WORKING P A P E R

Unconditional Quantile Regression for Panel Data with Exogenous or Endogenous Regressors

DAVID POWELL

WR-710-1

November 2010

This paper series made possible by the NIA funded RAND Center for the Study of Aging (P30AG012815) and the NICHD funded RAND Population Research Center (R24HD050906).

This product is part of the RAND Labor and Population working paper series. RAND working papers are intended to share researchers' latest findings and to solicit informal peer review. They have been approved for circulation by RAND Labor and Population but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.

LABOR AND POPULATION

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

David Powell
RAND[†]

November 28, 2010

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, redefines the quantiles. This paper introduces a quantile estimator for panel data which conditions on fixed effects for identification but allows the parameters of interest to be interpreted in the same manner as cross-sectional quantile estimates. The quantile treatment effects are unconditional in the fixed effect but identification originates from differences in the covariates or instruments. The fixed effects are never estimated and the estimator is consistent for small T .

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 entire 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 all of the 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 produces consistent estimates for small T . An IV version is also introduced (IVQRFE).

Estimating QTEs in the presence of fixed effects can be problematic. Panel data 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. 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, including repeated cross-sections where fixed effects are based on cells.

Let $\mathbf{d}$ denote the policy or treatment variables. This terminology is used in Powell [2010]. With unconditional QTEs, we are interested in the distribution of $y_{it}|\mathbf{d}_{it}$, not the distribution of $y_{it}|\mathbf{d}_{it}, \alpha_i$. In this paper, I focus on linear quantiles so the model of interest

is

$$y_{it} = \mathbf{d}'_{it}\beta(\alpha_i + \epsilon_{it}), \quad (1)$$

where $\alpha_i = h(\mathbf{d}_{i1}, \dots, \mathbf{d}_{iT}, \mu_i)$. $\alpha_i + \epsilon_{it}$ is the disturbance, interpreted by Doksum [1974] as individual ability or proneness. It is useful to convert $\alpha_i + \epsilon_{it}$ into a “rank variable,” $u^*_{it} \sim U(0, 1)$.¹ An observation with a higher u^* is more prone to the outcome variable, for a given $\mathbf{d}$. If the outcome variable is an individual’s wage, then people with higher u^* have higher individual ability in the labor market. Thus, the equation of interest can be written as

$$y_{it} = \mathbf{d}'_{it}\beta(u^*_{it}), \quad u^*_{it} \sim U(0, 1). \quad (2)$$

To use similar notation as Powell [2010], we can rewrite this equation as $y_{it} = \mathbf{d}'_{it}\beta(u^*_{it}(\alpha_i))$ to illustrate the relationship between u^*_{it} and α_i . The two measures of ability are related concepts - u^*_{it} is the underlying individual ability while α_i in fixed effect models is commonly thought of as a measure of fixed individual ability.

Quantile regression allows the coefficients of interests to vary based on “unobserved proneness.” By including additional variables - such as individual fixed effects - some of this unobserved proneness becomes observed, altering the interpretation of the coefficients. In Powell [2010], I discuss the merits of estimating effects which vary based on “total proneness” so that the interpretation does not vary as one adds covariates. The estimator introduced in this paper allows for individual fixed effects to be conditioned on, but the coefficients vary based on “total proneness.”

With mean regression, the disturbance does not take on such an important interpretation since distinguishing between observed and unobserved proneness is unnecessary. Consider the specification

$$y_{it} = \alpha_i + \mathbf{d}'_{it}\delta + \zeta_{it}.$$

An OLS regression of y on $\mathbf{d}$ will provide consistent estimates if $\mathbf{d}_{it}$ is orthogonal to α_i and the disturbance. Including α_i in the regression does not affect the consistency or interpretation of the estimates. With quantile estimation, however, the inclusion of α_i changes the interpretation even when it is orthogonal to $\mathbf{d}_{it}$. Put differently, the 90th percentile of ϵ is likely different from the 90th percentile of $\alpha + \epsilon$.

To adopt similar terminology as Chernozhukov and Hansen [2008], the Structural

¹ u^* is simply the CDF of $\alpha + \epsilon$: $u^*_{it} = P(\alpha + \epsilon \leq \alpha_i + \epsilon_{it})$.

Quantile Function (SQF) of interest for equation (2) is

$$S_y(\tau|\mathbf{d}) = \mathbf{d}'\beta(\tau), \quad \tau \in (0, 1). \quad (3)$$

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

1.1 Motivation

Cross-sectional quantile estimators are useful for specifications such as

$$y_i = \mathbf{d}'_i\beta(\alpha_i + \epsilon_i). \quad (4)$$

With cross-sectional data, it is unnecessary to distinguish between α_i and ϵ_i , but this formulation is useful for comparative purposes later. If $u^*|\mathbf{d} \sim U(0, 1)$, then QR can estimate the relevant SQF

$$S_y(\tau|\mathbf{d}) = \mathbf{d}'\beta(\tau), \quad \tau \in (0, 1). \quad (5)$$

