A New Framework for Estimation of Quantile Treatment Effects

Nonseparable Disturbance in the Presence of Covariates

David Powell

RAND Labor & Population

WR-824-1

January 2013

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 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.

A New Framework for Estimation of Quantile Treatment Effects: Nonseparable Disturbance in the Presence of Covariates*

David Powell[†]

RAND

January 24, 2013

Abstract

This paper introduces a new framework for quantile estimation. Quantile regression techniques have proven to be extremely valuable in understanding the relationship between explanatory variables and the conditional distribution of the outcome variable. Quantile regression allows the effect of the explanatory variables to vary based on a nonseparable disturbance term, frequently interpreted as “unobserved proneness” for the outcome, and provides conditional quantile treatment effects. Researchers are typically interested in the impact of the treatment variables on the unconditional distribution of the outcome. Additional covariates may be necessary (or simply desirable) for identification but adding these variables alters the interpretation of the resulting estimates as some of the “unobserved proneness” becomes observed and the disturbance term is separated. The Generalized Quantile Regression (GQR) estimator provides unconditional quantile treatment effects - the impact of the treatment variables on the unconditional distribution of the outcome variables. The control variables are conditioned on for identification or variance reduction but without altering the interpretation of the estimates. This property parallels mean regression. An IV version (IV-GQR) is also introduced. The estimator is extremely straightforward to implement using standard statistical software. Quantile Regression and Instrumental Variables Quantile Regression are special cases of the introduced estimation technique, but the proposed technique provides additional flexibility in the estimation of quantile treatment effects.

Keywords: Quantile Treatment Effects, Quantile Regression, Nonseparable disturbance

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 Abby Alpert, Helen Hsi, Erik Meijer, Pierre-Carl Michaud, Kathleen Mullen, Amanda Pallais, and Hui Shan.

[†]dpowell@rand.org

1 Introduction

Mean regression methods such as ordinary least squares (OLS) provide the relationship between the explanatory variables and the *unconditional* population mean. Many empirical applications have found quantile regression methods useful because they provide richer evidence than mean regression estimators by allowing for heterogenous effects and estimating the distributional impacts of the explanatory variables. However, traditional quantile estimation techniques provide estimates of the relationship between the covariates and the *conditional* distribution of the outcome variable. These estimates cannot be used to understand the impact on the unconditional outcome distribution. This paper introduces a new framework and a corresponding estimation technique for the estimation of quantile treatment effects to understand the impact of treatment variables on the unconditional outcome distribution.

Quantile estimators, such as the Koenker and Bassett [1978] quantile regression (QR), are useful for the estimation of conditional quantile treatment effects (QTEs). QR allows the relationship between the explanatory variables and the outcome variable to vary based on a nonseparable disturbance term. This disturbance term is frequently interpreted as “unobserved proneness” for the outcome variable (Doksum [1974]). As covariates are added, some of the unobserved proneness becomes observed and, consequently, the interpretation of the estimates from a quantile regression changes. This paper introduces a new framework for quantile estimation which divides the explanatory variables into “treatment variables” and “control variables.” The estimator produces unconditional quantile treatments effects - the impact of the treatment variables on the distribution of the outcome variable. The control variables are used to aid identification in a way that parallels mean regression. I introduce a Generalized Quantile Regression (GQR) technique and an instrumental variable version (IV-GQR) which are extremely straightforward to implement using standard statistical software. I use the term “generalized” because QR and IV-QR are special cases of GQR. The framework introduced in this paper allows for more flexibility in estimating quantile treatment effects.

Conditional QTEs describe the effect of treatment variables on the conditional distribution of the outcome variable. While conditional QTEs can be useful, we may be interested in *unconditional* quantile treatment effects. Unconditional QTEs describe the difference in the quantiles for different values of the treatment variables without reference to the control

variables. A simple way to obtain unconditional QTEs with QR is simply to not control for the other covariates. A problem arises, however, when it is necessary or simply desirable to condition on a separate set of covariates for the purposes of identification. With mean regression, adding covariates does not change the interpretation of the results since the estimates can still be interpreted as the effect on the unconditional mean. This paper provides a quantile estimator with an equivalent property.

Empirical researchers are frequently interested in the impact of changes in treatment variables on the distribution of the outcome variable such as how job training affects the distribution of labor earnings. Conditioning on additional variables such as measures of labor market ability may be necessary or simply desirable for this estimation. With mean regression, it is straightforward to add these covariates. With QR, these control variables alter the interpretation of the parameters of interest. The quantile estimates vary based only on unobserved labor market ability and conditioning on observed labor market ability alters this disturbance term. The “high quantile” estimates refer to the impact of job training on individuals with high ability given their *observed* labor market ability. Many of these individuals may have low overall labor market ability. Put differently, observations at the top of the conditional distribution may be at the bottom of the unconditional distribution.

Researchers frequently include additional covariates for the purposes of identification, but the desired interpretation of the parameters of interest does not change. The generalized quantile regression (GQR) method introduced in this paper estimates unconditional QTEs while conditioning on a separate set of covariates. Unlike conditional quantile estimators which assume that all treatment variables or instruments are orthogonal to “proneess,” GQR allows the covariates to inform the distribution of the disturbance. The estimation technique can be used for exogenous (GQR) and endogenous (IV-GQR) treatment variables. This framework is very flexible as it allows the researcher to choose which variables are treatment variables and which are control variables. A primary motivation of this paper is the frequency with which conditional QTEs are interpreted as the effect of the treatment variables on the unconditional distribution. The prevalence of this interpretation suggests demand for a more flexible framework.

