_{i0})]\} = O_p(d_n)$, *where* $d_n = \frac{\log\delta_n}{T} \vee \sqrt{\frac{\delta_n|\log\delta_n|}{T}}$.

Proof. Without loss of generality, we assume that $\boldsymbol{\theta}_{i0} = \mathbf{0}$ for all i as in Kato et al. (2012). Let $g_{\boldsymbol{\theta}}(\boldsymbol{X}_{it}^*) = [\mathbf{1}\{u_{it} \leq 0\} - \mathbf{1}\{u_{it} - \boldsymbol{X}_{it}^\top\boldsymbol{\theta} \leq 0\}]\boldsymbol{X}_{it}$, where $\boldsymbol{X}_{it}^* = (u_{it}, \boldsymbol{X}_{it})$. It then follows that

$$
\begin{aligned}
&\mathbb{S}_{iT}(\hat{\boldsymbol{\theta}}_i) - \mathbb{S}_{iT}(\mathbf{0}) \\
=&\frac{1}{T}\sum_{t=1}^{T}(\tau - \mathbf{1}\{y_{it} - \boldsymbol{X}_{it}^\top\hat{\boldsymbol{\theta}}_i \leq 0\})\boldsymbol{X}_{it} - \frac{1}{T}\sum_{t=1}^{T}(\tau - \mathbf{1}\{y_{it} \leq 0\})\boldsymbol{X}_{it} \\
=&-\frac{1}{T}\sum_{t=1}^{T}[\mathbf{1}\{u_{it} - \boldsymbol{X}_{it}^\top\hat{\boldsymbol{\theta}}_i \leq 0\} - \mathbf{1}\{u_{it} \leq 0\}]\boldsymbol{X}_{it} \\
=&\frac{1}{T}\sum_{t=1}^{T}g_{\hat{\boldsymbol{\theta}}_i}(\boldsymbol{X}_{it}^*).
\end{aligned}
$$

So the assertion to be shown becomes $\max_{1\leq i\leq n}\sum\limits_{t=1}^{T}[g_{\hat{\boldsymbol{\theta}}_i}(\boldsymbol{X}_{it}^*) - \mathrm{E}g_{\boldsymbol{\theta}}(\boldsymbol{X}_{it}^*)|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i}] = O_p(Td_n)$. Let $\tilde{\mathcal{G}}_{i\delta_n} = \{g_{\boldsymbol{\theta}} - \mathrm{E}[g_{\boldsymbol{\theta}}(\boldsymbol{X}_{i1}^*)] : ||\boldsymbol{\theta}|| < \delta_n\}$. Thus, it suffices to show that $\max_{1\leq i\leq n}\mathrm{E}Z_i = O(d_nT)$, where $Z_i := \left|\left|\sum\limits_{t=1}^{T}g(\boldsymbol{X}_{it}^*)\right|\right|_{\tilde{\mathcal{G}}_{i\delta}}$.

To apply Proposition B.1. of Kato et al. (2012), we verify the conditions for $\tilde{\mathcal{G}}_{i\delta}$. First,it is pointwise measurable, and is bounded by $4M$. Now we provide an upper bound of the covering number for the class $\tilde{\mathcal{G}}_{i\delta}$. Note that since $\tilde{\mathcal{G}}_{i\delta}$ is a subset of $\{g - \mathrm{E}[g(\boldsymbol{X}_{it}^*)] : g \in \mathcal{G}_\infty\}$, where $\mathcal{G}_\infty = \{g_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathbb{R}^{k+1}\}$, we could instead estimate its covering number. Since $\mathcal{G}_\infty$ is a subgraph class by Lemma 2.6.15 of van der Vaart and Wellner (1996), it follows that $N(4M\epsilon, \tilde{\mathcal{G}}_\infty, L_2(Q)) \leq \left(\frac{A}{\epsilon}\right)^v$, where $A \geq 3\sqrt{e}$ and $v \geq 1$ are independent of i and n, for every $0 < \epsilon < 1$ and every probability measure Q by Theorem 2.6.7 of van der Vaart and Wellner (1996). Therefore, we obtain $N(4M\epsilon, \tilde{\mathcal{G}}_{i\delta}, L_2(Q)) \leq \left(\frac{A}{\epsilon}\right)^v$. Moreover, since $\mathrm{E}g_{\boldsymbol{\theta}}(X_{it}^*)^2 = \mathrm{E}[|F_i(\boldsymbol{X}_{it}^\top \boldsymbol{\theta}|\boldsymbol{X}_{it}) - F_i(0|\boldsymbol{X}_{it})|\boldsymbol{X}_{it}] \leq C_u M^3 \delta$, $\tilde{\mathcal{G}}_{i\delta}$ satisfies all the conditions in Proposition B.1. of Kato et al. (2012) with $U = 4M + C_u T M^3 \delta$ and $\sigma^2 = C_u M^3 \delta$. Therefore, we have $\max_{1 \leq i \leq n} \mathrm{E} Z_i = O(d_n T)$. □

Lemma 8. *Under the conditions of Lemma 7, we have* $\max_{1\leq i\leq n} ||\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}|| = O_p\left(\sqrt{\frac{\log n}{T}}\right)$.

Proof. Expanding $S_i(\hat{\boldsymbol{\theta}}_i)$ around $\boldsymbol{\theta}_{i0}$ and using Lemma 2.12 of van der Vaart (1998), we obtain

$$S_i(\hat{\boldsymbol{\theta}}_i) = S_i(\boldsymbol{\theta}_{i0}) + \frac{\partial S_i(\boldsymbol{\theta}_{i0})}{\partial \boldsymbol{\theta}_i}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}) + O_p((\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0})^2).$$

After rearranging and noting that $\widetilde{\Gamma}_i = \frac{\partial S_i(\boldsymbol{\theta}_{i0})}{\partial \boldsymbol{\theta}_i}$, we have

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0} =& \widetilde{\Gamma}_i^{-1}\left(S_i(\hat{\boldsymbol{\theta}}_i) - S_i(\boldsymbol{\theta}_{i0}) + O_p((\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0})^2)\right) \\
=& \widetilde{\Gamma}_i^{-1}\Big(-\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0}) + (S_i(\hat{\boldsymbol{\theta}}_i) - S_i(\boldsymbol{\theta}_{i0})) - (\mathbb{S}_{iT}(\hat{\boldsymbol{\theta}}_i) - \mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})) + \mathbb{S}_{iT}(\hat{\boldsymbol{\theta}}_i) \\
& + O_p((\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0})^2)\Big) \\
=& -\widetilde{\Gamma}_i^{-1}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0}) - \widetilde{\Gamma}_i^{-1}[(\mathbb{S}_{iT}(\hat{\boldsymbol{\theta}}_i) - \mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})) - (S_i(\hat{\boldsymbol{\theta}}_i) - S_i(\boldsymbol{\theta}_{i0}))] \\
& + \widetilde{\Gamma}_i^{-1}\mathbb{S}_{iT}(\hat{\boldsymbol{\theta}}_i) + \widetilde{\Gamma}_i^{-1}O_p((\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0})^2).
\end{aligned} \tag{A10}$$