It may be the case that $\mathbf{d}$ is endogenous such that $u^*|\mathbf{d} \not\sim U(0, 1)$. 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. Existing panel data quantile estimators use a location-shift model where the fixed effect is held constant for all quantiles or it is allowed to vary by quantile

$$y_{it} = \alpha_i + \mathbf{d}'_{it}\beta(\epsilon_{it}) \quad \text{or} \quad y_{it} = \alpha_i(\epsilon_{it}) + \mathbf{d}'_{it}\beta(\epsilon_{it}). \quad (6)$$

The underlying equation of interest has changed as these estimators separate the components of the residual. For this reason, these estimators cannot be used for equations such as (4). Using a location-shift model in these circumstances is similar to differencing

one’s data and then using quantile regression. The “high quantiles” refer to observations experiencing large *increases* in the outcome variable. 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 can only vary based on the non-fixed component of underlying ability.

This paper is interested in specifications with nonseparable disturbances such as

$$y_{it} = \mathbf{d}_{it}'\beta(\alpha_i + \epsilon_{it}). \quad (7)$$

The coefficients of interests vary based on the “total disturbance” in the same way as equation (4) and the SQF is still represented by equation (5). 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 (5) is not possible with cross-sectional data, but the researcher does not want to alter the SQF by using panel data. The distinction is that τ in equation (5) refers to the τ^{th} quantile of $\alpha_i + \epsilon_{it}$. A location-shift model assumes that the SQF is $S_y(\tau|\mathbf{d}_{it}, \alpha_i) = \alpha_i + \mathbf{d}_{it}'\beta(\tau)$ where τ refers to the τ^{th} quantile of ϵ_{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.

1.1.1 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 potentially independently affect 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. Does the program help low-achieving students more than high-achieving students?

Let α_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

$$T_{it} = \delta_t(\alpha_i + \epsilon_{it}) + v_{it}\beta(\alpha_i + \epsilon_{it}). \quad (8)$$

The SQF is

$$S_{T_{it}}(\tau|v_{it}) = \delta_t(\tau) + v_{it}\beta(\tau). \quad (9)$$

Once we estimate the SQF, we can generate counterfactual distributions of test scores. 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. If we are interested in how vouchers affect high ability and low ability students, we cannot difference the data. Similarly, simply including individual fixed effects in a quantile regression or using a location-shift model causes problems since this implicitly “differences out” the individual’s placement in the distribution.

Instead, we want to condition on α without changing the interpretation of the parameters. The high quantiles should refer to observations in the top of the cross-sectional distribution given their policy variables. This paper’s estimator allows unconditional QTEs to be estimated in the presence of individual fixed effects.

2 Existing Literature

A small literature has focused on the extension of quantile estimation to panel data. In the motivating example above, the underlying equation has the form

$$y_{it} = \mathbf{d}_{it}'\beta(\alpha_i + \epsilon_{it}). \quad (10)$$

The coefficients of interests are a function of the “total disturbance,” including the fixed effect. Estimating this equation allows the results to be interpreted in the same manner as cross-sectional quantile regression results, which also vary based on the total disturbance. Many existing quantile panel data estimators, however, do not estimate the above equation. Instead, they use a location-shift model, separately estimating α_i so that the parameters of interest vary based only on ϵ_{it} , the observation-specific disturbance. These estimators are useful in contexts when we want to define high quantiles by observations with large 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. Location-shift models are preferable in certain situations. The estimator in this paper is preferable in situations where panel data are used for identification purposes only and the researcher is not interested in the distribution of the *difference* in the outcome variable. In other words, QRFE is useful when the researcher would like to do a simple cross-sectional quantile regression of y on $\mathbf{d}$, but panel data are necessary for identification.

Koenker [2004] introduces a quantile fixed effects estimator which separately estimates a fixed effect under the assumption that

$$y_{it} = \alpha_i + \mathbf{d}'_{it}\beta(\epsilon_{it}). \quad (11)$$

Similarly, Harding and Lamarche [2009] introduce an IV quantile panel data estimator under the assumption that

$$y_{it} = \alpha_i(\epsilon_{it}) + \mathbf{d}'_{it}\beta(\epsilon_{it}). \quad (12)$$

In both cases, the coefficient of interest (β) varies only with ϵ . For illustrative purposes, assume that α is known and provided to the econometrician. These estimators are equivalent to a traditional quantile regression of $(y - \alpha)$ on $\mathbf{d}$. Using the above example, it relates voucher receipt to the test scores of students at the top of the distribution *relative to their own underlying fixed skill level*. Assume that $(\alpha_1 = 40, T_{11} = 50)$, $(\alpha_2 = 80, T_{21} = 90)$, and $v_{11} = v_{21}$. Since $T_{i1} - \alpha_i = 10$ in each case (and voucher status is the same), existing panel data estimators assume that both of these students have the same “ability” and would react the same to receipt of a voucher. However, cross-sectionally, student 1 is low-achieving and student 2 is high-achieving. The estimates cannot be interpreted in the same manner as cross-sectional estimates because the SQF has changed to $S_{y_{it}}(\tau|\mathbf{d}_{it}, \alpha_i) = \alpha_i + \mathbf{d}'_{it}\beta(\tau)$ or $S_{y_{it}}(\tau|\mathbf{d}_{it}, \alpha_i) = \alpha_i(\tau) + \mathbf{d}'_{it}\beta(\tau)$ where τ relates to ϵ only.

