Unconditional Quantile Regression for Panel Data with Exogenous or Endogenous Regressors*

David Powell
RAND†
February 29, 2012

Abstract

Unconditional quantile treatment effects are difficult to estimate in the presence of fixed effects. Panel data are frequently used because fixed effects or differences are necessary to identify the parameters of interest. The inclusion of fixed effects or differencing of data, however, changes the interpretation of the estimates. This paper introduces a quantile estimator for panel data which use differences for identification but allows the parameters of interest to be interpreted in the same manner as cross-sectional quantile estimates. Many existing quantile panel data estimators include a separate additive term for the fixed effect. This paper includes the fixed effect in a nonseparable disturbance term. The fixed effects are never estimated and the estimator is consistent for small T. An IV version is also introduced.

Keywords: Quantiles, Panel Data, Fixed Effects, Instrumental Variables

JEL classification: C13, C31, C33, C51

---

*I am very grateful to Jerry Hausman and Whitney Newey for their guidance. I also want to thank David Autor, Tal Gross, Jon Gruber, Amanda Pallais, Jim Poterba, Nirupama Rao, and Hui Shan for their helpful comments. Special thanks to Cecilia Rouse for making her MPCP data available.

†dpowell@rand.org
1 Introduction

Many empirical applications have found quantile regression analysis useful when the variables of interest potentially have varying effects at different points in the conditional distribution of the outcome variable. While mean regression provides a valuable summary of the impact of the covariates, it does not describe the effects on different parts of the distribution. Quantile estimation, such as quantile regression (QR) introduced by Koenker and Bassett [1978], is capable of describing the effects throughout the conditional outcome distribution. Traditional quantile estimators are useful for the estimation of conditional quantile treatment effects (QTEs). We may, however, be interested in unconditional QTEs - understanding how the variables of interest impact the distribution of the outcome variable, not the distribution of the outcome variable conditional on other covariates. Panel data offer a special case where traditional methods are incapable of estimating unconditional QTEs. This paper introduces an estimation technique for panel data with fixed effects (QRFE) which allows the estimates to be interpreted in the same manner as traditional cross-sectional estimates. The estimator maintains the nonseparable disturbance property of QR. The estimator produces consistent estimates for small $T$. An IV version is also introduced (IVQRFE). The estimators are straightforward to implement using standard statistical software. The motivation of this estimator is that in many situations in empirical work, the researcher is interested in unconditional quantile treatment effects, but identification requires using “within-group” variation.

Quantile regression allows the coefficients of interests to vary based on a nonseparable disturbance term, interpreted by Doksum [1974] as “unobserved proneness.” By including additional variables - such as individual fixed effects - some of this unobserved proneness becomes observed, altering the interpretation of the coefficients. Panel data typically allow researches to identify solely off of within-group variation in the covariates or instruments.
This method allows for an arbitrary correlation between the fixed effects and the covariates or instruments. In a quantile framework, however, these fixed effects alter the interpretation of the results by including an additive fixed effect, separating the disturbance term into different components. Many quantile panel data estimators in the literature also include a separate additive term for the fixed effect, separating the fixed effect from the rest of the disturbance term. This property may be undesirable in contexts where panel data are used for identification purposes only and not to redefine the meaning of each quantile. This paper introduces a quantile estimator which uses within-group variation for identification but does not alter the interpretation of the coefficients. For the remainder of this paper, I refer to the fixed effects as “individual fixed effects” and assume that the data have multiple observations for each individual. This is done to simplify the discussion, though the estimator is also applicable in other contexts, such as repeated cross-sections where fixed effects are based on cells.

Let \( d \) denote the policy or treatment variables. With unconditional QTEs, we are interested in the distribution of \( y_{it}|d_{it} \), the outcome distribution for a given \( d_{it} \). Instead, many quantile estimators for panel data in the literature estimate the distribution of \( y_{it}|d_{it}, \alpha_i \). In this paper, I focus on linear quantiles so the model of interest is

\[
y_{it} = d'_{it} \beta(u^*_{it}), \quad u^*_{it} \sim U(0, 1),
\]

where \( u^*_{it} \) is the disturbance term, interpreted by Doksum [1974] as individual ability or proneness. An observation with a higher \( u^* \) is more prone to the outcome variable, for a given \( d \). If the outcome variable is an individual’s wage, then people with higher \( u^* \) have higher individual ability in the labor market. This paper places no functional form on the underlying components of the disturbance: \( u^*_{it} = f(u_{it}, \alpha_i) \). The disturbance is a function of the fixed effect and an observation-specific disturbance which, for comparative
purposes with the existing literature, I will consider distributed uniformly: \( u_{it} \sim U(0,1) \).

The policy variables are allowed to have an arbitrary correlation with the fixed effects, i.e.
\[
\alpha_i = h(d_{i1}, \ldots, d_{iT}, \nu_i).
\]

Traditional quantile estimators and many quantile panel data estimators in the literature are limited by assuming that both the conditional and unconditional quantiles of the disturbance term have the same distribution: \( u_{it} \sim U(0,1) \) and \( u_{it}|d_{it}, \alpha_i \sim U(0,1) \). The estimator introduced in this paper relaxes this assumption by assuming that the unconditional distribution is uniform (\( u_{it}^* \sim U(0,1) \)), but relaxing the assumptions on the conditional distributions by using within-individual comparisons: \( u_{it}^* - u_{is}^* | d_{it} - d_{is} \).

With mean regression, the disturbance does not take on such an important interpretation since distinguishing between conditional and unconditional expectations is unnecessary. The resulting estimates from OLS can be interpreted as the impact of the explanatory variables on the unconditional population mean. A key motivation of the estimation strategy introduced in this paper is to provide a quantile estimator for panel data with an equivalent property. The estimators maintain the nonseparable disturbance property implicit in cross-sectional quantile estimators. This differs from many existing quantile panel data estimators which include a separate additive term for the fixed effect. By including a separate term - such as in a location-shift model, these estimators assume that the coefficients of interest vary based only on \( u_{it} \), not \( u_{it}^* \). The interpretation of these estimators is different since the \( \tau^{th} \) percentile of \( u \) is likely different from the \( \tau^{th} \) percentile of \( u^* \).

To adopt similar terminology as Chernozhukov and Hansen [2008], the Structural Quantile Function (SQF) of interest for equation (1) is

\[
S_y(\tau|d) = d'\beta(\tau), \quad \tau \in (0,1).
\]  (2)
The SQF defines the quantile of the latent outcome variable $y_d = d' \beta(u^*)$ for a fixed $d$ and a randomly-selected $u^* \sim U(0,1)$. In other words, it describes the $\tau^{th}$ quantile of $y$ for a given $d$. Notice that once the SQFs are estimated, counterfactual distributions of the outcome variables can be generated for any given values of $d$. For known or estimated SQFs, knowledge of the distribution of $\alpha$ is unnecessary to generate this counterfactual distribution.

1.1 Motivation

Cross-sectional quantile estimators are useful for specifications such as

$$y_i = d_i' \beta(u_i^*), \quad u_i^* \sim U(0,1).$$

(3)

If $u^*|d \sim U(0,1)$, then QR can estimate the relevant SQF

$$S_y(\tau|d) = d' \beta(\tau), \quad \tau \in (0,1).$$

(4)

It may be the case that $d$ is endogenous such that $u^*|d \not\sim U(0,1)$. With mean regression, it could be possible to use panel data and condition on fixed effects for identification. Conditioning on individual fixed effects is not as straightforward with quantile estimation. Many existing panel data quantile estimators use a location-shift model where the fixed effect is held constant for all quantiles or, similarly, include a separate term for the fixed effect

$$y_{it} = \alpha_i + d' \beta(u_{it}) \quad \text{or} \quad y_{it} = \alpha_i(u_{it}) + d' \beta(u_{it}).$$

(5)

The underlying equation of interest has changed as these estimators separate the
components of the disturbance. For this reason, these estimators cannot be used for equations such as (3). The “high quantiles” refer to observations experiencing large increases in the outcome variable (relative to their fixed effect). These are not necessarily observations at the top of the cross-sectional outcome distribution. By separately including a term representing fixed ability, location shift models separate the disturbance into different components and the coefficients vary based on the non-fixed component of underlying ability. In many applications, this is not desirable. The fixed effect and disturbance are, in many contexts, related concepts. The disturbance is interpreted as underlying ability while fixed effects are frequently considered measures of fixed ability.