Because of the computational property of quantile regression estimators, $\mathbb{S}_{iT}(\hat{\boldsymbol{\theta}}_i) = O_p\left(\frac{1}{T}\right)$ uniformly across i, and therefore $\max_{1\leq i\leq n} ||\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}||$ is bounded by

$$\text{const.} \cdot \left[\max_{1\le i\le n} ||\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})|| + \max_{1\le i\le n} ||(\mathbb{S}_{iT}(\hat{\boldsymbol{\theta}}_i) - \mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})) - (S_i(\hat{\boldsymbol{\theta}}_i) - S_i(\boldsymbol{\theta}_{i0}))|| \right] + O_p\left(\frac{1}{T}\right)$$

with probability approaching one by Lemma 6. Now we determine the order of the terms in the brackets.

First, by Hoeffding's inequality, for any fixed $K > 0$,

$$\mathrm{P}\left\{ \max_{1\le i\le n} ||\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})|| > \sqrt{\frac{\log n}{T}} K \right\} \le \sum_{i=1}^{n} \mathrm{P}\left\{ ||\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})|| > \sqrt{\frac{\log n}{T}} K \right\} \le 2n^{1-\frac{K^2}{2M^2}}.$$

Thus, $\max_{1\le i\le n} ||\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})|| = O_P\left(\sqrt{\frac{\log n}{T}}\right)$.

Next, we show that

$$\max_{1\le i\le n} ||(\mathbb{S}_{iT}(\hat{\boldsymbol{\theta}}_i) - \mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})) - (S_i(\hat{\boldsymbol{\theta}}_i) - S_i(\boldsymbol{\theta}_{i0}))|| = o_p\left(\sqrt{\frac{\log n}{T}}\right).$$

Without loss of generality, we set $\boldsymbol{\theta}_{i0} = \mathbf{0}$ for $1 \le i \le n$ as in Kato et al. (2012). So we need to show that

$$\max_{1\le i\le n} \left|\left| \frac{1}{T} \sum_{t=1}^{T} g_{\hat{\boldsymbol{\theta}}_i}(\boldsymbol{X}_{it}^*) - \mathrm{E}[g_{\boldsymbol{\theta}}(\boldsymbol{X}_{it}^*)]|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i} \right|\right| = o_p\left(\sqrt{\frac{\log n}{T}}\right),$$

which is equivalent to that for any $\epsilon > 0$,

$$\mathrm{P}\left(\max_{1\le i\le n} \left|\left| \frac{1}{T} \sum_{t=1}^{T} g_{\hat{\boldsymbol{\theta}}_i}(\boldsymbol{X}_{it}^*) - \mathrm{E}[g_{\boldsymbol{\theta}}(\boldsymbol{X}_{it}^*)]|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i} \right|\right| > \epsilon\sqrt{\frac{\log n}{T}} \right) = o(1).$$

To this end, we only need to show that

$$\max_{1\le i\le n} \mathrm{P}\left(\left|\left| \frac{1}{T} \sum_{t=1}^{T} g_{\hat{\boldsymbol{\theta}}_i}(\boldsymbol{X}_{it}^*) - \mathrm{E}[g_{\boldsymbol{\theta}}(\boldsymbol{X}_{it}^*)]|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i} \right|\right| > \epsilon\sqrt{\frac{\log n}{T}} \right) = o\left(\frac{1}{n}\right).$$

Because the QR estimators are uniformly consistent by Lemma 4, it suffices to show that for any $\epsilon > 0$, there is a sufficiently small $\delta > 0$ such that

$$\max_{1\leq i\leq n} \mathrm{P}\left(\left|\left|\sum_{t=1}^{T} g(\boldsymbol{X}_{it}^*)\right|\right|_{\tilde{\mathcal{G}}_{i\delta}} > \epsilon\sqrt{T\log n}\right) = o_p\left(\frac{1}{n}\right) \tag{A11}$$

where $\tilde{\mathcal{G}}_{i\delta} = \{g_{\boldsymbol{\theta}} - \mathrm{E}[g_{\boldsymbol{\theta}}(\boldsymbol{X}_{i1}^*)] : ||\boldsymbol{\theta}|| < \delta\}$.

By Proposition B.2. of Kato et al. (2012) and setting $s = \sqrt{2\log n}$, we have

$$\mathrm{P}\Big\{Z_i \geq \mathrm{E}Z_i + \sqrt{4\log n\{TC_f M^3\delta + 2(4M + C_f M^3\delta)\mathrm{E}Z_i\}} + \frac{2\log n}{3}(4M + C_f M^3\delta)\Big\} \leq \frac{1}{n^2},$$

where $Z_i := \left|\left|\sum_{t=1}^{T} g(\boldsymbol{X}_{it}^*)\right|\right|_{\tilde{\mathcal{G}}_{i\delta}}$. Therefore,

$$\mathrm{P}\left\{Z_i \geq \max_{1\leq i\leq n}\mathrm{E}Z_i + \sqrt{4\log n\{TC_f M^3\delta + 2(4M + C_f M^3\delta)\max_{1\leq i\leq n}\mathrm{E}Z_i\}} + \frac{2\log n}{3}(4M + C_f M^3\delta)\right\} \leq \frac{1}{n^2}.$$

Notice that in Lemma 7 we have shown that $\max_{1\leq i\leq n}\mathrm{E}Z_i \leq \text{const.} \times (\log\delta + \sqrt{T\delta|\log\delta|})$. Thus, for fixed $\epsilon > 0$, we can find δ sufficiently small and n_0 such that when $n \geq n_0$, we have

$$\max_{1\leq i\leq n}\mathrm{E}Z_i + \sqrt{4\log n\{TC_f M^3\delta + 2(4M + C_f M^3\delta)\max_{1\leq i\leq n}\mathrm{E}Z_i\}} + \frac{2\log n}{3}(4M + C_f M^3\delta) \leq \epsilon\sqrt{T\log n}.$$

So we have equation (A11), and hence we have $\max_{1\leq i\leq n}||\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}|| = O_p(\delta_n)$, where $\delta_n = \sqrt{\frac{\log n}{T}}$. □

Remark 5. *From Lemmas 7 and 8, we see that* $d_n = \frac{(\log n)^{1/4}\sqrt{|\log\delta_n|}}{T^{3/4}}$.