This paper introduces an estimator that estimates the relevant SQF (equation (9)) while conditioning on individual fixed effects. The fixed effects are used for identification purposes only, allowing for an arbitrary correlation between the fixed effects and the policy variables (or instruments). Typically, 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. Location-shift models let the coefficients of interest vary based on ϵ only. The SQF for quantile τ refers to the τ^{th} quantile

of ϵ . With the QRFE and IVQRFE estimators of this paper, the SQF for quantile τ refers to the τ^{th} quantile of $\alpha + \epsilon$. 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 (5).

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

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 let the quantiles be defined by the total disturbance (including the fixed effect). Abrevaya and Dahl [2008] introduce this technique. Graham and Powell [2008] discuss a similar estimation strategy.

Similarly, Chernozhukov et al. [2009] discuss identification of bounds on quantile effects in nonseparable panel models where the quantiles are defined by $(\alpha_i, \epsilon_{it})$.

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 .

3 Model

The specification of interest is

$$y_{it} = \mathbf{d}_{it}'\beta(u_{it}^*), \quad u_{it}^* \sim U(0, 1). \quad (13)$$

The motivation of this paper is that for situations where $u_{it}^*|\mathbf{d}_{it} \not\sim U(0, 1)$, QR cannot be used. Simply including individual fixed effects in a quantile regression does not solve the problem because it assumes a different specification. Also, note that u_{it}^* is a function of α_i , implying that $u_{it}^*|\mathbf{d}_{it}, \alpha_i \not\sim U(0, 1)$. Instead, exogeneity is defined differently. In words, the exogeneity assumption is that within-individual changes in the policy variables do not

provide information about changes in the underlying proneness. This suggests using pairwise comparisons between observations with the same fixed effect.

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 (or any set of fixed effects which saturate the space) as exogenous policy variables. If the outcome variable is the individual’s wage, then year fixed effects allow the wage distribution to shift across years. An individual with a large u^* will make a higher 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 $\mathbf{d}_i \equiv (\mathbf{d}_{i1}, \dots, \mathbf{d}_{iT})$ and $\bar{\mathbf{d}}_i \equiv \frac{1}{T} \sum_{t=1}^T \mathbf{d}_{it}$.

3.2.1 Assumptions

The following conditions hold jointly with probability one:

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

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

A3 Full Rank: $E[\mathbf{d}_i]$ is rank k .

A4 Continuity: y_{it} continuously distributed conditional on $\mathbf{d}_i, \alpha_i$.

The first assumption (**A1**) is a standard monotonicity condition for quantile estimators. **A2** is an independence assumption. This assumption is slightly different from the

equivalent condition found in Powell [2010]. **A2** could be replaced by $P(u_{it}^* \leq \tau | \mathbf{d}_i, \alpha_i) = P(u_{it}^* \leq \tau | \alpha_i)$ and an assumption of stationarity so that the distributions of u_{it}^* and u_{is}^* are the same. Instead, I use a slightly weaker assumption. The distribution of u_{it}^* can change over time, but this change must be independent of $\alpha_i, \mathbf{d}_i$.

A3 requires within-individual variation in the policy variables. **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 α_i . Furthermore, there are no assumptions on the relationship between α_i and $\mathbf{d}_{it}$.

3.2.2 Moment Conditions

These assumptions lead to two separate moment conditions. Both conditions must be used simultaneously.

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

$$E \left\{ \mathbf{d}_{it} \left[\mathbf{1}(y_{it} \leq \mathbf{d}_{it}'\beta(\tau)) - \frac{1}{T} \sum_{s=1}^T \mathbf{1}(y_{is} \leq \mathbf{d}_{is}'\beta(\tau)) \right] \right\} = 0, \quad (14)$$

$$E[\mathbf{1}(y_{it} \leq \mathbf{d}_{it}'\beta(\tau)) - \tau] = 0. \quad (15)$$

Proof of (14):

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

$$\begin{aligned}
&= E \left[E \left\{ \mathbf{d}_{it} \left[\mathbf{1}(y_{it} \leq \mathbf{d}'_{it}\beta(\tau)) - \frac{1}{T} \sum_{s=1}^T \mathbf{1}(y_{is} \leq \mathbf{d}'_{is}\beta(\tau)) \right] \middle| \alpha_i, \mathbf{d}_i \right\} \right] \\
&= E \left[E \left\{ \mathbf{d}_{it} \left[\mathbf{1}(\mathbf{d}'_{it}\beta(u_{it}^*) \leq \mathbf{d}'_{it}\beta(\tau)) - \frac{1}{T} \sum_{s=1}^T \mathbf{1}(\mathbf{d}'_{is}\beta(u_{is}^*) \leq \mathbf{d}'_{is}\beta(\tau)) \right] \middle| \alpha_i, \mathbf{d}_i \right\} \right] \quad \text{by A1} \\
&= E \left[\mathbf{d}_{it} E \left\{ \mathbf{1}(u_{it}^* \leq \tau) - \frac{1}{T} \sum_{s=1}^T \mathbf{1}(u_{is}^* \leq \tau) \middle| \alpha_i, \mathbf{d}_i \right\} \right] \quad \text{by A1} \\
&= 0 \quad \text{by A2}
\end{aligned}$$