This paper is interested in specifications with nonseparable disturbances such as

$$y_{it} = d_{it}' \beta(u_{it}^*)$$.

(6)

This specification maintains the nonseparable disturbance in the same way as equation (3) and the SQF is still represented by equation (4). Thus, the resulting estimates can be interpreted in the same manner as cross-sectional quantile estimates. This estimator is useful in circumstances where identification of equation (4) is not possible with cross-sectional data, but the researcher does not want to alter the SQF. The distinction is that $\tau$ in equation (4) refers to the $\tau^{th}$ quantile of $u_{it}^*$. A location-shift model assumes that the SQF is $S_y(\tau | d_{it}, \alpha_i) = \alpha_i + d_{it}' \beta(\tau)$ where $\tau$ refers to the $\tau^{th}$ quantile of $u_{it}$. Location-shift models are, of course, useful in certain applications. However, there are cases where these models are undesirable. An example should illustrate the value of an unconditional quantile estimator for panel data.
1.2 Motivating Example: Vouchers and Student Achievement

Rouse [1998] studies whether receipt of a voucher in the Milwaukee Parental Choice Program (MPCP) increases the mean test score of students. The vouchers were randomly-assigned conditional on individual characteristics, which are potentially correlated with test scores. Using panel data, Rouse is able to condition on individual fixed effects to eliminate this source of bias. The impact of the vouchers at different parts of the test score distribution should also be interesting, making quantile estimation potentially useful.

Let \( \alpha_i \) represent the underlying fixed skill of student \( i \), \( T_{it} = \) test score for student \( i \) at time \( t \), \( v_{it} = \) an indicator for the receipt of a voucher. The underlying model is

\[
T_{it} = \delta_t(u_{it}^*) + v_{it} \beta(u_{it}^*).
\]  

(7)

where \( u_{it}^* \) represents a student’s underlying ability and is an unknown function of \( \alpha_i \).

The SQF is

\[
S_{T_{it}}(\tau|v_{it}) = \delta_t(\tau) + v_{it} \beta(\tau).
\]  

(8)

For illustrative purposes, assume there are only 2 time periods in the data. With mean regression, researchers would typically difference the data. Differencing, however, changes the distribution of the outcome variable. The “high-performing” students in differenced data refer to those experiencing the largest gains in test scores. Some of these students may, cross-sectionally, be in the lower part of the test distribution. Similarly, simply including individual fixed effects in a quantile regression or using a location-shift model implicitly “differences out” the individual’s placement in the distribution. A location-shift model estimates the distribution \( T_{it} - \alpha_i | v_{it} \). The QRFE estimator below allows for estimation of the distribution of \( T_{it} | v_{it} \), which parallels the interpretation of QR results. The estimator does
this while using only within-student changes in voucher receipt.

Instead of explicitly including a fixed effect or a location-shift term, the estimators below make within-person comparisons of the placement within the outcome distribution. There are two important conditions. First, \( P(y_{it} \leq d'_{it} \beta(\tau)) = \tau \), defining the unconditional distribution. Second, within-person changes in \( P(y_{it} \leq d'_{it} \beta(\tau)) \) are independent of within-person changes in \( d_{it} \). \( \alpha_i \) is never estimated or even specified. Notice that these conditions relax the condition that \( P(y_{it} \leq d'_{it} \beta(\tau)|d_{it}) = \tau \) for all \( i \). Instead, each person’s placement in the distribution is allowed to be implicitly informed by \( \alpha_i \). The motivation for relaxing this condition is that \( u^*_{it}|\alpha_i \) is likely not uniformly distributed for all \( i \) since \( u^*_{it} \) is a function of \( \alpha_i \). This implies that \( P(y_{it} \leq d'_{it} \beta(\tau)|\alpha_i) \neq \tau \).

2 Existing Literature

A small literature has extended quantile estimation to panel data. These estimation techniques tend to include separate terms for the fixed effects, either by assuming a location-shift model or allowing the fixed effects to vary by quantile. In both cases, the coefficients of interest vary based only on \( u_{it} \). These estimators typically make the assumption that \( u_{it}|\alpha_i, d_i \sim U(0, 1) \). This paper does not make such an assumption. Instead, it only assumes that the total disturbance is distributed uniformly, unconditional on fixed effects and policy variables: \( u^*_{it} \sim U(0, 1) \). The restriction made below is that within-individual changes in the policy variables, \( d_{it} - d_{is} \), are independent of within-individual changes in the disturbance term, \( u^*_{it} - u^*_{is} \).

Location-shift estimators are useful in contexts when we want to define high (low) quantiles by observations with large (small) values of \( y \) relative to their fixed level. It is important to highlight this point because the estimator in this paper is not theoretically
better than existing quantile panel data estimators without reference to a specific application. The estimator in this paper is preferable in situations where panel data are used for identification purposes only and the researcher wants to maintain the same interpretation as cross-sectional quantile estimates. In other words, QRFE is useful when the researcher would like to do a simple cross-sectional quantile regression of $y$ on $d$, but panel data are necessary for identification. QRFE estimates the relevant SQF (equation (2)) while using differences for identification. Frequently, researchers employ panel data and fixed effects models because they do not believe their model is identified cross-sectionally. However, they do not necessarily want to change the interpretation of their results.

Koenker [2004] introduces a quantile fixed effects estimator which separately estimates an additive fixed effect for the specification

$$y_{it} = \alpha_i + d_{it}'\beta(u_{it}).$$

Similarly, Harding and Lamarche [2009] introduce an IV quantile panel data estimator for the specification

$$y_{it} = \alpha_i(u_{it}) + d_{it}'\beta(u_{it}).$$

In both cases, the coefficients of interest ($\beta$) vary only with $u$, and the total disturbance term is split into its two components. These estimators make the assumptions $u_{it}|d_{it},\alpha_i \sim U(0,1)$ and $u_{it}|z_{it},\alpha_i \sim U(0,1)$, respectively. For illustrative purposes, assume that $\alpha$ is known and provided to the econometrician. The Koenker [2004] estimator is equivalent to a traditional quantile regression of $(y - \alpha)$ on $d$. The estimates cannot be interpreted in the same manner as cross-sectional estimates because the implicit SQF has changed to $S_yu(\tau|d_{it},\alpha_i) = \alpha_i + d_{it}'\beta(\tau)$ or, for the Harding and Lamarche [2009] estimator, $S_yu(\tau|d_{it},\alpha_i) = \alpha_i(\tau) + d_{it}'\beta(\tau)$ where $\tau$ relates to $u$ only. Notice that even when
$u_{it}^*|d_{it} \sim U(0, 1)$ and QR can be used, estimators with additive fixed effect terms cannot provide consistent estimates of equation (4) since it is likely that $u_{it}^*|d_{it}, \alpha_i \sim U(0, 1)$.

With the QRFE and IVQRFE estimators of this paper, the SQF for quantile $\tau$ refers to the $\tau^{th}$ quantile of $u^*$. This is the same SQF that cross-sectional quantile estimators like QR (or IV-QR estimators such as Chernozhukov and Hansen [2008]) estimate, as seen in equation (4).

Other existing quantile estimators for panel data include separate additive terms for the fixed effect too. These include Canay [2010], Galvao [2008], Ponomareva [2010], and Graham et al. [2009].

A related literature uses a correlated random effects approach for exogenous covariates. These papers impose structure on the relationship between the covariates and the fixed effects. Importantly, however, they maintain the nonseparable disturbance property of QR. Abrevaya and Dahl [2008] introduce this technique. Graham and Powell [2008] discuss a similar estimation strategy.

