WORKING PAPER

Unconditional Quantile Treatment Effects in the Presence of Covariates

DAVID POWELL

WR-816

December 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).

RAND LABOR AND POPULATION

Unconditional Quantile Treatment Effects in the Presence of Covariates*

David Powell†

December, 2010

Abstract

Many economic applications have found quantile models useful when the explanatory variables may have varying impacts throughout the distribution of the outcome variable. Traditional quantile estimators provide conditional quantile treatment effects. Typically, we are interested in unconditional quantiles, characterizing the distribution of the outcome variable for different values of the treatment variables. Conditioning on additional covariates, however, may be necessary for identification of these treatment effects. With conditional quantile models, the inclusion of additional covariates changes the interpretation of the estimates. This paper discusses identification of unconditional quantile treatment effects when it is necessary or simply desirable to condition on covariates. I discuss identification for both exogenous and endogenous treatment variables, which can be discrete or continuous, without functional form assumptions.

Keywords: Unconditional Quantile Treatment Effects, Quantile Regression, Instrumental Variables
JEL classification: C14, C31, C51

*I thank Whitney Newey and Jerry Hausman for their guidance when I first began working with this topic. I also had helpful discussions with Helen Hsi, Erik Meijer, Pierre-Carl Michaud, Amanda Pallais, and Hui Shan.
†RAND, dpowell@rand.org

1 Introduction

Quantile estimation is useful in describing the impact of variables throughout the outcome distribution. While mean regression is more popular in empirical work, it only provides estimates of how the explanatory variables impact the conditional mean of the outcome variable. This summary statistic may be useful, but it is also possible that it cannot explain the effect at any part of the outcome distribution. Quantile estimation, introduced by Koenker and Bassett [1978], allows the researcher to understand the effects throughout the entire distribution. For example, we may be interested in how maternal smoking affects birthweight, but we are likely primarily interested in how maternal smoking impacts the lower part of the birthweight distribution due to the problems associated with low birthweight newborns. Quantile regression allows this parameter to be estimated.

Traditional quantile estimators, such as the Koenker and Bassett [1978] quantile regression (QR), are useful for the estimation of conditional quantile treatment effects (QTEs). Conditional QTEs describe the effect of treatment variables on the conditional distribution of the outcome variable. For example, they characterize the impact on the upper end of the distribution where the "upper end" is defined by observations with large values of the outcome variable given the covariates. This may include observations that are in the lower part of the outcome distribution for their treatment variables. While conditional QTEs can be useful, we are typically interested in *unconditional* quantile treatment effects. Unconditional QTEs describe the difference in the quantiles for different values of the variables of interest, unconditional on other covariates. With exogenous treatment variables, one way to resolve this issue is simply to not control for the other covariates. Then, QR provides unconditional QTEs.

A problem arises, however, when it is necessary or simply desirable to condition on a separate set of covariates for the purposes of identification. The disturbance plays a special role in conditional quantile regression by characterizing the unobserved "proneness" for the outcome variable. As covariates are added, some of the unobserved proneness becomes observed. Consequently, the interpretation of the estimates from a quantile regression changes as covariates are added. The covariates "shift" an observation's placement in the conditional distribution.

This quality is not always desirable. Researchers typically include additional covariates for the purposes of identification, but the desired interpretation of the parameters of interest does not change. This paper considers unconditional quantile treatment effects in the presence of control variables. These effects tell us the difference between the unconditional quantiles for different values of the variables of interest. I discuss identification of unconditional QTEs for both exogenous and endogenous variables. Furthermore, I do not limit the discussion to cases where these is a single binary treatment variable. Finally, identification does not originate from any functional form assumptions.

I also compare the results of this paper to the conditions for conditional QTEs. Identification of conditional QTEs requires assumptions on the relationship between the additional covariates and the disturbance since the effect of these covariates must be separately identified. With unconditional QTEs, no such assumptions are required, making the conditions for unconditional QTEs less restrictive than those needed for conditional QTEs. Similarly, I discuss how conditional QTEs are a special case of unconditional QTEs.

In the next section, I introduce a new terminology which makes it easier to discuss unconditional QTEs. Section 3 reviews the literature. Section 4 discuss unconditional treatment effects for both exogenous and endogenous treatment variables. Section 5 concludes.

2 Terminology

Quantile regression is useful for nonseparable models such as

$$y = q(\mathbf{d}, u^*), \quad u^* \sim U(0,1). \tag{1}$$

Doksum [1974] interprets the disturbance u^* as unobserved (or underlying) "proneness" or individual ability. It is a "rank variable" which describes the underlying rank within the distribution of the outcome variable. Observations with a large u^* are "prone" to a large y for a given $\mathbf{d}$. For example, if the dependent variable is an individual's wage, then individuals with a high u^* are "high ability" individuals in the labor market. This paper finds this terminology slightly limiting but builds on this interpretation. The concern is that researchers may need or want to account for control variables $\mathbf{x}$ which are related to proneness.

This paper considers a nonseparable model of the form

$$y = q(\mathbf{d}, u^*(\mathbf{x})), \quad u^* \sim U(0,1). \tag{2}$$

$\mathbf{d}$ denotes the policy variables or treatment variables. These are the variables of interest that we believe affect the outcome variable for a given level of proneness. $\mathbf{x}$ denotes measures of "observed proneness" (or observed skill). These are the control variables that we want to condition on for identification purposes.