Proof of Theorem 3.2. Consider the following

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{MD} - \boldsymbol{\beta}_0) = \left(\frac{1}{n}\sum_{i=1}^{n}\hat{V}_i^{-1}\right)^{-1}\left(\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\hat{V}_i^{-1}\Xi\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0})\right)$$

$$= \left(\left(\frac{1}{n}\sum_{i=1}^{n}V_i^{-1}\right)^{-1} + O_p\left(T^{-1/2}h_n^{-1/2}\right)\right)$$

$$\times\left(\frac{\sqrt{n}}{n}\sum_{i=1}^{n}V_i^{-1}\Xi\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}) + O_p\left((\log n)^{1/2}n^{1/2}T^{-1/2}h_n^{-1/2}\right)\right)$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n}V_i^{-1}\right)^{-1}\frac{\sqrt{n}}{n}\sum_{i=1}^{n}V_i^{-1}\Xi\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}) + O_p\left((\log n)^{1/2}n^{1/2}T^{-1/2}h_n^{-1/2}\right).$$

By Lemmas 7 and 8, we have for each i,

$$\Xi(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_0) = -\,\Xi\widetilde{\Gamma}_i^{-1}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0}) + O_p(d_n) + O_p\left(\frac{1}{T}\right) + \Xi\widetilde{\Gamma}_i^{-1}O_p((\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0})^2)$$

$$= -\,\Xi\widetilde{\Gamma}_i^{-1}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0}) + O_p\left(d_n\right). \tag{A12}$$

Therefore, it follows that

$$\frac{\sqrt{n}}{n}\sum_{i=1}^{n}V_i^{-1}\Xi\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}) = -\frac{\sqrt{n}}{n}\sum_{i=1}^{n}V_i^{-1}\Xi\sqrt{T}\widetilde{\Gamma}_i^{-1}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0}) + O_p(\sqrt{nT}d_n).$$

The second term is $o_p(1)$ by the assumption of the relative rates of n and T in the theorem. For the first term, by Lyapunov central limit theorem, it converges in distribution to $N(0, \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}V_i^{-1})$. Hence, by Slutsky Theorem,

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{MD} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, V)$$

as $(n, T) \to \infty$ and $\sqrt{nT}d_n \to 0$. □

Asymptotic Normality of $\hat{\boldsymbol{\beta}}_{FE}$ under Sequential Asymptotics

Next we present the proof of Theorems 4.1.

Proof of Theorem 4.1. For the proof of the asymptotic normality for the sequential asymptotics, define

$$\begin{aligned}
\mathbb{H}_{iT}^{(1)}(\alpha_i, \boldsymbol{\beta}) &= \frac{1}{T}\sum_{t=1}^{T}\{\tau - \mathbf{1}(y_{it} \le \alpha_i + \boldsymbol{x}_{it}^{\top}\boldsymbol{\beta})\} \\
H_{iT}^{(1)}(\alpha_i, \boldsymbol{\beta}) &= \mathrm{E}[\mathbb{H}_{iT}^{(1)}(\alpha_i, \boldsymbol{\beta})] = \mathrm{E}[\tau - F_i(\alpha_i - \alpha_{i0} + \boldsymbol{x}_{i1}^{\top}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|\boldsymbol{x}_{it})] \\
\mathbb{H}_{nT}^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\{\tau - \mathbf{1}(y_{it} \le \alpha_i + \boldsymbol{x}_{it}^{\top}\boldsymbol{\beta})\}\boldsymbol{x}_{it} \\
H_{nT}^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \mathrm{E}\mathbb{H}_{nT}^{(2)}(\alpha_i, \boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\{\tau - F_i(\alpha_i - \alpha_{i0} + \boldsymbol{x}_{i1}^{\top}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|\boldsymbol{x}_{it})\}\boldsymbol{x}_{it}],
\end{aligned}$$

where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_n)^{\top}$. Note that the definitions listed above are similar to but different from the ones defined in the proof of Theorem 3.2 of Kato et al. (2012) in that our definitions depend on T directly.

Expanding $H_{iT}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}_{FE})$ and $H_{nT}^{(2)}(\boldsymbol{\alpha}, \hat{\boldsymbol{\beta}}_{FE})$ around $(\alpha_{i0}, \boldsymbol{\beta}_0)$, and $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$, respectively, we have

$$\begin{aligned}
H_{iT}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}_{FE}) =& - f_i(0)(\hat{\alpha}_i - \alpha_{i0}) \\
& - \boldsymbol{\gamma}_i f_i(0)(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0) + O_p((\hat{\alpha}_i - \alpha_{i0})^2 \vee ||\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0||^2) \qquad \text{(A13)} \\
H_{nT}^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{FE}) =& - \frac{1}{n}\sum_{i=1}^{n} f_i(0)\boldsymbol{\gamma}_i(\hat{\alpha}_i - \alpha_{i0}) - \left\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}f(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}^{\top}\right\}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0) \\
& + O_p((\hat{\alpha}_i - \alpha_{i0})^2 \vee ||\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0||^2) \qquad \text{(A14)}
\end{aligned}$$

Solving $\hat{\alpha}_i - \alpha_{i0}$ from equation (A13)

$$\begin{aligned}
\hat{\alpha}_i - \alpha_{i0} =& -f_i(0)^{-1}H_{iT}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}_{FE}) \\
& - \boldsymbol{\gamma}_i^{\top}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0) + O_p((\hat{\alpha}_i - \alpha_{i0})^2 \vee ||\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0||^2),
\end{aligned}$$

and plugging in equation (A14), we have

$$
\begin{aligned}
H_{nT}^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{FE}) &= \frac{1}{n}\sum_{i=1}^{n} H_{iT}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}_{FE})\boldsymbol{\gamma}_i + \frac{1}{n}\sum_{i=1}^{n} f_i(0)\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^{\top}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0) \\
&- \frac{1}{n}\sum_{i=1}^{n} \mathrm{E} f(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}^{\top}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0) + O_p((\hat{\alpha}_i - \alpha_{i0})^2 \vee ||\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0||^2) \\
=&\frac{1}{n}\sum_{i=1}^{n} H_{iT}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}_{FE})\boldsymbol{\gamma}_i + \Gamma_n(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0) + O_p((\hat{\alpha}_i - \alpha_{i0})^2 \vee ||\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0||^2)
\end{aligned}
$$

where $\Gamma_n = \frac{1}{n}\sum_{i=1}^{n}\Gamma_i$. Solving $\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0$, we obtain

$$
\begin{aligned}
&\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0 + o_p(||\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0||) = \Gamma_n^{-1}\left[H_{nT}^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{FE}) - \frac{1}{n}\sum_{i=1}^{n} H_{iT}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}_{FE})\boldsymbol{\gamma}_i\right] \\
=&\Gamma_n^{-1}\left\{-n^{-1}\sum_{i=1}^{n}\mathbb{H}_{iT}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\boldsymbol{\gamma}_i + \mathbb{H}_{nT}^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\right\} \\
&- \Gamma_n^{-1}\left[n^{-1}\sum_{i=1}^{n}\boldsymbol{\gamma}_i\{\mathbb{H}_{iT}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - H_{iT}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{iT}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\}\right] \quad \text{(A15)} \\
&+ \Gamma_n^{-1}\{\mathbb{H}_{nT}^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - H_{nT}^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - \mathbb{H}_{nT}^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\} \quad \text{(A16)} \\
&+ O_p(1/T \vee \max_{1\le i\le n}|\hat{\alpha}_i - \alpha_{i0}|^2),
\end{aligned}
$$

The second equality follows from the computational property of the QR estimator $|\mathbb{H}_{iT}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}_{FE})| = O_p(T^{-1})$ and $||\mathbb{H}_{nT}^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{FE})|| = O_p(T^{-1})$.