Rosen [2009] introduces a method which provides consistent bounds for short panels with exogenous variables and distinguishes itself from the location-shift models found in the literature. The Rosen [2009] estimator relies on the assumption $P(u_{it} \leq \tau|d_i, \alpha_i) = \tau$ for all $i$. The QRFE estimator below relaxes this assumption, replacing it with $E[P(u_{it}^* \leq \tau|d_i, \alpha_i)] = \tau$. Furthermore, the QRFE and IVQRFE estimators provide point estimates instead of bounds.

Similarly, Chernozhukov et al. [2009] discuss identification of bounds on quantile effects in nonseparable panel models where the quantiles are defined by $(\alpha_i, u_{it})$ and the variables are exogenous. They show that these bounds tighten as $T$ increases.

There is also a small literature on unconditional quantile regression. Firpo et al.
[2009] introduce an unconditional quantile regression technique for exogenous variables. Firpo [2007] and Frölich and Melly [2009] propose unconditional quantile estimators for a binary treatment variable and discuss identification. These estimators re-weight the traditional check function to get consistent estimates. These estimators are discussed further in Powell [2010]. It is unlikely that these estimators could be used with fixed effects for small $T$.

The estimator below is, to my knowledge, the first quantile panel data estimator to provide point estimates which be interpreted in the same manner as cross-sectional regression results while allowing an arbitrary correlation between the fixed effects and the policy variables. It is also one of the few IV quantile panel data estimators.

3 Model

The specification of interest is

$$y_{it} = d_{it}' \beta(u_{it}^*), \quad u_{it}^* \sim U(0, 1).$$ (11)

The motivation of this paper is that for situations where $u_{it}^* | d_{it} \not\sim U(0, 1)$, QR cannot be used. Simply including individual fixed effects in a quantile regression does not solve the problem because it assumes an additive fixed effect term. Instead, this paper maintains the nonseparable disturbance property implicit in QR. The exogeneity assumption is that within-individual changes in the policy variables do not provide information about changes in the disturbance term, $u^*$. This suggests using pairwise comparisons between observations for the same individual.
3.1 Year Fixed Effects

With panel data, it is customary to include year fixed effects. This paper assumes the inclusion of year fixed effects as exogenous policy variables to shift the outcome distribution. An individual with a large $u^*$ will make a different wage than a person with the same $u^*$ in a different year. While year fixed effects are not necessary (a constant is sufficient), the assumptions below are more plausible when they are included. Year fixed effects define the “high quantiles” as observations at the top of the cross-sectional distribution within a year. Furthermore, it implies that $k > T$ (where $k$ is the number of policy variables), which has ramifications for identification. The practical implications of including year fixed effects will be detailed during the estimation discussion.

3.2 Exogenous Policy Variables

First, some notation: let $d_i \equiv (d_{i1}, \ldots, d_{iT})$ and $\bar{d}_i \equiv \frac{1}{T} \sum_{t=1}^{T} d_{it}$.

3.2.1 Assumptions

The following conditions hold jointly with probability one:

A1 Potential Outcomes and Monotonicity: $y_{it} = d'_{it} \beta(u^*_{it})$ where $d'_{it} \beta(u^*_{it})$ is increasing in $u^*_{it} \sim U(0, 1)$.

A2 Independence: $E[1(u_{it}^* \leq \tau) - 1(u_{is}^* \leq \tau) | d_{it} - d_{is}] = 0$ for all $s, t$.

A3 Full Rank: $E[d_i]$ is rank $k$.

A4 Continuity: $y_{it}$ continuously distributed conditional on $d_i$. 
The first assumption (A1) is a standard monotonicity condition for quantile estimators (see Chernozhukov and Hansen [2008] for one example). A2 is an independence assumption. A2 could be replaced by \( P(u_{it}^* \leq \tau | d_i, \alpha_i) = P(u_{it}^* \leq \tau | \alpha_i) \) and an assumption of stationarity so that the distributions of \( u_{it}^* | \alpha_i \) and \( u_{is}^* | \alpha_i \) are the same. Instead, I use a slightly weaker assumption. The distribution of \( u_{it}^* | \alpha_i \) can change over time, but this change must be independent of \( d_{it} - d_{is} \).

A3 requires within-individual variation in the policy variables. Note that I have assumed the inclusion of year fixed effects such that \( k > T \), implying that A3 is not trivially violated. A4 is necessary for identification and typical in the context of quantile estimators.

It is important to note that no restrictions have been placed on the relationship between \( u_{it}^* \) and \( \alpha_i \). There are also no explicit restrictions on \( u_{it} \), which distinguishes this paper’s estimator from most of the quantile panel data literature. Furthermore, there are no assumptions on the relationship between \( \alpha_i \) and \( d_i \), paralleling fixed effect mean regression models.

### 3.2.2 Moment Conditions

These assumptions lead to two separate moment conditions. Both conditions will be important for identification.

**Theorem 3.1** (Moment Conditions). *Suppose A1 and A2 hold. Then for each \( \tau \in (0,1) \),

\[
E \{ (d_{it} - d_{is}) [1(y_{it} \leq d_{it}' \beta(\tau)) - 1(y_{is} \leq d_{is}' \beta(\tau))] \} = 0 \quad \text{for all } s,t, \tag{12}
\]

\[
E[1(y_{it} \leq d_{it}' \beta(\tau)) - \tau] = 0. \tag{13}
\]
Proof of (12):

\[
E \left\{ (d_{it} - d_{is}) \left[ 1(y_{it} \leq d'_{it} \beta(\tau)) - 1(y_{is} \leq d'_{is} \beta(\tau)) \right] \right\}
\]

\[
= E \left[ E \left\{ (d_{it} - d_{is}) \left[ 1(y_{it} \leq d'_{it} \beta(\tau)) - 1(y_{is} \leq d'_{is} \beta(\tau)) \right] \right| (d_{it} - d_{is}) \right]\]
\]

\[
= E \left[ E \left\{ (d_{it} - d_{is}) \left[ 1(d'_{it} \beta(u^{*}_{it}) \leq d'_{it} \beta(\tau)) - 1(d'_{is} \beta(u^{*}_{is}) \leq d'_{is} \beta(\tau)) \right] \right| (d_{it} - d_{is}) \right] \text{ by A1}
\]

\[
= E \left[ (d_{it} - d_{is}) E \left\{ 1(u^{*}_{it} \leq \tau) - 1(u^{*}_{is} \leq \tau) \right| (d_{it} - d_{is}) \right] \text{ by A1}
\]

\[
= 0 \text{ by A2}
\]

Proof of (13):

\[
E[1(y_{it} \leq d'_{it} \beta(\tau))] = E[1(d'_{it} \beta(u^{*}_{it}) \leq d'_{it} \beta(\tau))] \text{ by A1}
\]

\[
= P[u^{*}_{it} \leq \tau] \text{ by A1}
\]

\[
= \tau \text{ by A1}
\]

Equation (12) is a useful formulation since it shows that the estimator is simply a series of within-individual comparisons. However, it is also useful to consider equivalent conditions. Specifically, equation (12) can be replaced by

\[
E \left\{ \frac{1}{T} \sum_{t=1}^{T} d_{it} \left[ 1(y_{it} \leq d'_{it} \beta(\tau)) - \frac{1}{T} \sum_{s=1}^{T} 1(y_{is} \leq d'_{is} \beta(\tau)) \right] \right\} = 0
\]

Estimation details will be discussed below, but the corresponding sample moments are

14
Sample Moment 1

\[ g_{it}(b) = \frac{1}{T} \sum_{t=1}^{T} d_{it} \left[ 1(y_{it} \leq d_{it}'b) - \frac{1}{T} \sum_{s=1}^{T} 1(y_{is} \leq d_{is}'b) \right], \]

Sample Moment 2

\[ h(b) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} 1(y_{it} \leq d_{it}'b) - \tau. \]