Let u represent "unobserved proneness" (or unobserved skill). Consequently, we can define "total proneness" as $u^* \equiv f(\mathbf{x}, u), u^* \sim U(0,1)$, where the relationship between $\mathbf{x}$ and u is never specified and, importantly, may be arbitrarily correlated with one another. Unconditional QTEs describe the impact of the policy variables at different levels of "total proneness." To clarify this terminology further and illustrate its value, $\mathbf{x}$ and u determine placement in the outcome distribution assuming all observations have the same $\mathbf{d}$. The policy variables impact the outcome variable of each observation based on that observation's u^*. If we are interested in how job training affects earnings, the policy variable is a dummy variable for "training." Other variables such as race, gender, and education affect each person's earnings regardless of training status. These are the control variables or "observed proneness." The terminology should be clear from equation (2). The policy variables ($\mathbf{d}$) impact the distribution of y. The control variables or "observed proneness" are determinants (or correlates) of the rank variable.

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

$$y = \alpha + \mathbf{d}'\delta + \mathbf{x}'\phi + \zeta.$$

An OLS regression of y on $\mathbf{d}$ will provide consistent estimates if $\mathbf{d}$ is orthogonal to $\mathbf{x}$ and the disturbance. Including $\mathbf{x}$ in the regression does not affect the consistency or interpretation of the estimates.

With quantile estimation, however, the inclusion of $\mathbf{x}$ changes the interpretation even when $\mathbf{x}$ is orthogonal to $\mathbf{d}$. Consider a case where each quantile is specified as linear function of the covariates. Since $P(y \leq \mathbf{d}'\gamma(\tau) + \mathbf{x}'\beta(\tau))$ is not necessarily equal to $P(y \leq \mathbf{d}'\tilde{\gamma}(\tau))$, the interpretation of the parameters changes. The quantiles refer to the distribution of the outcome variable conditional on the included covariates. Adding covariates turns some of the unobserved proneness into observed proneness. Since traditional quantile estimators only allow the parameters of interest to vary based on unobserved proneness, the interpretation of the estimates changes. Put differently, the 90^{th} percentile of u is likely different from the

90^{th} percentile of u^*. QR and other conditional QTE estimators are limited by assuming that all variables are "policy variables."

To adopt similar terminology as Chernozhukov and Hansen [2008], the Structural Quantile Function (SQF) of interest for equation (2) is

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

The SQF defines the quantile of the latent outcome variable $y_d = q(\mathbf{d}, u^*)$ 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 $\mathbf{x}$ is unnecessary to generate this counterfactual distribution. The unconditional QTEs are defined by $q(\mathbf{d}'', \tau) - q(\mathbf{d}', \tau)$ for some $\mathbf{d}'', \mathbf{d}'$.

Imbens and Newey [2009] use slightly different notation to discuss SQFs. They discuss conditional quantile models as useful for nonseparable equations such as

$$y = g(\mathbf{d}, \mathbf{x}, \epsilon). \tag{4}$$

The SQF is $S_y(\tau|\mathbf{d}, \mathbf{x}) = g(\mathbf{d}, \mathbf{x}, m_\epsilon(\tau))$ where $m_\epsilon(\tau)$ is the τ^{th} quantile of ϵ. This condition contrasts to the SQF of this paper which, to use similar terminology, is equal to $g(\mathbf{d}, m_{\epsilon^*}(\tau))$ where $\epsilon^* = f(\mathbf{x}, \epsilon)$.

Notice that with a conditional SQF such as $g(\mathbf{d}, \mathbf{x}, \tau)$, generating counterfactual distributions for a given $\mathbf{d}$ requires holding $\mathbf{x}$ constant as well. This technique generates a distribution for $y|\mathbf{d}, \mathbf{x}$ when we are interested in the distribution of $y|\mathbf{d}$. Alternatively, one could vary $\mathbf{x}$ *and* u, but this would require some structural assumptions on the relationship between $\mathbf{x}$ and u. Similarly, QTEs are defined by the difference in the SQF for different values of $\mathbf{d}$ and a fixed $\mathbf{x}$. With the SQF in equation (3), it is possible to generate the distribution of $y|\mathbf{d}$ with no restrictions on the relationship between $\mathbf{x}$ and u^* and to estimate unconditional QTEs.

Conditional quantile estimators such as QR map "unobserved proneness" into quantiles. With some abuse of notation, QR makes the implicit mapping $u \leftrightarrow \tau$. The top quantiles are relevant to observations with high values of the outcome values given all of the policy variables and covariates. More generally, however, we may be interested in estimation techniques which map "total proneness" into quantiles: $u^* \leftrightarrow \tau$. Unconditional QTEs are concerned with exactly this mapping. It should become clear that conditional QTEs are a special case of unconditional QTEs when all variables are simply considered policy variables. Unconditional QTEs are more general because they allow for conditioning on other covariates. This is useful because we are typically interested in how a policy affects the distribution of y, not the distribution of $y|\mathbf{x}$.