To derive the asymptotic normality of the FE-QR estimator, we set n to be fixed now. Note that for each fixed n, the regression quantiles are $\sqrt{T}$-consistent, therefore, we have $O_p(1/T \vee \max_{1\le i\le n}|\hat{\alpha}_i - \alpha_{i0}|^2) = O_p(1/T)$ and $o_p(||\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0||) = o_p(1/\sqrt{T})$. Using Lemmas 6–7, the third and fourth terms in the rightmost equation are ignorable, thus we have

$$
\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0 = \Gamma_n^{-1}\left\{-n^{-1}\sum_{i=1}^{n}\mathbb{H}_{iT}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0)\boldsymbol{\gamma}_i + \mathbb{H}_{nT}^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\right\} + o(1/\sqrt{T})
$$

for fixed n and $T \to \infty$. The term in the braces can be rewritten as

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\{\tau - \mathbf{1}(y_{it} \leq \alpha_i + \boldsymbol{x}_{it}^\top \boldsymbol{\beta})\}(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i).$$

As T tends to infinity, the term in the display above converges in distribution to $\frac{1}{n\sqrt{T}}\sum_{i=1}^{n} N(0, \tau(1-\tau)\mathrm{E}[(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)^\top]) = \frac{1}{n\sqrt{T}}N(0, \tau(1-\tau)\sum_{i=1}^{n}\mathrm{E}[(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)^\top])$ by Lindeberg-Levy Central Limit Theorem. Therefore, for fixed n, we have $\sqrt{T}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0) \xrightarrow{d} \frac{1}{n}N(0, \tau(1-\tau)\Gamma_n^{-1}\sum_{i=1}^{n}\mathrm{E}[(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)^\top]\Gamma_n^{-1})$ as $T \to \infty$ by Slutsky's Theorem.

Now we let n lend to infinity, and therefore,

$$\sqrt{nT}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}_0) \xrightarrow{d} N\left(0, \tau(1-\tau)\lim_{n\to\infty}\Gamma_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)(\boldsymbol{x}_{it} - \boldsymbol{\gamma}_i)^\top]\Gamma_n^{-1}\right)$$

as $n \to \infty$, as desired. □

Asymptotic Normality of $\hat{\boldsymbol{\beta}}_{FE}$ under Joint Asymptotics

For the proof of Theorems 4.2 , see the proofs of Theorem 3.2 of Kato et al. (2012).

Appendix B2: Proofs of the Theorems in Chapter 2, Section 4

Consistency and Asymptotic Normality of $\hat{\boldsymbol{\beta}}_{MD}$ under Sequential Asymptotics

Now we present the proof of Theorem 5.1.

Proof of Theorem 5.1. The proof of consistency result is very similar to that of Theorem 1.1. The only difference is the proof of the pointwise LLN for $\mathbb{M}_{iT}(\boldsymbol{\theta})$. Instead of using the LLN for i.i.d. data, we use an LLN for β-mixing data.

For weak convergence, we need to show for each individual, we have the weak convergence

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \widetilde{V}_i)$$

as $T \to \infty$. This can be seen from equation (A10) while holding n fixed. □

Consistency and Asymptotic Normality of $\hat{\boldsymbol{\beta}}_{MD}$ under Joint Asymptotics

For the proof of the consistency with β-mixing data, we need the following lemmas.

Lemma 9. *Under Conditions A2–A3 and B1, we have* $\max_{1\le i\le n} ||\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}|| = o_p(1)$ *as* $(n, T) \to \infty$ *and* $\frac{\log n}{T} \to 0$.

Proof. The proof is similar to that of Lemma 4. We only need to prove line (A8) for β-mixing data. If data are β-mixing stationary, we apply Lemma C.1 and Corollary C.1 of Kato et al. (2012). Following their strategy, let $s = 2\log n$ and $q = [\sqrt{T}]$ in Corollary C.1, for any $\epsilon > 0$ and large n (and therefore, large T),

$$\begin{aligned}
&\mathrm{P}\left\{|\mathbb{Q}_{iT}(\boldsymbol{\theta}^j) - Q_i(\boldsymbol{\theta}^j)| > \frac{\epsilon}{3}\right\} \\
\le &\mathrm{P}\left\{|\mathbb{Q}_{iT}(\boldsymbol{\theta}^j) - Q_i(\boldsymbol{\theta}^j)| > const. \left\{\sqrt{\frac{2\log n}{T}}\sigma_q(f) + \frac{2\log n}{T}[\sqrt{T}]M\right\}\right\}
\end{aligned}$$

since $\frac{\log n}{\sqrt{T}} \to 0$ and $\sigma_q(f)$ is bounded by the discussion after Lemma C.1 of Kato et al. (2012). The right hand side of the inequality above is bounded by

$$const. \times \left(n^{-2} + \sqrt{T}Ba^{[\sqrt{T}]}\right) = o(n^{-1}).$$

Since all the K terms are of $o(n^{-1})$, we have completed the proof of line (A8) for β-mixing data. □

Lemma 10. *Assume conditions A2–A5 and B1–B2 hold. For any* $c \in (0,1)$ *and* δ_N *such that* $|\log \delta_N| \asymp \log N$*, we have we have* $\max_{1\le i\le n}\{[\mathbb{S}_{iT}(\hat{\boldsymbol{\theta}}_i) - \mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})] - [S_i(\hat{\boldsymbol{\theta}}_i) - S_i(\boldsymbol{\theta}_{i0})]\} = O_p(T^{-(1-c)}(\log n) \vee T^{-1/2}\delta_n^{1/2}(\log n)^{1/2})$.