This estimator is useful for models such as

$$y_i = \mathbf{d}_i' \beta(u_i^*), \quad u_i^* \sim U(0, 1), \quad (1)$$

where $\mathbf{d}_i$ represents the treatment variables and u_i^* is the disturbance or unobserved rank variable, interpreted by Doksum [1974] as underlying proneness or ability. QR requires the assumption $u_i^*|\mathbf{d}_i \sim U(0,1)$ to estimate a Structural Quantile Function (SQF). This paper relaxes this assumption and replaces it with $u_i^*|\mathbf{d}_i, \mathbf{x}_i \sim u_i^*|\mathbf{x}_i$ where $\mathbf{x}_i$ represents the control variables. These covariates inform the distribution of the disturbance. If we are interested in understanding how job training affects the distribution of earnings, then the disturbance represents an individual’s underlying earnings potential or ability in the labor market. If we also have information about the individual’s education level, we should use this information during estimation. However, simply including “education” in QR changes the interpretation of the estimates as some of the unobserved ability becomes observed. We are likely interested in the distribution of y given job training status, but QR defines “high quantiles” as observations with high earnings given job training status and education. This may include people near the bottom of the earnings distribution.

Traditional quantile estimators rely on conditions such as $P[y_i \leq \mathbf{d}_i'\beta(\tau)|\mathbf{d}_i] = \tau$. The GQR estimator relaxes this condition, replacing it with $P[y_i \leq \mathbf{d}_i'\beta(\tau)|\mathbf{d}_i, \mathbf{x}_i] = P[y_i \leq \mathbf{d}_i'\beta(\tau)|\mathbf{x}_i] \equiv \tau_{\mathbf{x}_i}$. The covariates are allowed to inform the probability that y is smaller than the quantile function. This is the benefit of conditioning on additional covariates, paralleling the mean regression case. Instead of assuming that the probability that the outcome is below the quantile function is the same for each observation, the proposed framework allows this probability to vary based on covariates. A highly-educated individual is more likely to be in the top part of the earnings distribution so conditioning on this information is valuable. These conditional distributions are used for identification, but the parameters refer to the unconditional distribution since $P[y_i \leq \mathbf{d}_i'\beta(\tau)] = \tau$ (i.e., $E[\tau_{\mathbf{x}_i}] = \tau$).

Simulations in section 6 illustrate the usefulness of the GQR estimator. Even with randomly-assigned treatment variables, the GQR estimator performs better than the QR estimator. This is not surprising given that GQR uses more information. I also estimate unconditional QTEs using data originally analyzed by Abadie et al. [2002]. I look at the distribution impact of additional school days as well using data from Duflo et al. [2012]. In both applications, the estimated unconditional QTEs suggest different distributional effects than the conditional QTEs. The GQR estimator is easy to implement with standard statistical software, arguably easier than even QR.

In the next section, I discuss general terminology and related literature. Section 3 introduces the model and GQR estimator for the exogenous policy variable case. Section

4 introduces the IV equivalent. Section 5 looks at consistency, asymptotic normality, and inference of the estimators. Section 6 includes simulations and two applications. Section 7 concludes.

2 Unconditional Quantile Treatment Effects with Covariates

2.1 General Concepts

This paper introduces a broader terminology to discuss quantile treatment effects. The underlying equation has a nonseparable disturbance. I focus on linear quantiles in this paper:

$$y_i = \mathbf{d}_i' \beta(u_i^*), \quad u_i^* \sim U(0, 1). \quad (2)$$

Chernozhukov and Hansen [2008] introduce the Structural Quantile Function (SQF). For my purposes, I will write the SQF as

$$S_{y_i} = \mathbf{d}_i' \beta(\tau), \quad \tau \in (0, 1). \quad (3)$$

The SQF defines the τ^{th} quantile of $y_{\mathbf{d}} = \mathbf{d}'\beta(u^*)$ for a randomly-selected $u^* \sim U(0, 1)$. Importantly, the τ in the SQF refers to the τ^{th} quantile of u^* , the total disturbance. In many cases, identification may require conditioning on $\mathbf{x}$, a separate set of covariates. Let $u \sim U(0, 1)$ represent “unobserved proneness,” such that $u^* = f(\mathbf{x}, u)$. In words, “total proneness” is a function of observed and unobserved proneness.

I contrast this framework with traditional conditional linear quantile estimators. Simply including the covariates in a quantile regression assumes that the underlying equation is

$$y_i = \mathbf{d}_i' \beta(u_i) + \mathbf{x}_i' \delta(u_i), \quad u_i \sim U(0, 1), \quad (4)$$

where one assumes $u_i | \mathbf{d}_i, \mathbf{x}_i \sim U(0, 1)$. In the corresponding SQF, τ refers to the τ^{th} quantile of u . The interpretation has changed. Simply adding $\mathbf{x}$ to QR implicitly alters the parameters of interest since the τ^{th} quantile of u is likely not equal to the τ^{th} quantile of u^* . The primary motivation of this paper is to estimate the SQF represented by equation (3) even when $u^* | \mathbf{d} \not\sim U(0, 1)$ but additional covariates are available.

Note, however, that even if $\mathbf{d}$ were randomly-assigned, conditioning on $\mathbf{x}$ in a conditional quantile framework is problematic. When $u^*|\mathbf{d} \sim U(0, 1)$, conditioning on $\mathbf{x}$ is not necessary. But, it is typical in empirical work to condition on other covariates and it should be beneficial to use this information during estimation. Conditioning on $\mathbf{x}$ with QR (or IV-QR) can produce biased estimates of equation (3) even when the policy variables are randomly-assigned.