Given conditional quantile regression techniques, unconditional QTEs can be estimated by only including the policy variables under the assumption that $u^*|\mathbf{d} \sim U(0,1)$ (Koenker and Bassett [1978]) or, in the IV case, $u^*|\mathbf{z} \sim U(0,1)$ (Chernozhukov and Hansen [2008]). However, these conditions may not be met and conditioning on additional covariates may be necessary for identification purposes. Conditional quantile estimators fail in these cases. I discuss two motivating examples to clarify this terminology further.

2.1 Motivating Examples

2.1.1 Job Training and Earnings

Abadie et al. [2002] estimate conditional quantile treatment effects for job training programs on earnings. They use randomized assignment in the Job Training Partnership Act (JTPA) to understand how job training impacts earnings at different quantiles. The motivation for quantile analysis in this context is to understand how job training affects people of different skill levels in the labor market. Since the treatment is randomly-assigned, it is not necessary to control for other covariates, but it is typical in empirical work to condition on other variables. Abadie et al. [2002] control for variables such as race, high school graduation status, previous work experience, etc. The problem is that these covariates are components of underlying ability to earnings. People who are high school graduates are probably high skilled relative to people without high school degrees. We are likely interested in how job training affects, say, low skilled workers. The Abadie et al. [2002] estimator uses conditional quantiles, defining low quantiles as workers who are low skilled *given their high school graduation status*. This may include people who are relatively high skilled workers because they are more educated.

To illustrate this issue even further, assume that Abadie et al. [2002] have an "ability" variable which is strongly-correlated with the true ability level of each individual. This is useful information and should be used during estimation. Simply controlling for "ability," however, causes problems because it is difficult to interpret the resulting estimates. The high quantiles now refer to individuals with high earnings given their ability measures. These are not necessarily people with high underlying earnings potential as some of these people may be low-earners.

2.1.2 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 = 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

$$T_{it} = \delta_t(\alpha_i + \epsilon_{it}) + v'_{it}\beta(\alpha_i + \epsilon_{it}) \tag{5}$$

The SQF is

$$S_t(\tau|v) = \delta_t(\tau) + v'_{it}\beta(\tau) \tag{6}$$

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. With quantiles, however, differencing 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. In related work, Powell [2009] introduces a panel data estimator which conditions on the individual fixed effect for identification purposes but lets the resulting estimates be interpreted in the same manner as traditional cross-sectional (QR) quantile regression estimates. In other words, the high quantiles refer to observations in the top of the cross-sectional distribution given their policy variables. This estimator allows unconditional QTEs to be estimated in the presence of individual fixed effects. This panel data estimator is a special application of the moment conditions developed below.

3 Existing Literature

A growing literature has extended the use of conditional quantile estimation. For example, Chernozhukov and Hansen [2008] introduce an IV version. Relatedly, Chernozhukov and Hansen [2005] consider identification of conditional QTEs when covariates are endogenous. The identification discussions below are similar to those found in Chernozhukov and Hansen [2005] as many of the conditions are related. Estimation of conditional Structural Quantile Functions is also discussed in Imbens and Newey [2009].

A small literature has also specifically focused on the extension of quantile estimation to panel data. This literature is relevant because of its relationship to Powell [2009] which discusses the merits of using unconditional QTEs in the presence of individual fixed effects. In the second motivating example above, the underlying equation has the form

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

The coefficients of interests are a function of the "total residual," 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 residual. Many existing quantile panel data estimators, however, do not estimate the above equation. Instead, they use a location-shift model, separately estimating α 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.

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

$$y_{it} = \alpha_i + \mathbf{d}'_{it}\gamma(\epsilon_{it}) \tag{8}$$

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

$$y_{it} = \alpha_i(\epsilon_{it}) + \mathbf{d}'_{it}\gamma(\epsilon_{it}) \tag{9}$$

In both cases, the coefficient of interest (γ) varies only with ϵ and not the "total residual." 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 been changed to $S_{y_{it}}(\tau|\mathbf{d}_{it}, \alpha_i) = \alpha_i + \mathbf{d}'_{it}\gamma(\tau)$ or $S_{y_{it}}(\tau|\mathbf{d}_{it}, \alpha_i) = \alpha_i(\tau) + \mathbf{d}'_{it}\gamma(\tau)$ where τ relates to ϵ only.

Powell [2009] introduces an estimator that estimates the relevant SQF (equation (6)) 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). This makes sense. 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.

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 residual (including the fixed effect). Abrevaya and Dahl [2008] introduced 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 a smaller literature on unconditional quantile regression. Firpo et al. [2009] introduce an unconditional quantile regression technique for exogenous variables. They allow the effect of the variables of interest to vary based on the placement in the existing distribution of the outcome variable. While useful in many contexts, it is difficult to write the underlying equation in similar notation as equation (1) since the coefficient of interest is a function of y.

The difficulty is that some observations in the existing distribution are already "treated." The estimator essentially estimates conditional quantiles and integrates over the distribution of all the explanatory variables. This technique lumps together relatively low-ability individuals with policy variables that causally increase y with high-ability individuals without the same policy variables, simply because their current y are the same. However, we are typically interested in how the policy variables affect low-ability individuals and, separately, high-ability individuals.

Just as Koenker and Bassett [1978] and other conditional quantile estimators assume that all variables are "policy variables," the Firpo et al. [2009] estimator also fails to distinguish between policy variables and control variables. This is why the introduction of the terminology in the last section should be useful.