Proof. Without loss of generality, we assume that $\boldsymbol{\theta}_{i0} = \mathbf{0}$ for all i as in Kato et al. (2012). Let $g_{\boldsymbol{\theta}}(\boldsymbol{X}_{it}^*) = [\mathbf{1}\{u_{it} \le 0\} - \mathbf{1}\{u_{it} - \boldsymbol{X}_{it}^\top \boldsymbol{\theta} \le 0\}]\boldsymbol{X}_{it}$, where $\boldsymbol{X}_{it}^* = (u_{it}, \boldsymbol{X}_{it})$. We need to show $\max_{1\le i\le n} \sum_{t=1}^{T} [g_{\hat{\boldsymbol{\theta}}_i}(\boldsymbol{X}_{it}^*) - \mathrm{E} g_{\boldsymbol{\theta}}(\boldsymbol{X}_{it}^*)|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i}] = O_p(Td_n)$. Denote $\tilde{\mathcal{G}}_{i\delta_n} = \{g_{\boldsymbol{\theta}} - \mathrm{E}[g_{\boldsymbol{\theta}}(\boldsymbol{X}_{i1}^*)] : ||\boldsymbol{\theta}|| < \delta_n\}$.

To apply Proposition C.1. of Kato et al. (2012), we verify the conditions for $\tilde{\mathcal{G}}_{i\delta}$. First,it is pointwise measurable, and is bounded by $4M$. Now we provide an upper bound of the covering number for the class $\tilde{\mathcal{G}}_{i\delta}$. It can be show that $N(4M\epsilon, \tilde{\mathcal{G}}_{i,\delta_n}, L_1(Q)) \le \left(\frac{A}{\epsilon}\right)^v$, where $A \ge 5e$ and $v \ge 1$ are independent of i and n, for every $0 < \epsilon < 1$ and every probability measure Q. Setting $q = [T^c]$ and using Lemma C.1 of Kato et al. (2012), we have $\sup_{g\in\tilde{\mathcal{G}}_{i,\delta_n}} \mathrm{Var}\{\sum_{t=1}^{q} g(\boldsymbol{X}_{it}^*)/\sqrt{q}\} \le$ *const.* $\times\, \delta_n^{1/2}$. Using the condition on the dependence of the data, we obtain the conclusion. □

Lemma 11. *Assume conditions A2–A5 and B1–B2 hold. We have* $\max_{1\le i\le n} ||\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0}|| = O_p\left(\sqrt{\frac{\log n}{T}}\right)$ *for* β*-mixing data.*

Proof. The proof is similar to that of Lemma 8, except that instead of using Hoeffding inequality and Talagrand inequality, we apply a Bernstein inequality and Talagrand inequality. Setting c to be sufficiently small, we complete the proof. □

Proof of Theorem 5.2. The proof of the consistency result is similar to that of Theorem 1.2 and is an application of Lemma 9 for β-mixing data. As for the proof of the weak convergence, the only difference compared with that of Theorem 3.2 is the application of Lemmas 10–11 to obtain equation (A12). □

Consistency and Asymptotic Normality of $\hat{\boldsymbol{\beta}}_{FE}$ under Sequential Asymptotics

The argument is the same as that of the proof of the consistency part of Theorem 5.1. As for the weak convergence, all arguments are similar to the proof of the results for i.i.d. data, except that instead of using (A.10)–(A.11) of Kato et al. (2012), we use their (A.17)–(A.18) to show that terms (A15) and (A16) are negligible.

Consistency and Asymptotic Normality of $\hat{\boldsymbol{\beta}}_{FE}$ under Joint Asymptotics

The proof of Theorem 6.2 is given in Theorem 5.1 of Kato et al. (2012)

Appendix C

Appendix C1: Proofs of the Theorems in Chapter 3

Proof of Theorem 1.1. For the Swamy type test

$$
\begin{aligned}
\hat{S} =& \sum_{i=1}^{n} \left(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}\right)^{\top} \left(\frac{\hat{V}_i}{T}\right)^{-1} \left(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}\right) \\
=& \sum_{i=1}^{n} \left(\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}) - \sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right)^{\top} \hat{V}_i^{-1} \left(\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}) - \sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right) \\
=& \sum_{i=1}^{n} \sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})^{\top} \hat{V}_i^{-1} \sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})^{\top} - 2\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\top} \hat{V}_i^{-1} \sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}) \\
& + \sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \hat{V}_i^{-1} \sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
=& \sum_{i=1}^{n} \sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})^{\top} \hat{V}_i^{-1} \sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})^{\top} \\
& - 2\left[\left(\sum_{i=1}^{n} \widehat{V}_i^{-1}\right)^{-1} \sum_{i=1}^{n} \widehat{V}_i^{-1} (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})\right]^{\top} \sum_{i=1}^{n} \widehat{V}_i^{-1} \sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}) \\
& + \left[\left(\sum_{i=1}^{n} \widehat{V}_i^{-1}\right)^{-1} \sum_{i=1}^{n} \widehat{V}_i^{-1} (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})\right]^{\top} \sum_{i=1}^{n} \widehat{V}_i^{-1} \left[\left(\sum_{i=1}^{n} \widehat{V}_i^{-1}\right)^{-1} \sum_{i=1}^{n} \widehat{V}_i^{-1} (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})\right] \\
=& \sum_{i=1}^{n} \sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})^{\top} \hat{V}_i^{-1} \sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})^{\top} \\
& - \left[\sum_{i=1}^{n} \widehat{V}_i^{-1} (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})\right]^{\top} \left(\sum_{i=1}^{n} \widehat{V}_i^{-1}\right)^{-1} \sum_{i=1}^{n} \hat{V}_i^{-1} \sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}) \\
\xrightarrow{d}& \sum_{i=1}^{n} Z_i^{\top} V_i^{-1} Z_i - \sum_{i=1}^{n} Z_i^{\top} V_i^{-1} \left(\sum_{i=1}^{n} V_i^{-1}\right)^{-1} \sum_{i=1}^{n} V_i^{-1} Z_i,
\end{aligned}
$$

where Z_i are i.i.d. normal distributions with mean zero and variance V_i. The fact that $\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}) \xrightarrow{d} Z_i$ is by the standard argument as in Koenker (2005) for i.i.d. data. Thus, the asymptotic distribution of $\hat{S}$ is $\chi^2(k(n-1))$ for fixed n and T

tending to infinity. Intuitively, the degree of freedom of the χ^2 distribution should be kn instead of $k(n-1)$. However, because we are using $\hat{\boldsymbol{\beta}}$ for $\hat{S}$ rather than $\boldsymbol{\beta}$, the true value of the slope parameter, k degrees of freedom is lost. Therefore the degree of freedom of the χ^2 distribution is $k(n-1)$. For more details of the Swamy type test, see pp. 149–153 and 323–324 of Rao (1965). $\square$

Proof of Theorem 1.2. Now we consider the standardized Swamy test, $\hat{\Delta} = \sqrt{n}\frac{\frac{1}{n}\hat{S}-k}{\sqrt{2k}}$, for $(T,n)_{seq} \to \infty$. To show the result, we first fix n. From the proof of Theorem 1.1, we know that for fixed n, $\hat{S} \xrightarrow{d} \chi^2(k(n-1))$ as $T \to \infty$. By continuous mapping theorem, $\sqrt{n}\frac{\frac{1}{n}\hat{S}-k}{\sqrt{2k}} \xrightarrow{d} \sqrt{n}\frac{\frac{1}{n}\chi^2(k(n-1))-k}{\sqrt{2k}}$ as T tends to infinity.