Proof of (15):

$$\begin{aligned}
E[\mathbf{1}(y_{it} \leq \mathbf{d}'_{it}\beta(\tau))] &= E[\mathbf{1}(\mathbf{d}'_{it}\beta(u_{it}^*) \leq \mathbf{d}'_{it}\beta(\tau))] \quad \text{by A1} \\
&= P[u_{it}^* \leq \tau] \quad \text{by A1} \\
&= \tau \quad \text{by A1}
\end{aligned}$$

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

Sample Moment 1

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

Sample Moment 2

$$h(b) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbf{1}(y_{it} \leq \mathbf{d}'_{it}b) - \tau.$$

Sample Moment 1 is worth discussing further. Define $\tau_i(b) \equiv \frac{1}{T} \sum_{s=1}^T \mathbf{1}(y_{is} \leq \mathbf{d}'_{is}b)$. Note that the moment condition is similar to the cross-sectional quantile moment condition where τ is replaced by τ_i :

$$g_{it}(b) = \mathbf{d}_{it} \left[\mathbf{1}(y_{it} \leq \mathbf{d}'_{it}b) - \tau_i \right].$$

This makes intuitive sense. The individual fixed effect provides information about the distribution of the disturbance. An individual with high fixed ability (i.e., a large α_i) is more likely to be in the top of the wage distribution, impacting τ_i . Thus, instead of assuming that

$\tau_i = \tau$ for each individual, this condition implicitly makes within-individual comparisons of the individual's underlying disturbance. Consistent estimation of τ_i is not necessary since Sample Moment 1 is equivalent to a series of pairwise comparisons and can be replaced by

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

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

$$g_i(b) = \frac{1}{T} \left\{ \sum_{t=1}^T (\mathbf{d}_{it} - \bar{\mathbf{d}}_i) \left[\mathbf{1}(y_{it} \leq \mathbf{d}'_{it}b) \right] \right\}. \quad (16)$$

Sample Moment 2 essentially defines the quantile and 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, which is why the resulting estimates can be interpreted in the same manner as cross-sectional quantile estimates. With QR, one assumes that $u^* \sim U(0, 1)$ and $u^*|\mathbf{d} \sim U(0, 1)$. The QRFE estimator does *not* assume $u^*|\mathbf{d} \sim U(0, 1)$, 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

$$\hat{g}(b) = \frac{1}{N} \sum_{i=1}^N g_i(b). \quad (17)$$

It is necessary to use Sample Moment 2 as well. Define

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

Then,

$$\widehat{\beta(\tau)} = \arg \min_{b \in \mathcal{B}} \hat{g}(b)' \hat{A} \hat{g}(b) \quad (18)$$

for some weighting matrix $\hat{A}$.

There is a straightforward way to confine all guesses b to the set $\mathcal{B}$, but it is first

helpful to discuss year fixed effects.

Year Fixed Effects

Moment Condition 1 (equation (14)) represents k separate conditions. The inclusion of year fixed effects implies

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

Equation (15), then, implies

$$P(y_{it} \leq \mathbf{d}'_{it}\beta(\tau)) = \tau \quad \text{for all } t. \quad (19)$$

By assuming the inclusion of year fixed effects, I can use equation (19) for the sample moments. Let $\mathbf{d} \equiv (\mathbf{x}, \gamma_t)$ where $\mathbf{x}$ are the variables of interest and γ_t is a set of year fixed effects. For the rest of this section, let $\tilde{b}$ be the coefficients on $\mathbf{x}$ so that $\mathbf{d}'_{it}b = \gamma_t + \mathbf{x}'_{it}\tilde{b}$. The sample moments can be replaced by

Sample Moment 1'

$$g_i(b) = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_{it} \left[\mathbf{1}(y_{it} \leq \mathbf{d}'_{it}b) - \frac{1}{T} \sum_{s=1}^T \mathbf{1}(y_{is} \leq \mathbf{d}'_{is}b) \right], \quad (20)$$

Sample Moment 2'

$$h_t(b) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(y_{it} \leq \mathbf{d}'_{it}b) - \tau \quad \text{for all } t. \quad (21)$$

Sample Moment 2' defines the year fixed effects. The value of these fixed effects forces $y_{it} \leq \mathbf{d}'_{it}b$ 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

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

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

$$\hat{\gamma}_t(\tau, \tilde{b}) \quad \text{solves} \quad \frac{1}{N} \sum_i^N \mathbf{1}(y_{it} - \mathbf{x}'_{it}\tilde{b} \leq \hat{\gamma}_t(\tau, \tilde{b})) = \tau. \quad (22)$$

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 values $\hat{\gamma}_t(\tau, \tilde{b})$ are automatically known. This simplifies the estimation process. Thus, the first step is to guess $\tilde{b}$. Second, calculate the year fixed effects using equation (22). Then, evaluate $\hat{g}(b)' \hat{A} \hat{g}(b)$ where $g_i(b)$ is defined by equation (20). The b that minimizes this condition is $\widehat{\beta(\tau)}$.