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. In this paper, I discuss identification of unconditional QTEs in a broader context, including multivariate, non-binary variables. Furthermore, my approach is very different and does not involve re-weighting the check function. The moment conditions are very intuitive and provide a natural interpretation. In complementary work, Powell [2010] develops an unconditional quantile estimator using the conditions discussed in this paper.

4 Conditions and Identification

The specification of interest is

$$y = q(\mathbf{d}, u^*), \quad u^* \sim U(0,1). \tag{10}$$

The SQF is

$$q(\mathbf{d}, \tau), \quad \tau \in (0,1). \tag{11}$$

Unconditional QTEs, then, are the change in τ^{th} quantile of y due to a change in $\mathbf{d}$: $q(\mathbf{d}'', \tau) - q(\mathbf{d}', \tau)$ for some $\mathbf{d}'', \mathbf{d}'$.

The motivation for this analysis is the possibility that for $u^* \sim U(0,1)$, $u^*|\mathbf{d} \not\sim$

$U(0,1)$. The policy variables may not be exogenous. Conditioning on another set of covariates may be necessary. Note, however, that $u^*|\mathbf{d},\mathbf{x} \not\sim U(0,1)$ since $\mathbf{x}$ provides information about u^*. This condition is why QR fails in this situation. Instead, exogeneity here is defined as $u^*|\mathbf{d},\mathbf{x} \sim u^*|\mathbf{x}$. In other words, once $\mathbf{x}$ is conditioned on, $\mathbf{d}$ is exogenous. Again, note that QR is a special case. If all variables are policy variables so that we do not have $\mathbf{x}$, then this condition reduces to the QR assumption.

4.1 Exogenous Policy Variables

Define $\mathcal{X}$ as the support of $\mathbf{x}$: $\mathcal{X} \equiv \{\mathbf{x}|P(\mathbf{x}) > 0\}$. Similarly, $\mathcal{D} \equiv \{\mathbf{d}|P(\mathbf{d}) > 0\}$.

The following conditions hold jointly with probability one:

A1 Potential Outcomes and Monotonicity: $y = q(\mathbf{d}, u^*)$ where $q(\mathbf{d}, u^*)$ is increasing in $u^* \sim U(0,1)$.

A2 Independence: $P(u^* \leq \tau|\mathbf{d},\mathbf{x}) = P(u^* \leq \tau|\mathbf{x})$.

A3 Full Rank: For all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{d} \in \mathcal{D}$, either $P(\mathbf{x}|\mathbf{d}) < 1$ or $P(\mathbf{d}|\mathbf{x}) < 1$.

A4 Continuity: y continuously distributed conditional on $\mathbf{d},\mathbf{x}$.

The first assumption (**A1**) is a standard monotonicity condition for quantile estimators. **A2** states that $\mathbf{d}$ does not provide additional information about u^* once $\mathbf{x}$ is conditioned on. The purpose of **A3** is to rule out situations where $\mathbf{d}$ and $\mathbf{x}$ perfectly predict one another. It is equivalent to a full rank condition which requires that there be independent variation in $\mathbf{d}$ and $\mathbf{x}$. It is necessary for identification. **A4** is also necessary for identification and typical in the context of quantile estimators.

It is especially important to note that no restrictions have been placed on the relationship between u^* and $\mathbf{x}$. This is a significant advantage of the approach of this paper.

4.1.1 Moment Conditions

Theorem 4.1 (Moment Conditions)**.** *Suppose **A1** and **A2** hold. Then for each* $\tau \in (0,1)$*,*

$$
\begin{aligned}
P\left[y \leq q(\mathbf{d},\tau)|\mathbf{d},\mathbf{x}\right] &= P\left[y \leq q(\mathbf{d},\tau)|\mathbf{x}\right], && (12)\\
P\left[y \leq q(\mathbf{d},\tau)\right] &= \tau. && (13)
\end{aligned}
$$

Proof of (12):

$$
\begin{aligned}
P\left[y \leq q(\mathbf{d},\tau)|\mathbf{d},\mathbf{x}\right] &= P\left[q(\mathbf{d},u^*) \leq q(\mathbf{d},\tau)|\mathbf{d},\mathbf{x}\right] \qquad \text{by } \mathbf{A1}\\
&= P\left[u^* \leq \tau|\mathbf{d},\mathbf{x}\right] \qquad \text{by } \mathbf{A1}\\
&= P\left[u^* \leq \tau|\mathbf{x}\right] \qquad \text{by } \mathbf{A2}\\
&= P\left[y \leq q(\mathbf{d},\tau)|\mathbf{x}\right] \qquad \text{by } \mathbf{A1}
\end{aligned}
$$

Proof of (13):

$$
\begin{aligned}
P\left[y \leq q(\mathbf{d},\tau)\right] &= P\left[q(\mathbf{d},u^*) \leq q(\mathbf{d},\tau)\right] \qquad \text{by } \mathbf{A1}\\
&= P\left[u^* \leq \tau\right] \qquad \text{by } \mathbf{A1}\\
&= \tau \qquad \text{by } \mathbf{A1}
\end{aligned}
$$