Quantile estimates allow the parameters to vary based on a nonseparable disturbance. Covariates are used by researchers in order to explain some of the unobserved proneness for the outcome variables because this is necessary for identification, because it reduces the chances of bias due to spurious correlation, or because it reduces the variance of the estimates. Simply adding covariates in a conditional quantile framework specifies and separates the disturbance term, possibly counteracting the original intent of implementing a quantile estimator.

With mean regression, adding more control variables relaxes the identification assumption without changing the interpretation of the estimates. This paper introduces a quantile estimator equivalent. The GQR estimator allows one to include additional covariates while still estimating the SQF of interest (equation (3)). Covariates can be added, but the parameters of interest are still assumed to vary based on “total proneness” (u^*). Using the covariates allows one to relax the assumption that $u^*|\mathbf{d} \sim U(0, 1)$.

2.2 Treatment Variables vs. Control Variables

This framework uses the treatment variables and control variables differently. Instead of assuming that all variables are policy variables, the framework provides more flexibility than traditional quantile estimators by allowing the researcher to decide which variables belong in each category. The treatment variables are the variables that one wants to include in the SQF. These are the policy variables that we are interested in understanding how they shift the distribution of the outcome variable. We can imagine these variables being manipulated and predict how the outcome distribution will subsequently change.

The control variables are partial determinants of the rank in the distribution and are excluded from the SQF. Thus, once the SQF is estimated, the outcome distribution can be simulated for different values of the treatment variables, without reference to any of the control variables. I use “unconditional quantile treatment effects” to refer to the impact of

the treatment variables on the outcome at a certain quantile, unconditional on the control variables.

The distinction between the two sets of variables is likely best understood through equation (3). The treatment variables are included in the SQF. The control variables will be used to help estimate the SQF. In the presence of covariates, QR cannot be used to estimate equation (3).

The GQR estimator is a more general form of QR because it allows one to choose which variables are policy variables ($\mathbf{d}$) and which variables are covariates ($\mathbf{x}$). If all variables are considered policy variables, then the estimation technique of this paper reduces to QR (or IV-QR). Thus, QR and IV-QR are still available in this framework. The advantage of this framework, then, is that it provides flexibility to the researcher, allowing for the estimation of QTEs for whichever variables the researcher wants to consider the “treatment variables” while not requiring the researcher to throw out valuable information contained in the remaining control variables.

2.3 Motivating Example: 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. These covariates are components of underlying ability in the labor market. 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” in a linear quantile regression 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.

2.4 Motivating Example: Number of School Days and Student Achievement

Duflo et al. [2012] use a randomized experiment to study the impact of monitoring and incentives on teacher shirking. They find that monitoring and incentives increased teacher attendance and the number of days that schools are open. In one specification, they study the mean impact of the number of school days on student achievement, instrumenting the number of school days by the interventions. They also condition on three separate pre-test scores for each student. The distributional impacts of the number of school days are also interesting. Does an increase in the number of school days have a greater impact on low-achieving students or high-achieving students? Including controls for student achievement in the form of pre-test scores is problematic with IV-QR because we are interested in the impact on the unconditional achievement distribution. Controlling for measures of student achievement changes the interpretation of the results. Instead of providing estimates of the impact on low-achieving students, IV-QR estimates with controls must be interpreted as the effect on students that are low-achieving relative to their observed levels of achievement. Many of these students may be high-achieving students.

2.5 Illustrative Simulation

The difference between conditional and unconditional quantiles can also be seen graphically by comparing the disturbance terms. I generate a normally-distributed control variable x (eg., proxy for education) and a normally-distributed “unobserved disturbance” term ϵ (eg., unobserved labor market ability). I define the “total disturbance” (eg., observed and unobserved labor market ability) as $(x + \epsilon)$. We are interested in how job training affects the labor earnings based on the underlying total labor market ability of the individual. Conditional quantile estimators only allow the parameters to vary based on unobserved ability while we want the parameters to vary based on total ability. To illustrate the difference, I graph the relationship between the quantile of the “unobserved disturbance” and the distribution (in

quantiles) of the “total disturbance.” For a given value of the “unobserved disturbance,” there is a wide range of possible values for overall labor market ability. Figure 1 graphs the 95% confidence interval of the “total disturbance” term for a given “unobserved disturbance” value. Imagine estimating a traditional quantile regression for the 5th conditional quantile because of interest in the effect of job training on the bottom on the earnings distribution. In a conditional quantile framework, this 5th quantile refers to a group of people with a wide range of labor market ability, include people with above-median labor market ability. This property makes it impossible to interpret the estimates as the effect of job training on the bottom of the unconditional earnings distribution. The framework of this paper provides guidance for studying the 5th quantile of the unconditional distribution.

Figure 1: 95% Confidence Interval of “Total Disturbance” Term

2.6 Related Literature

Quantile estimation was introduced by Koenker and Bassett [1978] and a growing literature has extended it to new contexts. Chernozhukov and Hansen [2008] discuss IV estimation of conditional quantile treatment effects when the explanatory variables are endogenous.