Sample Moment 1 is worth discussing further. Define \( \tau_{i}(b) \equiv \frac{1}{T} \sum_{s=1}^{T} 1(y_{is} \leq d_{is}'b). \) Note that the moment condition is similar to the cross-sectional quantile moment condition where \( \tau \) is replaced by \( \tau_{i}: \)

\[ g_{it}(b) = d_{it} \left[ 1(y_{it} \leq d_{it}'b) - \tau_{i} \right]. \]

This makes intuitive sense. The individual fixed effect provides information about the distribution of the disturbance. Thus, instead of assuming that \( \tau_{i} = \tau \) for each individual, panel data allow us to relax this assumption. \( \tau_{i} \) varies by individual with \( E[\tau_{i}] = \tau. \)

For identification and other properties, it is easiest to use the following equivalent formulation

\[ g_{i}(b) = \frac{1}{T} \left\{ \sum_{t=1}^{T} \left( d_{it} - \overline{d}_{i} \right) \left[ 1(y_{it} \leq d_{it}'b) \right] \right\}. \tag{14} \]

Sample Moment 2 relies on the fact that the unconditional distribution of \( u^{*} \) is \( U(0,1). \) Notice that this sample moment also holds with traditional quantile estimators such as QR. With QR, one assumes both that \( u^{*} \sim U(0,1) \) and \( u^{*}|d \sim U(0,1). \) The QRFE estimator does not assume \( u^{*}|d \sim U(0,1), \) but replaces it with a weaker assumption. This is the gain from employing panel data.
3.2.3 Estimation

Estimation uses Generalized Method of Moments (GMM). Sample moments are defined by

\[ \hat{g}(b) = \frac{1}{N} \sum_{i=1}^{N} g_i(b). \] (15)

It is necessary to use Sample Moment 2 as well. Sample Moment 2 constrains \( b \) to ensure that the estimates refer to the \( \tau^{th} \) quantile. Define

\[ B \equiv \left\{ b \mid \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} 1(y_{it} \leq d'_{it} b) = \tau \right\}. \]

Then,

\[ \hat{\beta}(\tau) = \arg \min_{b \in B} \hat{g}(b)' \hat{A} \hat{g}(b) \] (16)

for some weighting matrix \( \hat{A} \).

There is a straightforward way to confine all guesses \( b \) to the set \( B \), but it is first helpful to discuss year fixed effects.

Year Fixed Effects

Moment Condition 1 (equation (12)) represents \( k \) separate conditions. The inclusion of year fixed effects implies

\[ P(y_{it} \leq d'_{it} \beta(\tau)) = P(y_{is} \leq d'_{is} \beta(\tau)) \quad \text{for all } s, t. \]

Equation (13), then, implies

\[ P(y_{it} \leq d'_{it} \beta(\tau)) = \tau \quad \text{for all } t. \] (17)
By assuming the inclusion of year fixed effects, I can use equation (17) for the sample moments. Let \( \mathbf{d} \equiv (\mathbf{x}, \gamma_t) \) where \( \mathbf{x} \) are the variables of interest and \( \gamma_t \) is a set of year fixed effects. Let \( \tilde{b} \) be the coefficients on \( \mathbf{x} \) so that \( \mathbf{d}'_it \tilde{b} = \gamma_t + \mathbf{x}'_it \tilde{b} \). The sample moments can be replaced by

**Sample Moment 1’**

\[
g_i(b) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_{it} \left[ 1(y_{it} \leq \mathbf{d}'_it b) - \frac{1}{T} \sum_{s=1}^{T} 1(y_{is} \leq \mathbf{d}'_is b) \right],
\]

\( (18) \)

**Sample Moment 2’**

\[
h_t(b) = \frac{1}{N} \sum_{i=1}^{N} 1(y_{it} \leq \mathbf{d}'_it b) - \tau \quad \text{for all } t.
\]

\( (19) \)

Sample Moment 2’ defines the year fixed effects. The value of these fixed effects forces \( y_{it} \leq \mathbf{d}'_it \tilde{b} \) to hold for \( 100\tau \%) \) of the observations in each year. The benefit of this approach is that it reduces the number of parameters that need to be estimated and offers a simple way to enforce the second sample moment by defining

\[
\mathcal{B} \equiv \left\{ b \mid \frac{1}{N} \sum_{i=1}^{N} 1(y_{it} \leq \mathbf{d}'_it b) = \tau \quad \text{for all } t \right\}.
\]

Define \( \gamma_t(\tau, \tilde{b}) \) as the \( \tau^{th} \) quantile of the distribution of \( y_{it} - \mathbf{x}'_it \tilde{b} \) in year \( t \):

\[
\hat{\gamma}_t(\tau, \tilde{b}) \quad \text{solves} \quad \frac{1}{N} \sum_{i} 1(y_{it} - \mathbf{x}'_it \tilde{b} \leq \hat{\gamma}_t(\tau, \tilde{b})) = \tau.
\]

\( (20) \)

This equation forces \( h_t(b) = 0 \) to hold for all \( t \), confining all guesses to \( \mathcal{B} \). In words, for any “guess” \( \tilde{b} \), the optimal values \( \hat{\gamma}_t(\tau, \tilde{b}) \) are known. This simplifies the estimation process. The steps are the following:
1. Choose $\tilde{b}$.

2. Calculate the year fixed effects using equation (20).

3. Evaluate $\hat{g}(b)\hat{A}g(b)$ where $g_i(b)$ is defined by equation (18).

The $b$ that minimizes this condition is $\hat{\beta}(\tau)$. In many economic applications, it is typical to have only one or two treatment variables (not counting the year fixed effects). In these cases, grid-searching is appropriate. Chernozhukov and Hansen [2008] make this same argument and recommend grid-searching. With more treatment variables, other optimization methods are likely necessary, such as Markov Chain Monte Carlo (see Chernozhukov and Hong [2003] for details).

3.2.4 Identification

Identification of unconditional QTEs is discussed extensively in Powell [2010]. This section includes a brief discussion of identification in the panel data case. Identification requires

$$\left( E[g_i(\tilde{\beta})] = 0, E[h(\tilde{\beta})] = 0 \right) \iff \tilde{\beta} = \beta(\tau) \text{ for } \tau \in (0, 1).$$

I use the equation (14) formulation in this section.

**Theorem 3.2** (Identification). If (i) $A1$-$A4$ hold; (ii) $E \left\{ \frac{1}{T} \sum_{t=1}^{T} (d_{it} - \bar{d}_i) \left[ 1(y_{it} \leq d'_{it}\tilde{\beta}) \right] \right\} = 0$; (iii) $E \left\{ 1(y_{it} \leq d'_{it}\tilde{\beta}) \right\} = \tau$, then $\tilde{\beta} = \beta(\tau)$.

**Proof.** Start with (ii): $E \left\{ \frac{1}{T} \sum_{t=1}^{T} (d_{it} - \bar{d}_i) \left[ 1(y_{it} \leq d'_{it}\tilde{\beta}) \right] \right\} = 0$. Using the Law of Iterated Expectations, we have $E \left\{ \frac{1}{T} \sum_{t=1}^{T} (d_{it} - \bar{d}_i) E \left[ 1(y_{it} \leq d'_{it}\tilde{\beta}) | d_i \right] \right\} = 0$.

Without loss of generality, assume that $P(y_{it} \leq d'_{it}\tilde{\beta}|d_i) = P(y_{it} \leq d'_{it}\beta(\tilde{\tau})|d_i)$ for some $\tilde{\tau} \in (0, 1)$. Under $A3$, $P(y_{is} \leq d'_{is}\tilde{\beta}|d_i) = P(y_{is} \leq d'_{is}\beta(\tilde{\tau})|d_i)$ for all $s$.

By $A4$, we know that $d'_{it}\tilde{\beta} = d'_{it}\beta(\tilde{\tau})$ for all $t$. $A3$ implies $\tilde{\beta} = \beta(\tilde{\tau})$. 