Theorem 4.1 gives us the two moment conditions. Equation (13) holds because the *unconditional* distribution of u^* is uniform. This condition holds for conditional quantile models as well. Equation (12) tells us that once we condition on $\mathbf{x}$, then $\mathbf{d}$ should not provide additional information about $P\left[y \leq q(\mathbf{d},\tau)\right]$. In Powell [2009], $\mathbf{x}$ only includes individual fixed effects, implying the use of within-individual pairwise comparisons for estimation. With two observations with the same values for $\mathbf{x}$, we have observations with the same "observed proneness." Conditioning on the fixed effect provides useful information since variation in $\mathbf{d}$ is more likely to be exogenous to differences in u^* after conditioning on determinants (or correlates) of u^*. In other contexts, this technique is more difficult to implement. These are discussed in Powell [2010].

These moment conditions make intuitive sense. Since u^* is a function of $\mathbf{x}$, conditioning on $\mathbf{x}$ provides information about u^*. Notice that the relationship between u^* and $\mathbf{x}$ is never specified or known. Consequently, the distribution of $u^*|\mathbf{x}$ is unknown. This is unnecessary information. By integrating out $\mathbf{x}$ using the Law of Iterated Expectations, we get condition (13).

The reason these conditions are different from those for conditional quantiles is because we do *not* have the condition $P\left[y \leq q(\mathbf{d}, \tau)|\mathbf{d}, \mathbf{x}\right] = \tau$ for all $\mathbf{d}, \mathbf{x}$. Instead, the above conditions require all comparisons to be made "within-$\mathbf{x}$." The conditional distributions are never specified. Certain covariates may predict, on average, smaller u^*. These conditions allow the covariates to inform the distribution. While we let the SQF vary based on "total proneness," we want to use "observed proneness" for identification by implicitly or explicitly comparing observations with the same observed proneness. Instead of assuming that $\mathbf{d}$ is orthogonal to the entire disturbance, we can relax this and only assume that $\mathbf{d}$ is orthogonal to the unobserved part of the disturbance. $u^*|\mathbf{d}, \mathbf{x} \sim u^*|\mathbf{x}$ can hold even when $u^*|\mathbf{d} \not\sim U(0, 1)$.

4.1.2 Identification

This section discusses the uniqueness of $q(\mathbf{d}, \tau)$ in meeting the moment conditions. Define $\mathcal{D}_{\mathbf{x}} \equiv \{\mathbf{d}|P(\mathbf{d}|\mathbf{x}) > 0\}$. $\mathcal{D}_{\mathbf{x}}$ is the support of $\mathbf{d}$ for a given $\mathbf{x}$.

Lemma 4.1. *If* $P[y \leq q(\mathbf{d}, \tau)|\mathbf{d}, \mathbf{x}] = P[y \leq \tilde{q}(\mathbf{d}, \tau)|\mathbf{d}, \mathbf{x}]$*, then* $q(\mathbf{d}, \tau) = \tilde{q}(\mathbf{d}, \tau)$*.*

This holds under **A4**.

Theorem 4.2 (Identification)**.** *If (a)* $P[y \leq \tilde{q}(\mathbf{d}, \tau)|\mathbf{d}, \mathbf{x}] = P[y \leq \tilde{q}(\mathbf{d}, \tau)|\mathbf{x}]$ *for all* $\mathbf{d}, \mathbf{x}$ *and (b)* $P[y \leq \tilde{q}(\mathbf{d}, \tau)] = \tau$ *and (c)* ***A1**-**A4*** *hold, then for all* $\mathbf{d} \in \mathcal{D}$*,* $\tilde{q}(\mathbf{d}, \tau) = q(\mathbf{d}, \tau)$*.*

Proof. By (a) and equation (12), we know that for all $\mathbf{d} \in \mathcal{D}_{\mathbf{x}}$, $P[y \leq \tilde{q}(\mathbf{d}, \tau)|\mathbf{d}, \mathbf{x}] = P[y \leq q(\mathbf{d}, \theta_{\mathbf{x}})|\mathbf{d}, \mathbf{x}]$ for some $\theta_{\mathbf{x}} \in [0, 1]$. This condition must hold for all $\mathbf{x}$.

A3 implies that $\theta_{\mathbf{x}} = \theta \in (0, 1)$. In words, $\theta_{\mathbf{x}}$ must be equal for all $\mathbf{x}$.[1]

By Lemma 4.1, $\tilde{q}(\mathbf{d}, \tau) = q(\mathbf{d}, \theta)$.

Because of (b), we know that $\theta = \tau$.

□

4.1.3 Relationship to Conditional Quantile Regression

Traditional quantile estimators, such as QR in Koenker and Bassett [1978], estimate conditional QTEs. They do not separately condition on $\mathbf{x}$. It should be clear that the moment conditions are the same when all variables are "policy variables" (i.e. there is no $\mathbf{x}$ to condition on). Combining equations (12) and (13) in this case:

$$P[y \leq q(\mathbf{d}, \tau)|\mathbf{d}] \overset{(12)}{=} P[y \leq q(\mathbf{d}, \tau)] \overset{(13)}{=} \tau. \tag{14}$$

[1]Note that this is *not* saying that $P[y \leq q(\mathbf{d}, \theta_{\mathbf{x}})|\mathbf{d}, \mathbf{x}]$ is the same for all $\mathbf{x}$.