A small literature considers estimation of QTEs with panel data and fixed effects. This literature is discussed in Powell [2012]. I address the literature here because it is a

special but insightful example of the usefulness of the proposed framework. Many quantile panel estimators, such as those introduced by Koenker [2004] and Harding and Lamarche [2009], use location-shift models which separately account for an individual fixed effect (α_i). The coefficients of interest vary based only on the observation-specific part of the disturbance (u_{it}). Other existing quantile estimators for panel data also include separate terms for the fixed effect. These include Canay [2010], Galvao Jr. [2011], and Ponomareva [2010]. Powell [2012] introduces an estimator for equations with a nonseparable disturbance such as

$$y_{it} = \mathbf{d}'_{it}\beta(u_{it}^*). \quad (5)$$

where $u_{it}^* = f(\alpha_i, u_{it})$. The coefficients vary based on the total disturbance. The Powell [2012] estimator can be thought of as a special case of the GQR estimator. Researchers frequently employ panel data and fixed effects because they do not believe that the parameters of interest are identified with cross-sectional data. They do not necessarily want to change the interpretation of their estimates. A location-shift model assumes a SQF of $S_{y_{it}} = \alpha_i + \mathbf{d}'_{it}\beta(\tau)$ or $S_{y_{it}} = \alpha_i(\tau) + \mathbf{d}'_{it}\beta(\tau)$ where τ refers to the τ^{th} quantile of u_{it} .

The Powell [2012] estimator assumes a SQF

$$S_{y_{it}} = \mathbf{d}'_{it}\beta(\tau),$$

where τ refers to the τ^{th} quantile of u_{it}^* . Fixed effects are conditioned on for identification purposes only, allowing for an arbitrary correlation between the fixed effects and the policy variables (or instruments). Thus, the estimates can be interpreted in the same manner as cross-sectional quantile estimates. GQR is a more general version of the Powell [2012] estimator. In Powell [2012], the individual fixed effects are the “control variables” and all other variables are “treatment variables.”

There is a smaller literature on unconditional quantile regression. Firpo et al. [2009] introduce an unconditional quantile regression (UQR) technique for exogenous variables in an especially insightful paper. The motivation for the Firpo et al. [2009] estimator is similar to the one discussed in this paper. I see four major differences between GQR and UQR.

First, 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 does not distinguish between policy variables and control variables. Instead, the Firpo et al. [2009] estimator treats all variables in the same way that I will treat the “control variables.”

Second, the UQR estimator is useful for estimating the impact of changes in the covariates on the existing distribution of the outcome variable. The existing distribution, however, is already “treated.” This technique lumps together relatively low-ability individuals with policy variables that causally increase the outcome variable with high-ability individuals without the same policy variables. The framework introduced in this paper assumes that we are interested in how the policy variables affect low-ability individuals and, separately, high-ability individuals. It is difficult to write down an equivalent SQF in the UQR framework.

Third, the UQR estimator assumes all variables are exogenous while GQR is straightforward to extend to an IV context. Fourth, the UQR estimator relies on kernel density estimation. The GQR estimator is relatively easy to implement and does not require any choice of bandwidth or correct nonparametric estimation of the density of the outcome variable.

The GQR estimator introduced below can be seen as a combination of the Koenker and Bassett [1978] estimator and the Firpo et al. [2009] estimator. The caveat is that I only derive estimates for the treatment variables and do not study the relationship between the control variables and rank in the outcome distribution. I do this because the control variables are simply determinants of placement in the distribution, which are typically less interesting than the impact of the policy variables. Future work can use the Firpo et al. [2009] technique to also derive and interpret the effects of the control variables.

Note that there are also proposed estimators which estimate the conditional quantiles and then integrate over the variables to obtain unconditional quantiles (eg., Mata and Machado [2005], Melly [2005]). These estimators are also different from GQR because they assume that the parameters of interest vary based only on unobserved proneness (u) and then integrate to understand the ramifications of these parameters on the unconditional distribution. In other words, these estimators assume that a high education person and a low education person with the same unobserved ability will experience the same impact from job training. GQR does not enforce this assumption.

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. I discuss estimation of unconditional QTEs in a broader context, including multivariate, non-binary variables. Furthermore, my approach is different and does not involve re-weighting the check function. The moment

conditions are very intuitive and provide a natural interpretation.

3 Model

The specification of interest is

$$y_i = \mathbf{d}_i' \beta(u_i^*), \quad u_i^* \sim U(0, 1). \quad (6)$$

The motivation of this paper is that for situations where $u_i^* | \mathbf{d}_i \not\sim U(0, 1)$, QR does not provide consistent estimates. Note also that since $\mathbf{x}_i$ provides information about u_i^* , $u_i^* | \mathbf{d}_i, \mathbf{x}_i \not\sim U(0, 1)$, even when $\mathbf{d}_i$ is randomly-assigned. Instead, exogeneity is defined differently. In words, the exogeneity assumption is that the policy variables do not provide additional information about u_i^* once we condition on $\mathbf{x}_i$. Define

$$\tau_{\mathbf{x}_i} \equiv P(y_i \leq \mathbf{d}_i' \beta(\tau) | \mathbf{x}_i). \quad (7)$$

The probability that y_i is less than $\mathbf{d}_i' \beta(\tau)$ is allowed to vary based on $\mathbf{x}_i$ in this formulation. The control variables provide information about “proneess.” A highly-educated person is more likely to have a higher labor earnings, affecting this probability. Contrast this with QR which requires this probability to always equal τ . In practice, it is necessary to estimate equation (7) with

$$\hat{\tau}_{\mathbf{x}_i} \equiv \hat{P}(y_i \leq \mathbf{d}_i' \beta(\tau) | \mathbf{x}_i).$$