18
Because of (iii), we know that \( \tilde{\tau} = \tau \), implying that \( \tilde{\beta} = \beta(\tau) \).

### 3.3 Endogenous Policy Variables

Even after conditioning on individual fixed effects, the policy variables may be endogenous. In this section, I consider estimation of unconditional QTEs for endogenous policy variables. I assume the existence of instruments which are exogenous conditional on individual fixed effects. Identification requires that the instruments impact the entire distribution of the policy variables. For this section, assume that both the policy variables and instruments are discrete. The policy vector has \( m \) possible values. A brief discussion of continuous variables is included in Powell [2010].

Let \( \tilde{z}_i \equiv (z_{i1} - \bar{z}_i, \ldots, z_{iT} - \bar{z}_i) \). Define \( \Pi_i \) as the relationship between \( z \) and \( d \),

\[
\Pi_i \equiv \begin{bmatrix}
P(d_{i1} = d^{(1)}|z_{i1}) & \cdots & P(d_{i1} = d^{(m)}|z_{i1}) \\
\vdots & \ddots & \vdots \\
P(d_{iT} = d^{(1)}|z_{iT}) & \cdots & P(d_{iT} = d^{(m)}|z_{iT})
\end{bmatrix}.
\]

Define \( D \) as a matrix of all possible values for \( d \),

\[
D \equiv \begin{bmatrix}
d^{(1)r} \\
\vdots \\
d^{(m)r}
\end{bmatrix}.
\]

Finally,

\[
\Gamma_i \equiv \begin{bmatrix}
P(y_{it} \leq d^{(1)r}\beta(\tau)|z_i) \\
\vdots \\
P(y_{it} \leq d^{(m)r}\beta(\tau)|z_i)
\end{bmatrix}.
\]
3.3.1 Assumptions

The following conditions hold jointly with probability one:

**IV-A1** Potential Outcomes and Monotonicity: \( y_{it} = d'_{it} \beta(u_{it}^*) \) where \( d'_{it} \beta(u_{it}^*) \) is increasing in \( u_{it}^* \sim U(0,1) \).

**IV-A2** Independence: \( E[1(u_{it}^* \leq \tau) - 1(u_{is}^* \leq \tau)|z_{it} - z_{is}] = 0 \) for all \( s, t \).

**IV-A3** Full Rank: \( D \) is rank \( k \).

**IV-A4** First Stage: \( E[\tilde{z}_i \Pi_i] \) is rank \( m \).

**IV-A5** Continuity: \( y_{it} \) continuously distributed conditional on \( z_i \).

The first stage assumption is stronger than the typical mean-IV assumption. The instruments must impact the entire distribution of the policy variables. Note that **IV-A4** is stronger than necessary as it assumes that there are more instruments than possible values for \( d \). With discrete variables, this is possible by creating dummy variables for each possible value of \( z \). However, this may not be necessary. The above conditions are similar to those found in Powell [2010] which establishes nonparametric identification. These conditions can be relaxed with linear quantiles.

Instead, say that there exists a subset of \( d \)

\[
\hat{D} \equiv \begin{bmatrix} d^{(j_1)} \, \vdots \, d^{(j_s)} \end{bmatrix},
\]

where \( s \leq m \). Also define
\[
\Pi_i \equiv \begin{bmatrix}
P(d_{i1} = d^{(j_1)}|z_{i1}) & \cdots & P(d_{i1} = d^{(j_s)}|z_{i1}) \\
\vdots & \ddots & \vdots \\
P(d_{iT} = d^{(j_1)}|z_{iT}) & \cdots & P(d_{iT} = d^{(j_s)}|z_{iT})
\end{bmatrix}.
\]

Identification will hold as long as there exists a subset of \(d\) such that (i) \(\hat{D}\) is full rank and (ii) \(E[\tilde{z}_i\Pi_i]\) is rank \(s\). In words, only a minimum number of possible values of the policy variables need to be identified.

### 3.3.2 Moment Conditions

The moment conditions are similar:

**Theorem 3.3** (Moment Conditions). Suppose \(IV-A1\) and \(IV-A2\) hold. Then for each \(\tau \in (0,1)\),

\[
E\left\{ (z_{it} - z_{is}) \left[ 1(y_{it} \leq d_{it}' \beta(\tau)) - 1(y_{is} \leq d_{is}' \beta(\tau)) \right] \right\} = 0 \text{ for all } s, t, \tag{21}
\]

\[
E[1(y_{it} \leq d_{it}' \beta(\tau)) - \tau] = 0. \tag{22}
\]

Notice that equation (22) is exactly the same as the exogenous case. With equation (21), \((d_{it} - d_{is})\) has simply been replaced by \((z_{it} - z_{is})\). The sample moments are also similar.

**IV Sample Moment 1**

\[
g_t(b) = \frac{1}{T} \sum_{t=1}^{T} z_{it} \left[ 1(y_{it} \leq d_{it}' b) - \frac{1}{T} \sum_{s=1}^{T} 1(y_{is} \leq d_{is}' b) \right].
\]
IV Sample Moment 2

\[ h(b) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} 1(y_{it} \leq d_{it}'b) - \tau. \]

With year fixed effects, we can replace the sample moments as before to limit the number of parameters. It is easier to discuss estimation properties with the following formulation

\[ g_i(b) = \frac{1}{T} \left\{ \sum_{t=1}^{T} (z_{it} - z_i) \left[ 1(y_{it} \leq d_{it}'b) \right] \right\}. \] (23)

Estimation follows as before.

3.3.3 Identification

An extensive discussion of the conditions necessary for identification of unconditional QTEs with endogenous policy variables is included in Powell [2010].

**Theorem 3.4** (Identification). If (i) IV-A1 - IV-A5 hold; (ii) \(E\left[ \frac{1}{T} \sum_{t=1}^{T} (z_{it} - z_i) \left[ 1(y_{it} \leq d_{it}'\tilde{\beta}) \right] \right] = 0\); (iii) \(E\left[ 1(y_{it} \leq d_{it}'\tilde{\beta}) \right] = \tau\), then \(\tilde{\beta} = \beta(\tau)\).

**Proof.** Starting with (ii), we have \(E[z_i\Pi_i\tilde{\Gamma}_i] = 0\).

By IV-A4, \(\tilde{\Gamma}_i = \tilde{\Gamma}_i\) for some \(\tilde{\tau} \in (0, 1)\).

By IV-A5, we know that \(d^{(j)'}\tilde{\beta} = d^{(j)'}\beta(\tilde{\tau})\) for all \(j\). IV-A3 implies \(\tilde{\beta} = \beta(\tilde{\tau})\).

Because of (iii), we know that \(\tilde{\tau} = \tau\), implying that \(\tilde{\beta} = \beta(\tau)\). \(\square\)
4 Properties

This section discusses consistency and asymptotic normality of the estimator. These properties are discussed for small $T$ as $N \to \infty$. I use the IV notation, where it is possible that $z = d$. Some additional assumptions are necessary:

**IV-A6** $(y_i, d_i, z_i)$ i.i.d.

**IV-A7** $\mathcal{B}$ is compact.

**IV-A8** $\left\| \frac{1}{T} \sum_{t=1}^{T} (z_{it} - \bar{z}_i) \right\|^{2+\delta} < \infty$ for some $\delta > 0$.

**IV-A9** $G \equiv E \left[ \frac{1}{T} \sum_{t=1}^{T} (z_{it} - z_i) d'_i f_y(d'_i \beta(\tau)|z_i) \right]$ exists such that $G'AG$ nonsingular.

4.1 Consistency

**Theorem 4.1** (Consistency). If **IV-A1 - IV-A8** hold and $\hat{A} \xrightarrow{p} A$ positive definite, then $\hat{\beta}(\tau) \xrightarrow{p} \beta(\tau)$.