4.1.4 Special Case: Linear Quantiles

Empirical research typically uses linear specifications. Identification in this case means that there are unique values for the coefficient on each variable in $\mathbf{d}$. The same conditions above must hold with the exception of **A3** which can be replaced by:

A3' $P(\mathbf{d}'c = 0|\mathbf{x}) = 1 \quad \Leftrightarrow \quad c = 0$

4.2 Endogenous Policy Variables

Even after conditioning on $\mathbf{x}$, the policy variables may not be exogenous. The conditions necessary for identification are more difficult to describe when the policy variables are endogenous. Instruments are needed and they must have a rich correlation with the entire distribution of the policy variables. In this section, I discuss the conditions when the policy variables and instruments are discrete. The policy variables will be indexed by $k \in \{1, \cdots, K\}$, the instruments by $l \in \{1, \cdots, L\}$ where $L \geq K$. I allow the variables in $\mathbf{x}$ to be continuous or discrete. The appendix discusses the conditions necessary for continuous policy variables.

First, I discuss some notation. I define a matrix of the relationship between $\mathbf{z}$ and $\mathbf{d}$ for a given $\mathbf{x}_m$. Similarly, I define a "stacked" matrix which contains this relationship for multiple values of $\mathbf{x}$.

$$\Pi'(\mathbf{x}_m) \equiv \begin{bmatrix} P(\mathbf{d} = \mathbf{d}_1|\mathbf{z}_1, \mathbf{x}_m) & \cdots & P(\mathbf{d} = \mathbf{d}_K|\mathbf{z}_1, \mathbf{x}_m) \\ \vdots & \ddots & \vdots \\ P(\mathbf{d} = \mathbf{d}_1|\mathbf{z}_L, \mathbf{x}_m) & \cdots & P(\mathbf{d} = \mathbf{d}_K|\mathbf{z}_L, \mathbf{x}_m) \end{bmatrix}, \quad \Pi'(\mathbf{x}_{m(1)}, \cdots, \mathbf{x}_{m(s)}) \equiv \begin{bmatrix} \Pi'(\mathbf{x}_{m(1)}) \\ \vdots \\ \Pi'(\mathbf{x}_{m(s)}) \end{bmatrix}$$

Similarly, I define matrices where the probabilities are not dependent on $\mathbf{z}$:

$$\overline{\Pi'}(\mathbf{x}_m) \equiv \begin{bmatrix} P(\mathbf{d}=\mathbf{d}_1|\mathbf{x}_m) & \cdots & P(\mathbf{d}=\mathbf{d}_K|\mathbf{x}_m) \end{bmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{1\times L} , \quad \overline{\Pi'}(\mathbf{x}_{m(1)}, \cdots, \mathbf{x}_{m(s)}) \equiv \begin{bmatrix} \overline{\Pi'}(\mathbf{x}_{m(1)}) \\ \vdots \\ \overline{\Pi'}(\mathbf{x}_{m(s)}) \end{bmatrix}$$

Finally, I define the following matrices:

$$\Gamma(\mathbf{x}_m) \equiv \begin{bmatrix} P(y \le q(\mathbf{d}_1,\tau)|\mathbf{d}=\mathbf{d}_1,\mathbf{x}_m) \\ \vdots \\ P(y \le q(\mathbf{d}_K,\tau)|\mathbf{d}=\mathbf{d}_K,\mathbf{x}_m) \end{bmatrix}, \quad \Gamma(\mathbf{x}_{m(1)}, \cdots, \mathbf{x}_{m(s)}) \equiv \begin{bmatrix} \Gamma(\mathbf{x}_{m(1)}) \\ \vdots \\ \Gamma(\mathbf{x}_{m(s)}) \end{bmatrix}$$

This notation is helpful since we are interested in $P\left[y \le q(\mathbf{d},\tau)|\mathbf{z},\mathbf{x}_m\right]$ for all $\mathbf{z}$ which can be written as $\Pi'(\mathbf{x}_m)\Gamma(\mathbf{x}_m)$. Finally, define $\mathcal{Z}_{\mathbf{x}} \equiv \{\mathbf{z}|P(\mathbf{z}|\mathbf{x}) > 0\}$, $\mathcal{Z} \equiv \{\mathbf{z}|P(\mathbf{z}) > 0\}$.

The following conditions hold jointly with probability one:

IV-A1 Potential Outcomes and Monotonicity: $y = q(\mathbf{d}, u^*)$ where $q(\mathbf{d}, u^*)$ is increasing in $u^* \sim U(0,1)$.

IV-A2 Independence: $P(u^* \le \tau|\mathbf{z},\mathbf{x}) = P(u^* \le \tau|\mathbf{x})$.

IV-A3 First Stage: For some $(\mathbf{x}_{m(1)}, \cdots, \mathbf{x}_{m(s)})$, $\Pi'(\mathbf{x}_{m(1)}, \cdots, \mathbf{x}_{m(s)})$ is full rank.

IV-A4 Full Rank: For all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{d} \in \mathcal{D}$, either $P(\mathbf{x}|\mathbf{d}) < 1$ or $P(\mathbf{d}|\mathbf{x}) < 1$.