Estimation error is inevitable in most contexts and I discuss below how such errors are not necessarily problematic. Define this difference as

$$\nu_i \equiv \hat{\tau}_{\mathbf{x}_i} - \tau_{\mathbf{x}_i}.$$

While relying on conditional distributions for identification, the unconditional distribution is unchanged. The estimation technique also uses

$$P(y_i \leq \mathbf{d}_i' \beta(\tau)) = \tau.$$

This condition is equivalent to

$$E [P(y_i \leq \mathbf{d}'_i \beta(\tau) | \mathbf{x}_i)] = \tau \quad (\text{i.e., } E[\tau_{\mathbf{x}_i}] = \tau)$$

The estimator conditions on $\mathbf{x}$ for identification purposes. However, it implicitly integrates out the covariates so that the parameters refer to the unconditional distribution.

3.1 Assumptions

The following conditions hold jointly with probability one:

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

A2 Conditional Independence:

$$(a) \quad P(u_i^* \leq \tau | \mathbf{d}_i, \mathbf{x}_i) = P(u_i^* \leq \tau | \mathbf{x}_i).$$

$$(b) \quad E[\mathbf{d}_i \nu_i] = 0.$$

A3 Full Rank: If $P(\mathbf{d}'c = 0 | \mathbf{x}_i) = 1$ for all $\mathbf{x}_i$, then $c = 0$.

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

The first assumption (**A1**) is a standard monotonicity condition for quantile estimators (see Chernozhukov and Hansen [2008] for one example). I model the disturbance term as a uniform variable. This is standard in this literature and simply a normalization.

A2 is an independence assumption with two parts. The first part states that the policy variables do not provide additional information about the disturbance once the covariates are conditioned on. Variation in $\mathbf{d}$ cannot predict heterogeneity in $u^* | \mathbf{x}$. Note that this relaxes the standard assumption $P(u_i^* \leq \tau | \mathbf{d}_i) = \tau$.

The second part is specific to the estimation of $\tau_{\mathbf{x}_i}$. Estimation error is not necessarily problematic as long as that error is orthogonal to the policy variables, an assumption which should not be restrictive in most contexts. In practice, the difference $\hat{\tau}_{\mathbf{x}_i} - \tau_{\mathbf{x}_i}$ may be due to misspecification, poor distributional assumptions, and simple estimation error. However, it is not clear that the policy variables should be correlated with these errors. The estimation discussion below will further consider the merits of this assumption. Practically, concerns with this assumption should motivate the choice of method to estimate $\tau_{\mathbf{x}_i}$ in such a way

that this assumption is plausibly met.

A3 is a full rank condition that requires variation in the policy variables that is independent of the covariates. **A4** is necessary for identification and typical in the context of quantile estimators.

3.2 Moment Conditions

These assumptions lead to two separate moment conditions.

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

$$E \left\{ \mathbf{d}_i \left[\mathbf{1}(y_i \leq \mathbf{d}_i' \beta(\tau)) - \hat{\tau}_{\mathbf{x}_i} \right] \right\} = 0, \quad (8)$$

$$E[\mathbf{1}(y_i \leq \mathbf{d}_i' \beta(\tau)) - \tau] = 0. \quad (9)$$

Proof. (Equation (9) only):

$$\begin{aligned} E[\mathbf{1}(y_i \leq \mathbf{d}_i' \beta(\tau))] &= E[\mathbf{1}(\mathbf{d}_i' \beta(u_i^*) \leq \mathbf{d}_i' \beta(\tau))] && \text{by **A1**} \\ &= P[u_i^* \leq \tau] && \text{by **A1**} \\ &= \tau && \text{by **A1**} \end{aligned}$$

□

The proof of equation (8) is similar and included in the Appendix. The first condition uses conditional quantile restrictions. The second condition uses only the unconditional restriction. Estimation is discussed extensively below, but the corresponding sample moments are presented here. Let $\hat{\tau}_{\mathbf{x}_i}(b) \equiv \hat{P}(y_i \leq \mathbf{d}_i' b | \mathbf{x}_i)$.

Sample Moment 1

$$g_i(b) = \mathbf{d}_i \left[\mathbf{1}(y_i \leq \mathbf{d}_i' b) - \hat{\tau}_{\mathbf{x}_i}(b) \right],$$

Sample Moment 2

$$h(b) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(y_i \leq \mathbf{d}_i' b) - \tau.$$

For comparative purposes, quantile regression relies on the sample moment

$$\tilde{g}_i(b) = \mathbf{d}_i \left[\mathbf{1}(y_i \leq \mathbf{d}_i' b) - \tau \right].$$

The GQR estimator replaces τ with a function of $\mathbf{x}$, allowing the covariates to inform the distribution of the disturbance. Precise estimation of $\hat{\tau}_{\mathbf{x}_i}$ is advantageous, but estimation error is not necessarily problematic. In practice, even an imprecisely estimated $\hat{\tau}_{\mathbf{x}_i}$ relaxes the QR assumption $u_i^* | \mathbf{d}_i \sim U(0, 1,)$.

The relationship between GQR and QR should be clear. If all variables are treatment variables (i.e., there are no control variables), then $\hat{\tau}_{\mathbf{x}_i} = \tau_{\mathbf{x}_i} = \tau$. Sample Moment 1 (and, similarly, equation (8)) reduces to the QR moment condition. Thus, QR is a special case of GQR and still available in this framework. The benefit of the GQR estimator is that it allows for the use of covariates for the purposes of identification.