The sample moments functions are discontinuous, but consistency can still be proven by relying on continuity of the expectation of the sample moments (see Lemma 2.4 in Newey and McFadden [1994]). Consistency follows from Theorem 2.6 of Newey and McFadden [1994] because these conditions are met:

1. Theorem 3.4 above proves identification.

2. Compactness of $\mathcal{B}$ holds by assumption **IV-A7**.

3. $g_i(b)$ is continuous at each $b$ with probability one under **IV-A5**.

4. $\|g_i(b)\| \leq \left\| \frac{1}{T} \sum_{t=1}^{T} (z_{it} - \bar{z}_i) \right\| < \infty$. 

23
4.2 Asymptotic Normality

The conditions for asymptotic normality are more difficult with discontinuous sample moments. Newey and McFadden [1994] discuss asymptotic normality results for discontinuous moment conditions. Stochastic equicontinuity is an important condition for these results and follows here from the fact that the functional class \( \{1(y_{it} \leq d_{it}b), b \in \mathbb{R}^k\} \) is Donsker and the Donsker property is preserved when the class is multiplied by a bounded random variable. Thus,

\[
\left\{ \frac{1}{T} \sum_{t=1}^{T} (z_{it} - \bar{z}_i) \left[ 1(y_{it} \leq d_{it}b) \right], b \in \mathbb{R}^k \right\}
\]

is Donsker with envelope \( 2 \max_{i,t} |z_{it} - \bar{z}_i| \). Stochastic equicontinuity follows from Theorem 1 in Andrews [1994]. Define \( \Sigma \equiv E[g(\beta(\tau))g(\beta(\tau))'] \).

**Theorem 4.2 (Asymptotic Normality).** If \textbf{IV-A1} - \textbf{IV-A9} hold and \( \hat{A} \xrightarrow{p} A \) positive definite, then \( \sqrt{N}(\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{d} N[0, (G'AG)^{-1}G'\Sigma AG(G'AG)^{-1}] \).

The appendix includes a more extensive discussion of this theorem.

4.2.1 Inference

It is well-known that there are difficulties in estimating the variance of quantile estimators. Let \( \mu_{it} \equiv y_{it} - d_{it}'\beta(\tau) \). With QR, is it common to make the assumption \( f_\mu(0|x_i) = f_\mu(0) \). The equivalent assumption here \( (f_\mu(0|z_i) = f_\mu(0)) \) is difficult since a main motivation of this paper is that \( z_i \) provides information about the value of \( \mu \). It is possible to use the histogram estimation technique suggested in Powell [1986] to obtain consistent estimates of \( G \).

\[
\hat{G} = \frac{1}{2Nh} \sum_{i=1}^{N} \left[ \frac{1}{T} \sum_{t=1}^{T} (z_{it} - \bar{z}_i)d_{it}1 \left( y_{it} - d_{it}'\hat{\beta}(\tau) \leq h \right) \right].
\]

Consistent estimates of \( \Sigma \) and \( A \) are possible by plugging in \( \hat{\beta}(\tau) \). Because of
the concerns with estimation of the variance for quantile estimators, a bootstrap method (resampling with replacement) is typically recommended. Hahn [1996] justifies the bootstrap in this context. Notice that the sample moments have been defined by the individual. Thus, bootstrapping requires sampling based on individual (not individual-year). This technique accounts for within-individual clustering.

5 Applications

5.1 Simulations

To illustrate the usefulness of the QRFE estimator, I generate the following data:

\[ t \in \{0, 1\} \]

Fixed Effect: \( \alpha_i \sim U(0, 1) \)

\[ u_{it} \sim U(0, 1) \]

Total Disturbance: \( u_{it}^* \equiv F_{\alpha+u}(\alpha_i + u_{it}) \Rightarrow u_{it}^* \sim U(0, 1) \)

Year Effect: \( \delta_0 = 1, \delta_1 = 2 \)

\[ \psi_{it} \sim U(0, 1) \]

Policy Variable: \( d_{it} = \alpha_i + \psi_{it} \)

Outcome: \( y_{it} = u_{it}^*(\delta_t + d_{it}) \)

Note that \( d \) is exogenous conditional on \( \alpha \). The impact of \( d \) is a function of \( \alpha + u \) and varies by observation. Consequently, the coefficient varies by quantile: \( \beta(\tau) = \tau \). Year fixed effects are also crucial as the distribution changes (differentially) across years. I generate these data for \( N = 500, T = 2 \). Grid-searching is used to minimize the GMM objective
function. Table 1 presents the results of the simulation for the coefficient of interest. To illustrate that these data require conditioning on individual fixed effects, I show results for both QR (left) and the estimator of this paper, QRFE (right).

The simulated data offer a difficult test since the effect of \( d \) changes continuously throughout the distribution. Even under these circumstances, the QRFE estimator of this paper performs well. Note that the QR estimator, as expected, performs poorly.

Table 1: QRFE Simulation (N=500, T=2)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Mean Bias</th>
<th>MAD</th>
<th>RMSE</th>
<th>Mean Bias</th>
<th>MAD</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.36712</td>
<td>0.36556</td>
<td>0.13790</td>
<td>-0.00008</td>
<td>0.04</td>
<td>0.00288</td>
</tr>
<tr>
<td>10</td>
<td>0.44862</td>
<td>0.44591</td>
<td>0.20238</td>
<td>0.00203</td>
<td>0.05</td>
<td>0.00518</td>
</tr>
<tr>
<td>15</td>
<td>0.50631</td>
<td>0.50135</td>
<td>0.25263</td>
<td>0.00284</td>
<td>0.06</td>
<td>0.00737</td>
</tr>
<tr>
<td>20</td>
<td>0.54408</td>
<td>0.54276</td>
<td>0.29835</td>
<td>-0.00314</td>
<td>0.07</td>
<td>0.01045</td>
</tr>
<tr>
<td>25</td>
<td>0.58467</td>
<td>0.58609</td>
<td>0.34433</td>
<td>-0.00245</td>
<td>0.08</td>
<td>0.01337</td>
</tr>
<tr>
<td>30</td>
<td>0.62180</td>
<td>0.62133</td>
<td>0.38924</td>
<td>-0.00207</td>
<td>0.08</td>
<td>0.01577</td>
</tr>
<tr>
<td>35</td>
<td>0.65418</td>
<td>0.65448</td>
<td>0.43067</td>
<td>-0.00186</td>
<td>0.10</td>
<td>0.01871</td>
</tr>
<tr>
<td>40</td>
<td>0.68126</td>
<td>0.68146</td>
<td>0.46706</td>
<td>0.00142</td>
<td>0.10</td>
<td>0.02036</td>
</tr>
<tr>
<td>45</td>
<td>0.69951</td>
<td>0.69866</td>
<td>0.49237</td>
<td>0.00281</td>
<td>0.10</td>
<td>0.02108</td>
</tr>
<tr>
<td>50</td>
<td>0.71083</td>
<td>0.71043</td>
<td>0.50842</td>
<td>0.00030</td>
<td>0.11</td>
<td>0.02249</td>
</tr>
<tr>
<td>55</td>
<td>0.71442</td>
<td>0.71505</td>
<td>0.51359</td>
<td>-0.00007</td>
<td>0.10</td>
<td>0.02230</td>
</tr>
<tr>
<td>60</td>
<td>0.70769</td>
<td>0.70789</td>
<td>0.50395</td>
<td>0.00208</td>
<td>0.10</td>
<td>0.02188</td>
</tr>
<tr>
<td>65</td>
<td>0.69264</td>
<td>0.69326</td>
<td>0.48291</td>
<td>0.00150</td>
<td>0.10</td>
<td>0.02022</td>
</tr>
<tr>
<td>70</td>
<td>0.66441</td>
<td>0.66610</td>
<td>0.44453</td>
<td>0.00274</td>
<td>0.09</td>
<td>0.01909</td>
</tr>
<tr>
<td>75</td>
<td>0.62463</td>
<td>0.62473</td>
<td>0.39291</td>
<td>0.00204</td>
<td>0.09</td>
<td>0.01731</td>
</tr>
<tr>
<td>80</td>
<td>0.57263</td>
<td>0.57329</td>
<td>0.33008</td>
<td>-0.00059</td>
<td>0.09</td>
<td>0.01617</td>
</tr>
<tr>
<td>85</td>
<td>0.50626</td>
<td>0.50624</td>
<td>0.25792</td>
<td>0.00058</td>
<td>0.08</td>
<td>0.01352</td>
</tr>
<tr>
<td>90</td>
<td>0.41310</td>
<td>0.41338</td>
<td>0.17197</td>
<td>0.00185</td>
<td>0.07</td>
<td>0.01060</td>
</tr>
<tr>
<td>95</td>
<td>0.27547</td>
<td>0.27242</td>
<td>0.07732</td>
<td>0.00270</td>
<td>0.05</td>
<td>0.00573</td>
</tr>
</tbody>
</table>