IV-A5 Continuity: y continuously distributed conditional on $\mathbf{z}, \mathbf{x}$.

The assumptions are generally the same as before. The main difference is that, even conditional on $\mathbf{x}$, $\mathbf{d}$ can be endogenous. The underlying assumption now is that $u^*|\mathbf{z},\mathbf{x} \sim u^*|\mathbf{x}$. **IV-A3** is a "first stage" assumption that states that $\mathbf{z}$ must impact the entire distribution of $\mathbf{d}$. This must hold for some set of values for $\mathbf{x}$. As before, no restrictions have been placed on the relationship between $\mathbf{x}$ and u^*. The moment conditions are also similar to the exogenous case:

Theorem 4.3 (IV Moment Conditions)**.** *Suppose **IV-A1** and **IV-A2** hold. Then for each* $\tau \in (0,1)$*,*

$$P\left[y \leq q(\mathbf{d},\tau)|\mathbf{z},\mathbf{x}\right] = P\left[y \leq q(\mathbf{d},\tau)|\mathbf{x}\right], \tag{15}$$

$$P\left[y \leq q(\mathbf{d},\tau)\right] = \tau. \tag{16}$$

Equation (15) implies that $\left[\Pi'(\mathbf{x}_m) - \overline{\Pi'}(\mathbf{x}_m)\right]\Gamma(\mathbf{x}_m) = 0$ for all $\mathbf{x}_m$. Again, the defining feature is that $P\left[y \leq q(\mathbf{d},\tau)|\mathbf{z},\mathbf{x}\right]$ is unknown and allowed to vary based on $\mathbf{x}$. The instrument $\mathbf{z}$ is exogenous after conditioning on $\mathbf{x}$.

4.2.1 Identification

We need the IV equivalent to Lemma 4.1, which holds due to **IV-A5**:

Lemma 4.2. *If* $P\left[y \leq q(\mathbf{d},\tau)|\mathbf{z},\mathbf{x}\right] = P\left[y \leq \tilde{q}(\mathbf{d},\tau)|\mathbf{z},\mathbf{x}\right]$*, then* $q(\mathbf{d},\tau) = \tilde{q}(\mathbf{d},\tau)$*.*

Theorem 4.4 (IV Identification)**.** *If (a)* $P\left[y \leq \tilde{q}(\mathbf{d},\tau)|\mathbf{z},\mathbf{x}\right] = P\left[y \leq \tilde{q}(\mathbf{d},\tau)|\mathbf{x}\right]$ *for all* $\mathbf{z},\mathbf{x}$ *and (b)* $P\left[y \leq \tilde{q}(\mathbf{d},\tau)\right] = \tau$ *and (c) **IV-A1** - **IV-A5** hold, then for all* $\mathbf{d} \in \mathcal{D}$*,* $\tilde{q}(\mathbf{d},\tau) = q(\mathbf{d},\tau)$*.*

Proof. $P\left[y \leq \tilde{q}(\mathbf{d},\tau)|\mathbf{z},\mathbf{x}\right] = P\left[y \leq \tilde{q}(\mathbf{d},\tau)|\mathbf{x}\right]$ implies that for all $\mathbf{x}_m, \Pi'(\mathbf{x}_m)\tilde{\Gamma}(\mathbf{x}_m) = \overline{\Pi'}(\mathbf{x}_m)\tilde{\Gamma}(\mathbf{x}_m)$.

By **IV-A3**, then, $P\left[y \leq \tilde{q}(\mathbf{d}_k, \tau) | \mathbf{d} = \mathbf{d}_k, \mathbf{x}\right] = \theta_{\mathbf{x}}$ for some $\theta_{\mathbf{x}} \in [0, 1]$ which is constant for all k.

By **IV-A4**, we know that $\theta_{\mathbf{x}} = \theta \in (0, 1)$ for all $\mathbf{x}$.

By Lemma 4.2, $\tilde{q}(\mathbf{d}, \tau) = q(\mathbf{d}, \theta)$ for all $\mathbf{d}$.

Because of (b), we know that $\theta = \tau$.

□

This section shows that $q(\mathbf{d}, \tau)$ is unique. See the appendix for a discussion about conditions for continuous variables.

4.3 Discussion

The conditions discussed above are very similar to those found in Chernozhukov and Hansen [2005] which considers identification of conditional quantile treatment effects. In fact, the assumptions necessary to identify unconditional QTEs are arguably much less restrictive than those necessary for conditional QTEs since no assumptions must be made regarding the exogeneity of $\mathbf{x}$. Instead, the main difference for identification is the addition of a moment condition. This condition uses conditional distributions for identification without affecting the unconditional distribution.