3.3 Estimation

In this section, I discuss implementation of the estimator. Later, I will introduce an IV version of the proposed estimator where it is possible that the number of moment conditions exceeds the number of parameters. For this reason, I discuss estimation using Generalized Method of Moments (GMM). Sample moments are defined by

$$\hat{g}(b) = \frac{1}{N} \sum_{i=1}^N g_i(b). \quad (10)$$

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

$$\mathcal{B} \equiv \left\{ b \mid \frac{1}{N} \sum_{i=1}^N \mathbf{1}(y_i \leq \mathbf{d}_i' b) = \tau \right\}. \quad (11)$$

There is a straightforward way to confine b to the set $\mathcal{B}$ by separating out the constant. Define $\mathbf{d}_i = (1, \tilde{\mathbf{d}}_i)$. Let $\gamma(\tau, \tilde{b})$ be the τ^{th} quantile of the distribution of $y_i - \tilde{\mathbf{d}}_i' \tilde{b}$:

$$\hat{\gamma}(\tau, \tilde{b}) \text{ solves } \frac{1}{N} \sum_i^N \mathbf{1}(y_i - \tilde{\mathbf{d}}_i' \tilde{b} \leq \hat{\gamma}(\tau, \tilde{b})) = \tau. \quad (12)$$

For any $\tilde{b}$, there is a corresponding estimate of the constant which confines b to $\mathcal{B}$. Thus, $\tilde{b}$ is the only free parameter. Alternatively, one could simultaneously minimize the two moment conditions. The advantage of confining guesses to $\mathcal{B}$ is that it is a straightforward means of reducing the number of parameters that need to be estimated. Though not necessary, this simplifies estimation.

The next step is to estimate $\hat{\tau}_{\mathbf{x}_i}(b)$ using

$$\hat{\tau}_{\mathbf{x}_i}(b) = \hat{P}\left(y_i \leq \hat{\gamma}(\tau, \tilde{b}) + \tilde{\mathbf{d}}'_i \tilde{b} | \mathbf{x}_i\right). \quad (13)$$

The above equation states that the probability that y_i is less than or equal to $\hat{\gamma}(\tau, \tilde{b}) + \tilde{\mathbf{d}}'_i \tilde{b}$ is dependent on $\mathbf{x}_i$. Certain variables predict that an observation is above or below the given quantile function. This probability must be estimated. There is little guidance for the correct specification of this relationship. A probit or logit regression is easy to implement and it does not seem restrictive to assume that misspecification at this step is orthogonal to the policy variables. Less parametric techniques may also be used. Simulations in section 6 suggest that probit or logit estimation at this step works extremely well. In the simulations and applications, I will use $\hat{P}\left(y_i \leq \hat{\gamma}(\tau, \tilde{b}) + \tilde{\mathbf{d}}'_i \tilde{b} | \mathbf{x}_i\right) \equiv \Phi(\mathbf{x}'_i \hat{\delta})$ or $\Lambda(\mathbf{x}'_i \hat{\delta})$.

The sample moment condition is

$$g_i(b) = \mathbf{d}_i \left[\mathbf{1}\left(y_i \leq \hat{\gamma}(\tau, \tilde{b}) + \tilde{\mathbf{d}}'_i \tilde{b}\right) - \hat{\tau}_{\mathbf{x}_i}(b) \right],$$

$$\hat{g}(b) = \frac{1}{N} \sum_{i=1}^N g_i(b). \quad (14)$$

Finally,

$$\widehat{\beta(\tau)} = \arg \min_{b \in \mathcal{B}} \hat{g}(b)' \hat{A} \hat{g}(b) \quad (15)$$

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

The estimation steps are as follows. For each possible $\tilde{b}$:

1. Calculate $\hat{\gamma}(\tau, \tilde{b})$ using equation (12).
2. Estimate $\hat{\tau}_{\mathbf{x}_i}(b)$ using equation (13) for all $\mathbf{x}_i$.
3. Calculate $\hat{g}(b)' \hat{A} \hat{g}(b)$.

The b that minimizes $\hat{g}(b)' \hat{A} \hat{g}(b)$ is $\widehat{\beta(\tau)}$. In many contexts, it is standard to only

have one or two treatment variables. Grid-searching is practical in these circumstances. A similar argument is made in Chernozhukov and Hansen [2008]. Standard statistical programs are capable of dealing with numerous control variables using the proposed estimation procedure. Note again the value in directly calculating the constant. By reducing the number of parameters that need to be estimated, even by one, grid-searching becomes much more practical. The suggested inference approach accounts for the relationship between the estimated parameters and the constant.

In this set-up, I have suggested using a probit regression to estimate $\hat{\tau}_{\mathbf{x}_i}$. For a given b , $\tau\%$ of the observations are “below” $\mathbf{d}'_i b$. With a probit regression, regress the dummy variable for “below” on the covariates. No distributional assumptions have ever been made and it is not clear that a probit regression should provide consistent estimates of $\tau_{\mathbf{x}_i}$. However, consistent estimation is not necessary and probit (or logit) regressions are recommended simply for computational convenience. These techniques may not provide consistent estimates for $\tau_{\mathbf{x}_i}$, but the estimator only requires that estimation error is orthogonal to the policy variables. Alternatively, there is nothing in the framework or above estimation procedure that precludes use of less parametric methods.

Note also the similarity of step 2 in this estimation procedure with the estimator proposed by Firpo et al. [2009]. As discussed in section 2.6, Firpo et al. [2009] is similar to the GQR estimator when all variables are considered control variables.