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 $(E[g_i(\check{\beta})] = 0, E[h(\check{\beta})] = 0) \Leftrightarrow \check{\beta} = \beta(\tau)$ for $\tau \in (0, 1)$. I use the equation (16) formulation in this section.

Lemma 3.1. *If $P[y_{it} \leq \mathbf{d}'_{it}\check{\beta} | \mathbf{d}_i, \alpha_i] = P[y_{it} \leq \mathbf{d}'_{it}\beta(\tau) | \mathbf{d}_i, \alpha_i]$, then $\mathbf{d}'_{it}\check{\beta} = \mathbf{d}'_{it}\beta(\tau)$.*

This follows directly from **A4**.

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

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

Under **A3**, this condition requires $P(y_{it} \leq \mathbf{d}'_{it}\check{\beta} | \mathbf{d}_i, \alpha_i) = P(y_{it} \leq \mathbf{d}'_{it}\beta(\tilde{\tau}) | \mathbf{d}_i, \alpha_i)$ for all t , for some $\tilde{\tau} \in (0, 1)$.

By Lemma 3.1, we know that $\mathbf{d}'_{it}\check{\beta} = \mathbf{d}'_{it}\beta(\tilde{\tau})$ for all t . **A3** implies $\check{\beta} = \beta(\tilde{\tau})$.

Because of (iii), we know that $\tilde{\tau} = \tau \Rightarrow \boxed{\check{\beta} = \beta(\tau)}$. □

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 in the presence of fixed effects 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].

I need to introduce some notation. Let $\tilde{\mathbf{z}}_i \equiv (\mathbf{z}_{i1} - \bar{\mathbf{z}}_i, \dots, \mathbf{z}_{iT} - \bar{\mathbf{z}}_i)$. Define Π_i as the relationship between $\mathbf{z}$ and $\mathbf{d}$,

$$\Pi_i \equiv \begin{bmatrix} P(\mathbf{d}_{i1} = \mathbf{d}^{(1)} | \alpha_i, \mathbf{z}_{i1}) & \cdots & P(\mathbf{d}_{i1} = \mathbf{d}^{(m)} | \alpha_i, \mathbf{z}_{i1}) \\ \vdots & \ddots & \vdots \\ P(\mathbf{d}_{iT} = \mathbf{d}^{(1)} | \alpha_i, \mathbf{z}_{iT}) & \cdots & P(\mathbf{d}_{iT} = \mathbf{d}^{(m)} | \alpha_i, \mathbf{z}_{iT}) \end{bmatrix}.$$

Define D as a matrix of all possible values for $\mathbf{d}$,

$$D \equiv \begin{bmatrix} \mathbf{d}^{(1)'} \\ \vdots \\ \mathbf{d}^{(m)'} \end{bmatrix}.$$

Finally,

$$\Gamma_i \equiv \begin{bmatrix} P(y_{it} \leq \mathbf{d}^{(1)'} \beta(\tau) | \alpha_i, \mathbf{z}_i) \\ \vdots \\ P(y_{it} \leq \mathbf{d}^{(m)'} \beta(\tau) | \alpha_i, \mathbf{z}_i) \end{bmatrix}.$$

3.3.1 Assumptions

The following conditions hold jointly with probability one:

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

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

IV-A3 Full Rank: D is rank k .

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

IV-A5 Continuity: y_{it} continuously distributed conditional on $\mathbf{z}_i, \alpha_i$.

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 $\mathbf{d}$. With discrete variables, this is possible by creating dummy variables for each possible value of $\mathbf{z}$. 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 $\mathbf{d}$

$$\dot{D} \equiv \begin{bmatrix} \mathbf{d}^{(j_1)'} \\ \vdots \\ \mathbf{d}^{(j_s)'} \end{bmatrix},$$

where $s \leq m$. Also define

$$\dot{\Pi}_i \equiv \begin{bmatrix} P(\mathbf{d}_{i1} = \mathbf{d}^{(j_1)} | \alpha_i, \mathbf{z}_{i1}) & \cdots & P(\mathbf{d}_{i1} = \mathbf{d}^{(j_s)} | \alpha_i, \mathbf{z}_{i1}) \\ \vdots & \ddots & \vdots \\ P(\mathbf{d}_{iT} = \mathbf{d}^{(j_1)} | \alpha_i, \mathbf{z}_{iT}) & \cdots & P(\mathbf{d}_{iT} = \mathbf{d}^{(j_s)} | \alpha_i, \mathbf{z}_{iT}) \end{bmatrix}.$$

Identification will hold as long as there exists a subset of $\mathbf{d}$ such that (i) $\dot{D}$ is full rank and (ii) $E[\tilde{\mathbf{z}}_i \dot{\Pi}_i]$ 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*

$\tau \in (0, 1)$,

$$E \left\{ \mathbf{z}_{it} \left[\mathbf{1}(y_{it} \leq \mathbf{d}'_{it}\beta(\tau)) - \frac{1}{T} \sum_{s=1}^T \mathbf{1}(y_{is} \leq \mathbf{d}'_{is}\beta(\tau)) \right] \right\} = 0, \quad (23)$$

$$E[\mathbf{1}(y_{it} \leq \mathbf{d}'_{it}\beta(\tau)) - \tau] = 0. \quad (24)$$

Notice that equation (24) is exactly the same as the exogenous case. With equation (23), $\mathbf{d}_{it}$ has simply been replaced by $\mathbf{z}_{it}$. The sample moments are also similar.