Now we work with $\sqrt{n}\frac{\frac{1}{n}\chi^2(k(n-1))-k}{\sqrt{2k}}$ and derive the asymptotic distribution as $n \to \infty$. To this end, we transform $\sqrt{n}\frac{\frac{1}{n}\chi^2(k(n-1))-k}{\sqrt{2k}}$ as follows.

$$
\begin{aligned}
\sqrt{n}\frac{\frac{1}{n}\chi^2(k(n-1))-k}{\sqrt{2k}} &= \frac{\chi^2(k(n-1))-nk}{\sqrt{2nk}} \\
=&\frac{\chi^2(k(n-1))-k(n-1)-k}{\sqrt{2(n-1)k}}\frac{\sqrt{2(n-1)k}}{\sqrt{2nk}} \\
=&\frac{\chi^2(k(n-1))-k(n-1)}{\sqrt{2(n-1)k}}\frac{\sqrt{2(n-1)k}}{\sqrt{2nk}} - \frac{k}{\sqrt{2(n-1)k}}\frac{\sqrt{2(n-1)k}}{\sqrt{2nk}}.
\end{aligned}
$$

Using the fact that $\frac{\chi^2(\nu)-\nu}{\sqrt{2\nu}} \xrightarrow{d} N(0,1)$ as $\nu \to \infty$, we have

$$
\sqrt{n}\frac{\frac{1}{n}\chi^2(k(n-1))-k}{\sqrt{2k}} \xrightarrow{d} N(0,1),
$$

as $n \to \infty$. Hence, we conclude that $\hat{\Delta} \xrightarrow{d} N(0,1)$ as $(T,n)_{seq} \to \infty$. $\square$

Now we are ready to prove Theorem 1.3.

Proof of Theorem 1.3. Using equality (A10), Lemmas 7 and 8, we have for each i,

$$\begin{aligned}\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_{i0} =& - \Xi\widetilde{\Gamma}_i^{-1}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0}) + O_p(d_n) + O_p\left(\frac{1}{T}\right) + \Xi\widetilde{\Gamma}_i^{-1}O_p((\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_{i0})^2)\\ =& - \Xi\widetilde{\Gamma}_i^{-1}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0}) + O_p\left(d_n\right)\end{aligned}$$

Note that the remainder term $O_p(d_n)$ depends on n only. So it is a uniform version of a Bahadur representation.

Let $\hat{V}_i = \Xi\hat{\widetilde{V}}_i\Xi^\top$ and note that $\hat{V}_i = V_i + O_p\left(T^{-1/2}h_n^{-1/2}\right)$. Now we rewrite $\frac{1}{\sqrt{n}}\hat{S}$.

$$\begin{aligned}&\frac{1}{\sqrt{n}}\hat{S} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}\right)^\top\left(\frac{\hat{V}_i}{T}\right)^{-1}\left(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}\right)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\boldsymbol{\beta}}_i^\top\left(\frac{\hat{V}_i}{T}\right)^{-1}\hat{\boldsymbol{\beta}}_i\\
&-\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n}\left(\frac{\hat{V}_i}{T}\right)^{-1}\hat{\boldsymbol{\beta}}_i\right)^\top\left(\sum_{i=1}^{n}\left(\frac{\hat{V}_i}{T}\right)^{-1}\right)^{-1}\left(\sum_{i=1}^{n}\left(\frac{\hat{V}_i}{T}\right)^{-1}\hat{\boldsymbol{\beta}}_i\right)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0\right)^\top\left(\frac{\hat{V}_i}{T}\right)^{-1}\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0\right)\\
&-\frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n}\left(\frac{\hat{V}_i}{T}\right)^{-1}\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0\right)\right)^\top\left(\sum_{i=1}^{n}\left(\frac{\hat{V}_i}{T}\right)^{-1}\right)^{-1}\left(\sum_{i=1}^{n}\left(\frac{\hat{V}_i}{T}\right)^{-1}\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0\right)\right)\end{aligned}$$

For the first term, we have

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0\right)^\top\left(\frac{\hat{V}_i}{T}\right)^{-1}\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0\right)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\Xi\widetilde{\Gamma}_i^{-1}\sqrt{T}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0}) + O_p(\sqrt{T}d_n))^\top\left(V_i^{-1} + O_p\left(T^{-1/2}h_n^{-1/2}\right)\right)\\
&\times(\Xi\widetilde{\Gamma}_i^{-1}\sqrt{T}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0}) + O_p(\sqrt{T}d_n))\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\Xi\widetilde{\Gamma}_i^{-1}\sqrt{T}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0}))^\top V_i^{-1}(\Xi\widetilde{\Gamma}_i^{-1}\sqrt{T}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})) + O_p\left(\sqrt{nT}d_n \vee T^{-1/2}h_n^{-1/2}\right).\end{aligned}$$

Because $\frac{\sqrt{n}(\log n)^{1/4}\sqrt{|\log\delta_n|}}{T^{1/4}} \to 0$, the second term in the line above is $o_p(1)$.