This estimation procedure is very straightforward to implement using standard statistical software and arguably easier than conditional quantile estimation techniques which require special optimization procedures. By focusing on the distributional impacts of the treatment variables - which are usually limited to one or two variables - simple grid-searching coupled with procedures already available in standard statistical programs (relying on optimization methods simpler than those required for QR) are enough to implement the above estimation technique. If the SQF includes more than two treatment variables, then it is likely that more advanced optimization techniques such as MCMC (see Chernozhukov and Hong [2003]) are necessary. While not presented here, MCMC optimization worked well in simulations with several treatment variables.

The motivation of the proposed estimators is not to compare its relative computational speed to other quantile methods, but it is useful to discuss the practicality of its implementation in relation to existing conditional quantile methods. The similarities with the optimization method suggested in Chernozhukov and Hansen [2008] should be clear since

they both recommend grid-searching over a small set of treatment variables and using techniques that are standard in most statistical software to account for other variables. Note that the Chernozhukov and Hansen [2008] requires use of quantile regression optimization for each guess in the grid-search while the proposed estimators require use of a probit regression technique. Given the relative speed of the latter, the proposed estimators are an order of magnitude faster than the Chernozhukov and Hansen [2008] estimator.

3.4 Identification

This section briefly considers the uniqueness of $\beta(\tau)$. Identification of unconditional QTEs is discussed extensively in Powell [2010]. Define $\hat{\tau}_{\mathbf{x}_i}(\check{\beta}) \equiv \hat{P}(y_i \leq \mathbf{d}'_i \check{\beta} | \mathbf{x}_i)$.

Theorem 3.2 (Identification). *If (i) **A1-A4** hold; (ii) $E\left\{\mathbf{d}_i \left[\mathbf{1}(y_i \leq \mathbf{d}'_i \check{\beta}) - \hat{\tau}_{\mathbf{x}_i}(\check{\beta})\right]\right\} = 0$; (iii) $E\left[\mathbf{1}(y_i \leq \mathbf{d}'_i \check{\beta})\right] = \tau$, then $\check{\beta} = \beta(\tau)$.*

Identification stems from assumption (ii) and the full rank assumption which together imply $E\left[\mathbf{1}(y_i \leq \mathbf{d}'_i \check{\beta}) | \mathbf{d}_i, \mathbf{x}_i\right]$ is constant for all possible values of $\mathbf{d}$ for a given $\mathbf{x}$. Due to the conditional continuity of y , we know that there is only one set of parameters for each τ which meet this condition.

Proof. From assumption (ii), we know that $E\left\{\mathbf{d}_i E\left[\mathbf{1}(y_i \leq \mathbf{d}'_i \check{\beta}) - \hat{\tau}_{\mathbf{x}_i}(\check{\beta}) | \mathbf{d}_i, \mathbf{x}_i\right]\right\} = 0$.

This condition combined with **A3** implies that $E\left[\mathbf{1}(y_i \leq \mathbf{d}'_i \check{\beta}) | \mathbf{d}_i, \mathbf{x}_i\right]$ is constant for all possible values of $\mathbf{d}$ for a given $\mathbf{x}$:

$$P(y_i \leq \mathbf{d}'_i \check{\beta} | \mathbf{d}_i, \mathbf{x}_i) = P(y \leq \mathbf{d}'_i \check{\beta} | \mathbf{x}_i).$$

A4 implies $\mathbf{d}'_i \check{\beta} = \mathbf{d}'_i \beta(\tilde{\tau})$ for some $\tilde{\tau} \in (0, 1)$. **A3** implies $\check{\beta} = \beta(\tilde{\tau})$.

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

4 Model for Endogenous Treatment Variables

Even after conditioning on $\mathbf{x}$, the policy variables may be endogenous. This section introduces an IV-GQR estimator which assumes the existence of instruments correlated with the

treatment variables. This estimator relaxes the assumptions required by IV-QR to obtain unconditional QTEs. Instead of $u_i^*|\mathbf{z}_i \sim U(0, 1)$, IV-GQR only requires $u_i^*|\mathbf{z}_i, \mathbf{x}_i \sim u_i^*|\mathbf{x}_i$.

The conditions in the section closely follow those found in Powell [2010]. 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 variables will be indexed by $k \in \{1, \dots, K\}$, the instruments by $l \in \{1, \dots, L\}$ where $L \geq K$. I allow the variables in $\mathbf{x}$ to be continuous or discrete and assume the existence of multiple values of $\mathbf{x}$. Conditions for continuous policy variables and instruments will be briefly discussed below. Define D as a matrix of all possible values for $\mathbf{d}$,

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

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 $\mathbf{x}$.

$$\Pi'(\mathbf{x}^{(m)}) \equiv \begin{bmatrix} P(\mathbf{d}_i = \mathbf{d}^{(1)}|\mathbf{z}_i = \mathbf{z}^{(1)}, \mathbf{x}^{(m)}) & \dots & P(\mathbf{d}_i = \mathbf{d}^{(K)}|\mathbf{z}_i = \mathbf{z}^{(1)}, \mathbf{x}^{(m)}) \\ \vdots & \ddots & \vdots \\ P(\mathbf{d}_i = \mathbf{d}^{(1)}|\mathbf{z}_i = \mathbf{z}^{(L)}, \mathbf{x}^{(m)}) & \dots & P(\mathbf{d}_i = \mathbf{d}^{(K)}|\mathbf{z}_i = \mathbf{z}^{(L)}, \mathbf{x}^{(m)}) \end{bmatrix}$$

$$\Pi'(\mathbf{x}^{m(1)}, \dots, \mathbf{x}^{m(s)}) \equiv \begin{bmatrix} \Pi'(\mathbf{x}^{m(1)}) \\ \vdots \\ \Pi'(\mathbf{x}^{m(s)}) \end{bmatrix}$$