IV Sample Moment 1

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

IV Sample Moment 2

$$h(b) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbf{1}(y_{it} \leq \mathbf{d}'_{it}b) - \tau.$$

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

$$g_i(b) = \frac{1}{T} \left\{ \sum_{t=1}^T (\mathbf{z}_{it} - \bar{\mathbf{z}}_i) \left[\mathbf{1}(y_{it} \leq \mathbf{d}'_{it}b) \right] \right\}. \quad (25)$$

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].

Lemma 3.2. *If $P[y_{it} \leq \mathbf{d}'_{it}\check{\beta}|\mathbf{z}_i, \alpha_i] = P[y_{it} \leq \mathbf{d}'_{it}\beta(\tau)|\mathbf{z}_i, \alpha_i]$, then $\mathbf{d}'_{it}\check{\beta} = \mathbf{d}'_{it}\beta(\tau)$.*

This follows directly from **IV-A5**.

Theorem 3.4 (Identification). *If (i) **IV-A1** - **IV-A5** hold;*

(ii) $E \left\{ \frac{1}{T} \sum_{t=1}^T (\mathbf{z}_{it} - \bar{\mathbf{z}}_i) \left[\mathbf{1}(y_{it} \leq \mathbf{d}'_{it} \check{\beta}) \right] \right\} = 0$; (iii) $E \left[\mathbf{1}(y_{it} \leq \mathbf{d}'_{it} \check{\beta}) \right] = \tau$, then $\check{\beta} = \beta(\tau)$.

Proof. Starting with (ii), we have $E[\check{\mathbf{z}}_i \Pi_i \check{\Gamma}_i] = 0$.

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

By Lemma 3.2, we know that $\mathbf{d}^{(j)'} \check{\beta} = \mathbf{d}^{(j)'} \beta(\tilde{\tau})$ for all j . **IV-A3** implies $\check{\beta} = \beta(\tilde{\tau})$.

Because of (iii), we know that $\tilde{\tau} = \tau \Rightarrow \boxed{\check{\beta} = \beta(\tau)}$. $\square$

4 Properties

This section discusses consistency and asymptotic normality of the estimator. I use the IV notation where it is possible that $\mathbf{z} = \mathbf{d}$. Let $\mu \equiv y_{it} - \mathbf{d}'_{it} \beta(\tau)$. Some additional assumptions are necessary:

IV-A6 $(y_i, \mathbf{d}_i, \mathbf{z}_i)$ i.i.d.

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

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

IV-A9 $G \equiv E \left[\frac{1}{T} \sum_{t=1}^T (\mathbf{z}_{it} - \bar{\mathbf{z}}_i) \mathbf{d}'_{it} f_{\mu}(0 | \alpha_i, \mathbf{z}_i) \right]$ exists and is nonsingular.

4.1 Consistency

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

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). Theorem 3.4 above proves identification under these assumptions. Note also that $g_i(b)$ is continuous at each b with probability one under **IV-A5**. Finally, $\|g_i(b)\| \leq \left\| \frac{1}{T} \sum_{t=1}^T (\mathbf{z}_{it} - \bar{\mathbf{z}}_i) \right\| < \infty$. Consistency follows immediately from Theorem 2.6 of Newey and McFadden.

4.2 Asymptotic Normality

The conditions for asymptotic normality are more difficult with discontinuous sample moments. First, note that under the given assumptions, the Central Limit Theorem tells us

$$\frac{1}{\sqrt{N}} \sum_i g_i(\beta(\tau)) \xrightarrow{d} N(0, \Sigma)$$

where $\Sigma = E[g_i(\beta(\tau))g_i(\beta(\tau))']$.

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

Newey and McFadden 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 $\{\mathbf{1}(y_{it} \leq \mathbf{d}'_{it}b), b \in \mathbb{R}^k\}$ is Donsker and the Donsker property is preserved when the class is multiplied by a bounded random variable. Thus,

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

is Donsker with envelope $2 \max_{(i,t)} |\mathbf{z}_{it} - \bar{\mathbf{z}}_i|$. Stochastic equicontinuity follows from Theorem 1 in Andrews [1986]. The rest of the conditions for Theorem 7.2 of Newey and McFadden are met, proving the theorem above.