MAD=Median Absolute Deviation, RMSE=Root Mean Squared Error
Next, I generate data which requires the use of IVQRFE:

\[ t \in \{0, 1\} \]

Fixed Effect: \( \alpha_i \sim U(0, 1) \)

\( u_{it} \sim U(0, 1) \)

Total Disturbance: \( u^*_{it} \equiv F(\alpha_i + u_{it}) \Rightarrow u^*_{it} \sim U(0, 1) \)

Year Effect: \( \delta_0 = 1, \delta_1 = 2 \)

\( \psi_{it} \sim U(0, 1) \)

Instrument: \( z_{it} = \alpha_i + \psi_{it} \)

Policy Variable: \( d_{it} = z_{it} + u_{it} \)

Outcome: \( y_{it} = u^*_{it}(\delta_t + d_{it}) \)

Note that \( d \) is a function of \( u \) so IV is necessary. \( z \) is exogenous conditional on \( \alpha \). I generate these data for \( N = 500, T = 2 \) and, as before, \( \beta(\tau) = \tau \). Grid-searching is used to minimize the GMM objective function. Table 2 presents the results of the simulation for the coefficient of interest. To illustrate that these data require conditioning on individual fixed effects, I show results for both IVQR (left) and IVQRFE (right). The IVQR estimator used here is introduced in Chernozhukov and Hansen [2008].

5.2 Empirical Example

I use the Milwaukee Parental Choice Program (MPCP) to test the estimator in a practical application. This data set was analyzed in Rouse [1998]. The MPCP instituted a lottery to provide low-income students with vouchers for private schools. Rouse [1998] studies whether attendance at a choice school increases test scores. One specification compares the test score
Table 2: IVQRFE Simulation (N=500, T=2)

<table>
<thead>
<tr>
<th>IVQR</th>
<th>IVQRFE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Bias</td>
</tr>
<tr>
<td>5</td>
<td>0.56057</td>
</tr>
<tr>
<td>10</td>
<td>0.70229</td>
</tr>
<tr>
<td>15</td>
<td>0.80304</td>
</tr>
<tr>
<td>20</td>
<td>0.87783</td>
</tr>
<tr>
<td>25</td>
<td>0.93577</td>
</tr>
<tr>
<td>30</td>
<td>0.98169</td>
</tr>
<tr>
<td>35</td>
<td>1.01647</td>
</tr>
<tr>
<td>40</td>
<td>1.04178</td>
</tr>
<tr>
<td>45</td>
<td>1.06114</td>
</tr>
<tr>
<td>50</td>
<td>1.06906</td>
</tr>
<tr>
<td>55</td>
<td>1.06489</td>
</tr>
<tr>
<td>60</td>
<td>1.04540</td>
</tr>
<tr>
<td>65</td>
<td>1.00899</td>
</tr>
<tr>
<td>70</td>
<td>0.96410</td>
</tr>
<tr>
<td>75</td>
<td>0.91812</td>
</tr>
<tr>
<td>80</td>
<td>0.86625</td>
</tr>
<tr>
<td>85</td>
<td>0.79638</td>
</tr>
<tr>
<td>90</td>
<td>0.70683</td>
</tr>
<tr>
<td>95</td>
<td>0.58787</td>
</tr>
</tbody>
</table>

MAD=Median Absolute Deviation, RMSE=Root Mean Squared Error

gains of those selected into the program to those not selected, conditioning on individual fixed effects. These fixed effects are important because the probability of selection in the lottery was not equal for each student.

Rouse studies the mean effect of the program, but distributional impacts are also interesting. The IVQRFE estimator is ideal for this analysis. I measure the effect of choice schools on math test scores. The mean regression specification of interest is

\[ T_{ijt} = \alpha_i + \gamma_{jt} + \beta_0 P_{ijt} + \beta_1 CP_{ijt} + \epsilon_{ijt}, \]  

where \( T_{ijt} \) is the math score for student \( i \) in grade \( j \) at time \( t \). \( P \) is an indicator variable for whether or not the student is attending a choice school. \( CP \) measures the cumulative number of years the student has attended a choice school.

\( P \) and \( CP \) are potentially endogenous. I employ the same instruments as Rouse [1998] - whether a student was randomly-selected to attend a choice school, and whether the
student was chosen interacted with the number of years since the application.

I include grade-year interactions. These are especially important for the quantile analysis. They define “high-performing” and “low-performing” within the grade and year. Using the IVQRFE estimator, I estimate the following SQF

\[ S_{T_{ijt}}(\tau|P_{ijt}) = \delta_{jt}(\tau) + \beta(\tau)P_{ijt}. \]  

(25)

The equation (24) IV estimates are shown in Table 3 and are similar to those found in Rouse [1998]. The IVQRFE results are found in Figure 1. For ease of interpretation, I focus on a specification which only includes the effect of currently attending a choice school. For reference, Table 4 includes both policy variables. The conclusions remain the same.

I bootstrap to derive 95% confidence intervals. Looking at Figure 1, there appears to be some heterogeneity in the effect of choice schools, but the mean effect cannot be rejected for most of the distribution. The effect is not monotonic, however. Choice schools generally have a positive impact on students below the median of the performance distribution. However, there is little effect for above-median students, until possibly the very top.

The results contrast with the MPCP results found in Harding and Lamarche [2009] which uses conditional quantiles. Harding and Lamarche [2009] find that the effect is largest for low-achieving students and monotonically decreases throughout the distribution.\(^1\) However, this interpretation is inaccurate because of the use of conditional quantiles. The Harding and Lamarche [2009] results show how choice schools affect students in years that they are low-achieving relative to their own fixed level of performance. The difference in these results illustrates the importance of using unconditional quantile regression.

\(^1\)The estimated specifications are slightly different. Harding and Lamarche [2009] control for the grade level, assuming that the grade has a linear effect on test scores. I replace these with grade-year interactions because I believe these define the quantiles “correctly.”
6 Conclusion

In this paper, I introduce an unconditional quantile estimator for panel data. The covariates or instruments can be arbitrarily correlated with the fixed effects. The estimators maintain the nonseparable disturbance property of traditional cross-sectional quantile estimators. These estimator should be extremely useful in contexts where identification requires differences and it is believed that the effect of the variable is heterogenous throughout the outcome distribution. The resulting estimates can be interpreted in the same manner as traditional cross-sectional quantile estimates. I extend the estimator to an IV context.

I apply the estimator to the analysis of Rouse [1998]. Importantly, the conclusions drawn from this analysis are very different from the conclusions found in Harding and Lamarche [2009] which uses conditional quantiles. These results stress the importance of using unconditional quantiles in certain contexts.

The estimators in this paper contrast with existing quantile panel data estimators which typically include a separate additive term for the fixed effect. Instead, this paper
maintains the nonseparable disturbance property of traditional cross-sectional quantile estimators. The estimators are consistent for small $T$ and straightforward to use with standard statistical software.
Appendix

Theorem 4.2 [Asymptotic Normality]: If IV-A1 - IV-A9 hold and \( \hat{A} \xrightarrow{p} A \) positive definite, then

\[
\sqrt{N}(\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{d} N \left[ 0, (G'AG)^{-1}G'\Sigma AG(G'AG)^{-1} \right].
\]

Define \( \beta_0 \equiv \beta(\tau), \hat{\beta} \equiv \hat{\beta}(\tau) \).