Regarding the second term, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\left(\frac{\sum_{i=1}^n\left(V_i^{-1}+O_p\left(T^{-1/2}h_n^{-1/2}\right)\right)\left(\Xi\widetilde{\Gamma}_i^{-1}\sqrt{T}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})+O_p(\sqrt{T}d_n)\right)}{\sqrt{n}}\right)^\top\\
&\times\left(\frac{\sum_{i=1}^n\left(V_i^{-1}+O_p\left(T^{-1/2}h_n^{-1/2}\right)\right)}{n}\right)^{-1}\\
&\times\left(\frac{\sum_{i=1}^n\left(V_i^{-1}+O_p\left(T^{-1/2}h_n^{-1/2}\right)\right)\left(\Xi\widetilde{\Gamma}_i^{-1}\sqrt{T}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})+O_p(\sqrt{T}d_n)\right)}{\sqrt{n}}\right)\\
=&\frac{1}{\sqrt{n}}\left(\frac{\sum_{i=1}^n V_i^{-1}\Xi\widetilde{\Gamma}_i^{-1}\sqrt{T}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})}{\sqrt{n}}+O_p\left(T^{-1/2}h_n^{-1/2}\vee\sqrt{nT}d_n\right)\right)^\top\\
&\times\left(\left(\frac{1}{n}\sum_{i=1}^n V_i^{-1}\right)^{-1}+O_p\left(T^{-1/2}h_n^{-1/2}\right)\right)\\
&\times\left(\frac{\sum_{i=1}^n V_i^{-1}\Xi\widetilde{\Gamma}_i^{-1}\sqrt{T}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})}{\sqrt{n}}+O_p\left(T^{-1/2}h_n^{-1/2}\vee\sqrt{nT}d_n\right)\right)\\
=&\frac{1}{\sqrt{n}}\left(\frac{\sum_{i=1}^n V_i^{-1}\Xi\widetilde{\Gamma}_i^{-1}\sqrt{T}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})}{\sqrt{n}}\right)^\top\left(\frac{1}{n}\sum_{i=1}^n V_i^{-1}\right)^{-1}\\
&\times\left(\frac{\sum_{i=1}^n V_i^{-1}\Xi\widetilde{\Gamma}_i^{-1}\sqrt{T}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})}{\sqrt{n}}\right)+O_p\left(T^{-1/2}h_n^{-1/2}\vee\sqrt{T}d_n\right)\\
=&O_p\left(\frac{1}{\sqrt{n}}\vee T^{-1/2}h_n^{-1/2}\vee\sqrt{T}d_n\right)
\end{aligned}$$

Therefore, the second term is $o_p(1)$ since $\frac{\sqrt{n}(\log n)^{1/4}\sqrt{|\log\delta_n|}}{T^{1/4}}\to 0$.

Consequently we can write

$$\begin{aligned}
\frac{1}{\sqrt{n}}\hat{S}&=\frac{1}{\sqrt{n}}\sum_{i=1}^n(\Xi\widetilde{\Gamma}_i^{-1}\sqrt{T}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0}))^\top V_i^{-1}(\Xi\widetilde{\Gamma}_i^{-1}\sqrt{T}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0}))+o_p\left(1\right)\\
&:=\frac{1}{\sqrt{n}}\sum_{i=1}^n\hat{z}_i+o_p(1).
\end{aligned}$$

Thus, for the standardized Swamy test, in the case when $(T, n) \to \infty$,

$$\begin{aligned}\hat{\Delta} := \sqrt{n}\frac{\frac{1}{n}\hat{S} - k}{\sqrt{2k}} &= \frac{\frac{1}{\sqrt{n}}\hat{S} - \sqrt{n}k}{\sqrt{2k}} \\ &= \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{z}_i + o_p(1) - \sqrt{n}k}{\sqrt{2k}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\hat{z}_i - k}{\sqrt{2k}}\right) + o_p(1).\end{aligned}$$

For some small positive constant $\bar{\epsilon}$, we have $\mathrm{E}|\hat{z}_i|^{2+\bar{\epsilon}} < \bar{K} < \infty$ by condition A2. Also, $\mathrm{E}\hat{z}_i = k$ for all i. Therefore, by Lindberg-Feller Central Limit Theorem, we have

$$\hat{\Delta} \xrightarrow{d} N(0, \hat{g}^2),$$

where $\hat{g}^2 := \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathrm{Var}(\hat{z}_i)}{2k}\right)$. Finally, it just remains to show that $\hat{g}^2 = 1$, and the proof is complete.

Since $\mathrm{Var}(\hat{z}_i) = \mathrm{E}(\hat{z}_i)^2 - k^2$, we need to verify that $\mathrm{E}(\hat{z}_i)^2 = 2k + k^2 + o(1)$ uniformly across i as $T \to \infty$. Let $\boldsymbol{Z}_{it} = V_i^{-1/2}\Xi\widetilde{\Gamma}_i^{-1}\psi_\tau(y_{it} - \boldsymbol{X}_{it}^\top\boldsymbol{\theta}_{i0})\boldsymbol{X}_{it}$, we have $\mathrm{E}\boldsymbol{Z}_{it} = \mathbf{0}$, and $\mathrm{E}\boldsymbol{Z}_{it}\boldsymbol{Z}_{it}^\top = \boldsymbol{I}$. Let Z_{itl} be the lth element of $\boldsymbol{Z}_{it}$. Since Z_{itl} is linear combination of the vector $\boldsymbol{X}_{it}$, and $\psi_\tau(\cdot)$ is a uniformly bounded function, it follows that $\mathrm{E}Z_{itl}^2Z_{itm}^2$ and $\mathrm{E}Z_{itl}^4$ are both uniformly bounded for any l and m. Note that we do not have that $\boldsymbol{Z}_{it}$ follows normal distribution. Consequently, the elements in each vector $\boldsymbol{Z}_{it}$ need not be independent, although they are uncorrelated. Nonetheless,

$\boldsymbol{Z}_{it}$ are independent across i and t by Condition A1. Now we compute $\mathrm{E}(\hat{z}_i)^2$.

$$
\begin{aligned}
&\mathrm{E}(\hat{z}_i)^2 = T^2\mathrm{E}[\mathbb{S}_{iT}^\top(\boldsymbol{\theta}_{i0})\widetilde{\Gamma}_i^{-1}\Xi^\top V_i^{-1}\Xi\widetilde{\Gamma}_i^{-1}\mathbb{S}_{iT}(\boldsymbol{\theta}_{i0})]^2\\
=&\frac{1}{T^2}\mathrm{E}\left[\sum_{t=1}^T \psi_\tau(y_{it}-\boldsymbol{X}_{it}^\top\boldsymbol{\theta}_{i0})\boldsymbol{X}_{it}^\top\widetilde{\Gamma}_i^{-1}\Xi^\top V_i^{-1/2}\sum_{t=1}^T V_i^{-1/2}\Xi\widetilde{\Gamma}_i^{-1}\psi_\tau(y_{it}-\boldsymbol{X}_{it}^\top\boldsymbol{\theta}_{i0})\boldsymbol{X}_{it}\right]^2\\
=&\frac{1}{T^2}\mathrm{E}\left[\left(\sum_{t=1}^T \boldsymbol{Z}_{it}\right)^\top\sum_{t=1}^T\boldsymbol{Z}_{it}\right]^2 = \frac{1}{T^2}\mathrm{E}\left[\sum_{l=1}^k\left(\sum_{t=1}^T Z_{itl}\right)^2\right]^2\\
=&\frac{1}{T^2}\sum_{l=1}^k\mathrm{E}\left(\sum_{t=1}^T Z_{itl}\right)^4+\frac{2}{T^2}\sum_{m<l}\mathrm{E}\left(\sum_{t=1}^T Z_{itl}\right)^2\left(\sum_{t=1}^T Z_{itm}\right)^2\\
=&\frac{1}{T^2}\sum_{l=1}^k\left(\sum_{t=1}^T\mathrm{E}Z_{itl}^4+3T(T-1)\mathrm{E}Z_{i1l}^2\mathrm{E}Z_{i1l}^2\right) \qquad \text{(A17)}\\
&+\frac{2}{T^2}\sum_{m<l}\mathrm{E}\left(\sum_{t=1}^T Z_{itl}^2+2\sum_{s<t}Z_{itl}Z_{isl}\right)\left(\sum_{t=1}^T Z_{itm}^2+2\sum_{s<t}Z_{itm}Z_{ism}\right)\\
=&3k+\frac{1}{T}\sum_{l=1}^k(\mathrm{E}Z_{itl}^4-3)+\frac{2}{T^2}\sum_{m<l}\mathrm{E}\left(\sum_{t=1}^T Z_{itl}^2\sum_{t=1}^T Z_{itm}^2+2\sum_{s<t}Z_{itl}Z_{isl}\sum_{t=1}^T Z_{itm}^2\right.\\
&\left.+2\sum_{s<t}Z_{itm}Z_{ism}\sum_{t=1}^T Z_{itl}^2+4\sum_{s<t}Z_{itl}Z_{isl}\sum_{s<t}Z_{itm}Z_{ism}\right)
\end{aligned}
$$