4.2.1 Inference

It is well-known that there are difficulties in estimating the variance of quantile estimators. With QR, it is common to make the assumption $f_\mu(0|\mathbf{x}_i) = f_\mu(0)$. The equivalent assumption here ($f_\mu(0|\alpha_i, \mathbf{z}_i) = f_\mu(0)$) is difficult since a main motivation of this paper is that α_i and $\mathbf{z}_i$ provide information about the value of μ . It is possible to use the histogram estimation technique suggested in Powell [1986]. However, using a bootstrap method (resampling with replacement) is recommended to estimate the variance. Theorem 23.7 of van der Vaart [2000] 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:

$$\begin{aligned} t &\in \{0, 1\} \\ \text{Fixed Effect: } \alpha_i &\sim U(0, 1) \\ \epsilon_{it} &\sim U(0, 1) \\ \text{Total Disturbance: } u_{it} &\equiv F_{\alpha+\epsilon}(\alpha_i + \epsilon_{it}) \Rightarrow u_{it} \sim U(0, 1) \\ \delta_0 &= 1, \delta_1 = 2 \\ \psi_{it} &\sim U(0, 1) \\ \text{Policy Variable: } d_{it} &= \alpha_i + \psi_{it} \\ \text{Outcome: } y_{it} &= u_{it}(\delta_t + d_{it}) \end{aligned}$$

Note that d is exogenous *conditional on* α . The impact of d is a function of $\alpha + \epsilon$ and varies by observation. Consequently, the coefficient varies by quantile: $\beta(\tau) = \tau$. Year fixed effects are also crucial as the distribution changes (non-uniformly) 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.

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

Quantile	QR			QRFE		
	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE
5	0.52845	0.52845	0.28583	-0.00143	0.05219	0.00410
10	0.65416	0.65416	0.43293	-0.00019	0.06969	0.00704
15	0.73725	0.73725	0.54772	-0.00050	0.07862	0.00935
20	0.79354	0.79354	0.63339	0.00189	0.08767	0.01165
25	0.83916	0.83916	0.70754	-0.00191	0.09412	0.01382
30	0.87951	0.87951	0.77696	-0.00283	0.10107	0.01597
35	0.91001	0.91001	0.83150	-0.00831	0.10829	0.01823
40	0.93158	0.93158	0.87102	-0.00576	0.11364	0.02056
45	0.94158	0.94158	0.88951	-0.00768	0.11620	0.02096
50	0.94184	0.94184	0.88992	-0.00866	0.12181	0.02305
55	0.93174	0.93174	0.87065	-0.00995	0.12211	0.02374
60	0.91044	0.91044	0.83114	-0.01111	0.11638	0.02179
65	0.87717	0.87717	0.77129	-0.01411	0.11574	0.02122
70	0.83315	0.83315	0.69573	-0.01378	0.11470	0.02111
75	0.77712	0.77712	0.60528	-0.00975	0.11495	0.02046
80	0.70873	0.70873	0.50367	-0.01028	0.10980	0.01866
85	0.63069	0.63069	0.39926	-0.00829	0.10197	0.01588
90	0.53509	0.53509	0.28784	-0.00829	0.09038	0.01245
95	0.41742	0.41742	0.17627	-0.00210	0.07227	0.00809

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

Next, I generate data which requires the use of IVQRFE:

$$t \in \{0, 1\}$$

$$\text{Fixed Effect: } \alpha_i \sim U(0, 1)$$

$$\epsilon_{it} \sim U(0, 1)$$

$$\text{Total Disturbance: } u_{it} \equiv F(\alpha_i + \epsilon_{it}) \Rightarrow u_{it} \sim U(0, 1)$$

$$\delta_0 = 1, \delta_1 = 2$$

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

$$\text{Instrument: } z_{it} = \alpha_i + \psi_{it}$$

$$\text{Policy Variable: } d_{it} = z_{it} + \epsilon_{it}$$

$$\text{Outcome: } y_{it} = u_{it}(\delta_t + d_{it})$$

Note that d is a function of ϵ so IV is necessary. z is exogenous *conditional on* α . 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].

Table 2: IVQRFE Simulation (N=500, T=2)

Quantile	IVQR			IVQRFE		
	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE
5	0.55662	0.55662	0.31739	-0.01129	0.05664	0.00531
10	0.69923	0.69923	0.49612	-0.01153	0.07952	0.00998
15	0.79869	0.79869	0.64402	-0.01207	0.09723	0.01512
20	0.87122	0.87122	0.76421	-0.01138	0.10967	0.01916
25	0.93058	0.93058	0.87020	-0.00738	0.12215	0.02306
30	0.97709	0.97709	0.95846	-0.00783	0.13164	0.02694
35	1.01333	1.01333	1.02997	-0.00643	0.13776	0.02942
40	1.03878	1.03878	1.08163	-0.00455	0.14576	0.03219
45	1.05635	1.05635	1.11795	-0.00177	0.14710	0.03274
50	1.06545	1.06545	1.13688	-0.00164	0.15148	0.03458
55	1.06090	1.06090	1.12707	-0.00203	0.15233	0.03597
60	1.03998	1.03998	1.08281	-0.01011	0.15253	0.03658
65	1.00351	1.00351	1.00810	-0.01011	0.15156	0.03695
70	0.95952	0.95952	0.92170	-0.00630	0.14755	0.03450
75	0.91314	0.91314	0.83489	-0.00602	0.14739	0.03442
80	0.86125	0.86125	0.74283	-0.00919	0.14373	0.03197
85	0.79035	0.79035	0.62592	-0.00653	0.13644	0.02948
90	0.70120	0.70120	0.49350	-0.00997	0.11933	0.02360
95	0.58312	0.58312	0.34366	-0.00819	0.09750	0.01481