Also, \( G(\beta) \equiv E \left[ \frac{1}{T} \sum_{t=1}^{T} (z_{it} - \bar{z}_i) d_{it}'f_y(d_{it}'\beta | \alpha_i, z_i) \right] \).

\( g(\beta) \equiv E \left[ \frac{1}{T} \sum_{t=1}^{T} (z_{it} - \bar{z}_i) \left[ 1(y_{it} \leq d_{it}'\beta) \right] \right] \).

Proof:

\( g(\beta)'Ag(\beta) \) is minimized at \( \beta_0 \) implying that

\[
G(\beta_0)'Ag(\beta_0) = 0.
\]

Expanding each element of \( g(\beta_0) \) around \( \hat{\beta} \) and multiplying by \( \sqrt{N} \) gives

\[
\sqrt{N}g(\beta_0) = \sqrt{N}g(\hat{\beta}) - G(\hat{\beta})\sqrt{N}(\hat{\beta} - \beta_0)
\]

where \( \hat{\beta} \) meets the condition \( \| \hat{\beta} - \beta_0 \| \leq \| \hat{\beta} - \beta_0 \| \) and takes on different values for each column of \( G(\hat{\beta}) \).

By assumption of i.i.d and continuity, \( G(\hat{\beta}) \xrightarrow{p} G(\beta_0) \).

Focus on the \( \sqrt{N}g(\hat{\beta}) \) term:

\[
-\sqrt{N}g(\hat{\beta}) = \left[ \sqrt{N}g_N(\hat{\beta}) - \sqrt{N}g(\hat{\beta}) \right] - \sqrt{N}g_N(\hat{\beta})
\]

\[
= \sqrt{N} \left[ g_N(\hat{\beta}) - g(\hat{\beta}) - g_N(\beta_0) \right] + \sqrt{N}g_N(\beta_0) - \sqrt{N}g_N(\hat{\beta}).
\]

32
(1): Define empirical process $v_N(\beta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [g_N(\beta) - \bar{g}(\beta)]$.

The functional class $\{1(y_{it} \leq d'_ib), b \in \mathbb{R}^k\}$ is Donsker and the Donsker property is preserved when the class is multiplied by a bounded random variable (see Theorem 2.10.6 in van der Vaart and Wellner [1996]). Thus,

$$\left\{ \frac{1}{T} \sum_{t=1}^{T} (z_{it} - \bar{z}_i) \left[1(y_{it} \leq d'_ib)\right], b \in \mathbb{R}^k \right\}$$

is Donsker with envelope $2 \max_{(i,t)} |z_{it} - \bar{z}_i|$. Stochastic equicontinuity of $v_N(\cdot)$ follows from IV-A8 and Theorem 1 in Andrews [1994]. Stochastic equicontinuity and consistency of $\hat{\beta}$ implies that part (1) is $o_p(1)$.

(2): By the Central Limit Theorem, $\sqrt{Ng_N(\beta_0)} \xrightarrow{d} N(0, \Sigma)$ where $\Sigma = E[g(\beta_0)g(\beta_0)']$.

(3): By consistency of $\hat{\beta}$, $\sqrt{Ng_N(\hat{\beta})} = o_p(1)$.

Plugging into (26) and using the assumption that $G'AG$ nonsingular

$$\sqrt{N}(\hat{\beta} - \beta(\tau)) \xrightarrow{d} N\left[0, (G'AG)^{-1}G'\Sigma AG(G'AG)^{-1}\right]$$
Table 3: Mean IV Results

<table>
<thead>
<tr>
<th></th>
<th>Dependent Variable: Math Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Currently enrolled in choice school</td>
<td>2.317**  -1.747</td>
</tr>
<tr>
<td></td>
<td>(1.219)  (1.216)</td>
</tr>
<tr>
<td>Cumulative number of years enrolled</td>
<td>3.200***</td>
</tr>
<tr>
<td></td>
<td>(0.618)</td>
</tr>
<tr>
<td>NT</td>
<td>7490              7490</td>
</tr>
</tbody>
</table>

Note: Standard Errors (in parentheses) are clustered by student. Significance levels: *10%, **5%, ***1%. Specification includes individual fixed effects and grade-year interactions.
<table>
<thead>
<tr>
<th>Quantile</th>
<th>Currently enrolled in choice school</th>
<th>Cumulative number of years enrolled</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2.8</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>(3.0)</td>
<td>(1.3)</td>
</tr>
<tr>
<td>10</td>
<td>0.8</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>(2.8)</td>
<td>(1.2)</td>
</tr>
<tr>
<td>15</td>
<td>1.7</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>(2.5)</td>
<td>(1.1)</td>
</tr>
<tr>
<td>20</td>
<td>-0.1</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>(2.5)</td>
<td>(1.0)</td>
</tr>
<tr>
<td>25</td>
<td>-1.0</td>
<td>2.0**</td>
</tr>
<tr>
<td></td>
<td>(2.5)</td>
<td>(0.9)</td>
</tr>
<tr>
<td>30</td>
<td>0.9</td>
<td>2.0**</td>
</tr>
<tr>
<td></td>
<td>(2.1)</td>
<td>(1.0)</td>
</tr>
<tr>
<td>35</td>
<td>2.7</td>
<td>2.1***</td>
</tr>
<tr>
<td></td>
<td>(2.0)</td>
<td>(0.8)</td>
</tr>
<tr>
<td>40</td>
<td>1.5</td>
<td>2.6***</td>
</tr>
<tr>
<td></td>
<td>(1.9)</td>
<td>(0.7)</td>
</tr>
<tr>
<td>45</td>
<td>0.3</td>
<td>2.4***</td>
</tr>
<tr>
<td></td>
<td>(1.9)</td>
<td>(0.7)</td>
</tr>
<tr>
<td>50</td>
<td>-0.7</td>
<td>2.7***</td>
</tr>
<tr>
<td></td>
<td>(1.7)</td>
<td>(0.6)</td>
</tr>
<tr>
<td>55</td>
<td>0.6</td>
<td>2.1***</td>
</tr>
<tr>
<td></td>
<td>(2.0)</td>
<td>(0.6)</td>
</tr>
<tr>
<td>60</td>
<td>-0.8</td>
<td>1.7***</td>
</tr>
<tr>
<td></td>
<td>(1.9)</td>
<td>(0.6)</td>
</tr>
<tr>
<td>65</td>
<td>-2.1</td>
<td>2.0***</td>
</tr>
<tr>
<td></td>
<td>(1.7)</td>
<td>(0.6)</td>
</tr>
<tr>
<td>70</td>
<td>-1.5</td>
<td>1.4**</td>
</tr>
<tr>
<td></td>
<td>(1.9)</td>
<td>(0.7)</td>
</tr>
<tr>
<td>75</td>
<td>-1.5</td>
<td>1.5**</td>
</tr>
<tr>
<td></td>
<td>(2.0)</td>
<td>(0.8)</td>
</tr>
<tr>
<td>80</td>
<td>-1.2</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>(1.9)</td>
<td>(0.9)</td>
</tr>
<tr>
<td>85</td>
<td>-0.5</td>
<td>2.4**</td>
</tr>
<tr>
<td></td>
<td>(2.2)</td>
<td>(1.2)</td>
</tr>
<tr>
<td>90</td>
<td>-0.2</td>
<td>3.1*</td>
</tr>
<tr>
<td></td>
<td>(2.8)</td>
<td>(1.6)</td>
</tr>
<tr>
<td>95</td>
<td>4.8</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>(3.1)</td>
<td>(1.9)</td>
</tr>
</tbody>
</table>

Note: Standard Errors (in parentheses) are clustered by student. Significance levels: *10%, **5%, ***1%. Specification includes individual fixed effects and grade-year interactions.
References