Line (A17) follows since the expectations of all other terms from the expansion of the fourth order polynomial are zeros. And the third term equals

$$
\begin{aligned}
&\frac{2}{T^2}\sum_{m<l}\mathrm{E}\left(\sum_{t=1}^T\sum_{s=1}^T Z_{itl}^2Z_{ism}^2+2\sum_{s<t}\sum_{r=1}^T Z_{itl}Z_{isl}Z_{irm}^2+2\sum_{s<t}\sum_{r=1}^T Z_{itm}Z_{ism}Z_{irl}^2\right.\\
&\left.+4\sum_{s<t}\sum_{r<q}Z_{itl}Z_{isl}Z_{irm}Z_{iqm}\right)\\
=&\frac{2}{T^2}\sum_{m<l}\left(\sum_{t=1}^T\mathrm{E}Z_{itl}^2Z_{itm}^2+T(T-1)+4\sum_{s<t}\mathrm{E}Z_{itl}Z_{itm}Z_{isl}Z_{ism}\right) \qquad \text{(A18)}\\
=&k(k-1)+\frac{2}{T^2}\sum_{m<l}\left(\sum_{t=1}^T\mathrm{E}Z_{itl}^2Z_{itm}^2-T\right) \qquad \text{(A19)}\\
=&k(k-1)+O(1/T).
\end{aligned}
$$

Equality (A18) follows because the second and the third terms in the parentheses in the line above are zeros. Equality (A19) follows since $\mathrm{E}Z_{itl}Z_{itm}Z_{isl}Z_{ism} = \mathrm{E}Z_{itl}Z_{itm}\mathrm{E}Z_{isl}Z_{ism} = 0$. Thus, $\mathrm{E}(\hat{z}_i)^2 = 3k + k(k-1) + O(1/T) = 2k + k^2 + O(1/T)$. □

Proof of Theorem 2.1. The proof is similar to that of Theorem 1.1. The only difference is how to get $\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}) \xrightarrow{d} Z_i$. Indeed, instead of using the standard results for i.i.d. data, we apply asymptotic results for stationary β-mixing data. □

Proof of Theorem 2.2. The proof uses the proof of Theorem 2.1 and follows the same argument as that of the proof of Theorem 1.2. □

Proof of Theorem 2.3. The proof is similar to that of Theorem 1.3. Instead of applying Lemmas 7–8, we apply Lemmas 10–11 of Appendix B. □

Appendix C2: the Definitions of the Variables in the Dataset

The following is quoted from Footnote 15 of Galvao and Montes-Rojas (2010). "MDR: market debt ratio = book value of (short-term plus long-term) debt (Compustat items [9]+[34])/market value of assets (Compustat items [9]+[34]+[199]*[25]). $EBITTA$: profitability: earnings before interest and taxes
(Compustat items [18]+[15]+[16]), as a proportion of total assets (Compustat item [6]). MB: market to book ratio of assets: book liabilities plus market value of equity (Compustat items [9]+[34]+[10]+[199]*[25]) divided by book value of total assets (Compustat item [6]). $DEPTA$: depreciation (Compustat item [14]) as a proportion of total assets (Compustat item [6]). $LnTA$: log of asset size, measured in 1983 dollars (Compustat item (6)*1,000,000, deflated by the consumer price index. $FATA$: fixed asset proportion: property, plant, and equipment (Compustat

item [8)]/total assets (Compustat Item [6]). *RDDum*: dummy variable equal to one if firm did not report R&D expenses. *RDTA*: R&D expenses (Compustat item (46)) as a proportion of total assets (Compustat item [6]).”

CURRICULUM VITAE

Liang Wang

Education

- **M.S.** Economics (Distinction), City University of Hong Kong, 2008.
- **B.S.** Mathematics, Zhejiang University, China, 2007.

Working Papers

- "A New Characterization of the Normal Distribution and Test for Normality", (with Anil Bera, Antonio Galvao, and Zhijie Xiao)
- "On Testing the Equality of Mean and Quantile Effects", (with Anil Bera and Antonio Galvao)
- "Tests for Skewness and Kurtosis in the One-Way Error Components Model", (with Antonio Galvao, Gabriel Montes-Rojas, and Walter Sosa-Escudero)

Publication

- "Economic Integration of Mainland China and the Hong Kong SAR – An Analysis of Growth Attributes," (with Kui-Wai Li, Tung Liu, and Hoi Kuan Lam), *The Chinese Economy*, 44, 4, 92–114, July 2011.

Teaching Experience

- Sole Instructor:
  - Economic Statistics (undergraduate): UWM, Fall & Spring 2012, Fall 2011
  - Principles of Macroeconomics (undergraduate): UWM, Spring 2011
  - Principles of Microeconomics (undergraduate): UWM, Spring 2013
- Teaching Assistant:
  - Foundation of Econometric Methods (graduate): UWM, Spring 2011
  - Econometric Methods I (graduate): UWM, Fall 2011, Fall 2012
  - Principles of Macroeconomics (undergraduate): UWM, Spring 2011, Fall 2010
  - Economics II (undergraduate): CityU HK, Spring 2009
  - Advanced Macroeconomics (graduate): CityU HK, Fall 2008

Awards, Grants, and Honors

- UW-Milwaukee, Graduate Student Travel Award, 2011
- UW-Milwaukee, Chancellor's Graduate Student Awards, 2010
- Zhejiang University, Entrance Scholarship, 2003

Journal Referee

- Agricultural Economics