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

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 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

$$T_{ijt} = \alpha_i + \gamma_{jt} + \beta_0 P_{ijt} + \beta_1 CP_{ijt} + \epsilon_{ijt}, \quad (26)$$

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

$$S_{T_{ijt}}(\tau|P_{ijt}) = \delta_{jt}(\tau) + \beta(\tau)P_{ijt}. \quad (27)$$

The equation (26) 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 heterogeneity in the effect of choice schools. 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 the very top. The highest-performing students receive the largest gains from choice schools.

Figure 1: Impact of Choice Scores on Math Scores

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.² 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.

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 quantiles are defined by the “total disturbance,” the fixed effect and the observation-specific residual. This 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 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.

²The 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.”

7 Tables

Table 3: Mean IV Results

Dependent Variable: Math Score		
Currently enrolled in choice school	2.317** (1.219)	-1.747 (1.216)
Cumulative number of years enrolled		3.200*** (0.618)
N	7490	7490

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

Table 4: Quantile IV Results: Effect of Choice Schools on Math Scores

Quantile	Currently enrolled in choice school	Cumulative number of years enrolled
5	2.8 (2.9)	1.6 (1.4)
10	0.9 (2.5)	1.1 (1.0)
15	0.9 (2.4)	1.1 (1.0)
20	-0.1 (2.5)	1.1 (1.1)
25	0.1 (2.3)	1.8* (1.1)
30	0.9 (2.0)	2.0** (0.9)
35	2.7 (2.0)	2.1** (0.9)
40	1.2 (1.9)	2.6*** (0.7)
45	0.3 (1.7)	2.4*** (0.7)
50	-0.7 (1.6)	2.7*** (0.6)
55	-0.2 (1.9)	2.1*** (0.6)
60	-2.1 (1.7)	2.1*** (0.6)
65	-1.6 (1.6)	1.5*** (0.5)
70	-1.5 (1.9)	1.4*** (0.5)
75	-1.2 (2.1)	1.1 (0.7)
80	-2.6 (1.9)	1.6** (0.8)
85	-0.5 (2.3)	2.4** (1.1)
90	-0.2 (2.6)	3.1** (1.4)
95	5.0 (3.2)	2.6 (1.7)

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

References

- Jason Abrevaya and Christian M Dahl. The effects of birth inputs on birthweight. *Journal of Business & Economic Statistics*, 26:379–397, 2008.
- Donald W.K. Andrews. Empirical process methods in econometrics. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4 of *Handbook of Econometrics*, chapter 37, pages 2247–2294. Elsevier, 03 1986.
- Ivan A. Canay. A note on quantile regression for panel data models. Working paper, Northwestern Economics Department, May 2010.
- Victor Chernozhukov and Christian Hansen. Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, 142(1):379–398, January 2008.
- Victor Chernozhukov, Ivan Fernandez-Val, and Whitney Newey. Quantile and average effects in nonseparable panel models. Cemmap working papers, Centre for Microdata Methods and Practice, Institute for Fiscal Studies, October 2009.
- Kjell Doksum. Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.*, 2(2):267–277, 1974.
- Sergio Firpo. Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 75(1):259–276, 01 2007.
- Sergio Firpo, Nicole M. Fortin, and Thomas Lemieux. Unconditional quantile regressions. *Econometrica*, 77(3):953–973, 05 2009.
- Markus Frölich and Blaise Melly. Unconditional quantile treatment effects under endogeneity. Technical report, Brown University, January 2009.
- Antonio F. Galvao. Quantile regression for dyanmic panel data with fixed effects. Working paper, University of Illinois, Urbana-Champaign, 2008.
- Bryan S. Graham and James L. Powell. Identification and estimation of ‘irregular’ correlated coefficient models. Working paper, University of California, Barkeley, 2008.
- Matthew Harding and Carlos Lamarche. A quantile regression approach for estimating panel data models using instrumental variables. *Economics Letters*, 104(3):133 – 135, 2009.
- Roger Koenker. Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89, October 2004.

- Roger W Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, January 1978.
- Whitney K. Newey and Daniel McFadden. *Large sample estimation and hypothesis testing*.
- Maria Ponomareva. Quantile regression for panel data models with fixed effects and small T: Identification and estimation. Working paper, Northwestern Economics Department, 2010.
- David Powell. Unconditional quantile treatment effects in the presence of covariates. Working papers, RAND Corporation Publications Department, 2010.
- James L. Powell. Censored regression quantiles. *Journal of Econometrics*, 32(1):143–155, June 1986.
- Cecilia Elena Rouse. Private school vouchers and student achievement: An evaluation of the milwaukee parental choice program. *The Quarterly Journal of Economics*, 113(2): 553–602, May 1998.
- A.W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 2000.