The defining aspect of any quantile model is the equation $P\left[y \leq q(\mathbf{d}, \tau)\right] = \tau$. Conditional quantile models are restrictive. For a SQF including both policy variables and control variables - i.e. $g(\mathbf{d}, \mathbf{x}, \tau)$ - these models require $P\left[y \leq g(\mathbf{d}, \mathbf{x}, \tau) | \mathbf{z}, \mathbf{x}\right] = \tau$ (where it is possible that $\mathbf{z} = \mathbf{d}$ for exogenous policy variables) to hold for *all* $\mathbf{z}, \mathbf{x}$. Identification of unconditional QTEs only requires the unconditional distribution to meet this condition. Instead, $\mathbf{z}$ and $\mathbf{x}$ are allowed to inform the distribution of the disturbance. An additional moment condition is needed, $P\left[y \leq q(\mathbf{d}, \tau)|\mathbf{z}, \mathbf{x}\right] = P\left[y \leq q(\mathbf{d}, \tau)|\mathbf{x}\right]$. This condition looks "within-$\mathbf{x}$" for identification. Importantly, no assumptions on the relationship between $\mathbf{x}$ and u^* are required. Conditional QTEs require additional assumptions on this relationship.

5 Conclusion

This paper discusses unconditional quantile treatment effects and the conditions under which they are identified. Unconditional QTEs are useful to characterize the impact of policy variables on the distribution of the unconditional outcome variable. I discuss identification for both exogenous and endogenous policy variables. Conditional quantile estimators can be used to estimate unconditional QTEs by not including additional covariates. However, sometimes identification requires conditioning on additional covariates or it might simply be desirable. This paper develops moment conditions for unconditional quantiles that are intuitive and allow for an arbitrary relationship between the disturbance and the control variables. Interestingly, this fact implies that unconditional QTEs require less restrictive assumptions than the equivalent conditional QTEs. Powell [2009] and Powell [2010] discuss estimation of unconditional quantiles in different contexts using the moment conditions discussed in this paper.

A Appendix: Identification with Continuous Endogenous Policy Variables

While the general intuition of the discrete variable case is valid, the conditions for identification with continuous policy variables are more complicated. I replace IV-A3 with

IV-A3' If $\int_{\mathbf{d}\in\mathcal{D}_{\mathbf{x}}} a(\mathbf{d}|\mathbf{x})\,[f(\mathbf{d}|\mathbf{z},\mathbf{x}) - f(\mathbf{d}|\mathbf{x})]\,d\mathbf{d} = 0$ for all $\mathbf{z} \in \mathcal{Z}_{\mathbf{x}}$, then $a(\mathbf{d}|\mathbf{x}) = a(\mathbf{x})$.

This condition is a "first stage" condition, requiring $\mathbf{z}$ to provide rich variation in $\mathbf{d}$ conditional on $\mathbf{x}$. **IV-A3'** states that $a(\mathbf{d}|\mathbf{x})$ must be constant for all $\mathbf{d} \in \mathcal{D}_{\mathbf{x}}$. With this assumption, we know that $P\,[y \leq \tilde{q}(\mathbf{d}_k, \tau)|\mathbf{d} = \mathbf{d}_k, \mathbf{x}] = \theta_{\mathbf{x}}$ for some $\theta_{\mathbf{x}} \in [0, 1]$ which is constant for all k.

The rest of the identification proof follows as before.

References

Alberto Abadie, Joshua Angrist, and Guido Imbens. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70(1):91–117, January 2002. URL `http://ideas.repec.org/a/ecm/emetrp/v70y2002i1p91-117.html`.

Jason Abrevaya and Christian M Dahl. The effects of birth inputs on birthweight. *Journal of Business & Economic Statistics*, 26:379–397, 2008. URL `http://ideas.repec.org/a/bes/jnlbes/v26y2008p379-397.html`.

Ivan A. Canay. A note on quantile regression for panel data models. Working paper, Northwestern Economics Department, May 2010.

Victor Chernozhukov and Christian Hansen. An iv model of quantile treatment effects. *Econometrica*, 73(1):245–261, 2005. URL `http://ideas.repec.org/a/ecm/emetrp/v73y2005i1p245-261.html`.

Victor Chernozhukov and Christian Hansen. Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, 142(1):379–398, January 2008. URL `http://ideas.repec.org/a/eee/econom/v142y2008i1p379-398.html`.

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. URL `http://ideas.repec.org/p/ifs/cemmap/29-09.html`.

Kjell Doksum. Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.*, 2(2):267–277, 1974. URL `http://projecteuclid.org/euclid.aos/1176342662`.

Sergio Firpo. Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 75(1):259–276, 01 2007. URL `http://ideas.repec.org/a/ecm/emetrp/v75y2007i1p259-276.html`.

Sergio Firpo, Nicole M. Fortin, and Thomas Lemieux. Unconditional quantile regressions. *Econometrica*, 77(3):953–973, 05 2009. URL `http://ideas.repec.org/a/ecm/emetrp/v77y2009i3p953-973.html`.

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. URL `http://ideas.repec.org/a/eee/ecolet/v104y2009i3p133-135.html`.

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

Roger Koenker. Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89, October 2004. URL `http://ideas.repec.org/a/eee/jmvana/v91y2004i1p74-89.html`.

Roger W Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, January 1978. URL `http://ideas.repec.org/a/ecm/emetrp/v46y1978i1p33-50.html`.

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 regression for panel data with exogenous or endogenous regressors. Working Papers 710, RAND Corporation Publications Department, October 2009. URL http://ideas.repec.org/p/ran/wpaper/710.html.

David Powell. Unconditional quantile regression in the presence of covariates. Working papers, RAND Corporation Publications Department, 2010.

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. URL http://ideas.repec.org/a/tpr/qjecon/v113y1998i2p553-602.html.