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

$$\overline{\Pi}'(\mathbf{x}^{(m)}) \equiv \begin{bmatrix} P(\mathbf{d} = \mathbf{d}^{(1)}|\mathbf{x}^{(m)}) & \dots & P(\mathbf{d} = \mathbf{d}^{(K)}|\mathbf{x}^{(m)}) \end{bmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{L \times 1}$$

Finally, I define the following matrix:

$$\Gamma(\mathbf{x}^{(m)}) \equiv \begin{bmatrix} P(y \leq \mathbf{d}^{(1)'}\beta(\tau)|\mathbf{z}_i, \mathbf{x}^{(m)}) \\ \vdots \\ P(y \leq \mathbf{d}^{(K)'}\beta(\tau)|\mathbf{z}_i, \mathbf{x}^{(m)}) \end{bmatrix}$$

This notation is useful since we are interested in $P[y_i \leq \mathbf{d}_i'\beta(\tau)|\mathbf{z}_i, \mathbf{x}^{(m)}]$, or $\Pi'(\mathbf{x}^{(m)})\Gamma(\mathbf{x}^{(m)})$.

4.1 Assumptions

The following conditions hold jointly with probability one:

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

IV-A2 Conditional Independence:

(a) $P(u_i^* \leq \tau|\mathbf{z}_i, \mathbf{x}_i) = P(u_i^* \leq \tau|\mathbf{x}_i)$.

(b) $E[\mathbf{z}_i\nu_i] = 0$.

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

IV-A4 Full Rank: D is full rank.

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

The assumptions are similar to those found before, except that $\mathbf{z}$ replaces $\mathbf{d}$. A first stage assumption is also needed. This assumption is more restrictive than the equivalent for mean-IV. The instruments must impact the entire distribution of the policy variables. In practice, this condition can be relaxed. Powell [2012] argues that only a subset of the possible policy values need to be impacted by the instruments, as long as this subset creates a matrix $\dot{D}$ which is full rank.

4.2 Moment Conditions

The moment conditions are similar to those as before and proofs are not included.

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

$$E \left\{ \mathbf{z}_i \left[\mathbf{1}(y_i \leq \mathbf{d}_i' \beta(\tau)) - \hat{\tau}_{\mathbf{x}_i} \right] \right\} = 0, \quad (16)$$

$$E[\mathbf{1}(y_i \leq \mathbf{d}_i' \beta(\tau)) - \tau] = 0. \quad (17)$$

The corresponding sample moments are

IV Sample Moment 1

$$g_i(b) = \mathbf{z}_i \left[\mathbf{1}(y_i \leq \mathbf{d}_i' b) - \hat{\tau}_{\mathbf{x}_i}(b) \right],$$

IV Sample Moment 2

$$h(b) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(y_i \leq \mathbf{d}_i' b) - \tau.$$

The second moment condition is identical to the GQR case. The first moment condition has replaced $\mathbf{d}_i$ with $\mathbf{z}_i$. Estimation proceeds as before using these sample moments.

Similar to the GQR case, if all variables are labeled policy variables so that there are no control variables, then IV-GQR reduces to IV-QR.

4.3 Identification

Theorem 4.2 (IV Identification). *If (i) **IV-A1** - **IV-A5** hold; (ii) $E \left\{ \mathbf{z}_i \left[\mathbf{1}(y_i \leq \mathbf{d}_i' \check{\beta}) - \hat{\tau}_{\mathbf{x}_i}(\check{\beta}) \right] \right\} = 0$; (iii) $E \left[\mathbf{1}(y_i \leq \mathbf{d}_i' \check{\beta}) \right] = \tau$, then $\check{\beta} = \beta(\tau)$.*

Proof. To simplify discussion of the proof, assume - without loss of generality - that $\Pi'(\mathbf{x}_i)$ is rank K .

$$E \left\{ \mathbf{z}_i \left[\mathbf{1}(y_i \leq \mathbf{d}_i' \check{\beta}) - \hat{\tau}_{\mathbf{x}_i}(\check{\beta}) \right] \right\} = 0 \text{ implies}$$

$$E \left\{ \mathbf{z}_i E \left[\mathbf{1}(y_i \leq \mathbf{d}_i' \check{\beta}) - \hat{\tau}_{\mathbf{x}_i}(\check{\beta}) \mid \mathbf{z}_i, \mathbf{x}_i \right] \right\} = 0.$$

This equation gives us $[\Pi'(\mathbf{x}_i) - \bar{\Pi}'(\mathbf{x}_i)] \check{\Gamma}(\mathbf{x}_i) = 0$.

Assumption **IV-A3** leads to $\check{\Gamma}(\mathbf{x}_i) = \tilde{\Gamma}(\mathbf{x}_i)$ for some $\tilde{\tau} \in (0, 1)$.

By **IV-A5**, we know that $D\check{\beta} = D\beta(\tilde{\tau})$. **A4** implies $\check{\beta} = \beta(\tilde{\tau})$.

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

5 Properties

This section briefly discusses consistency, asymptotic normality, and inference of the IV-GQR estimates (where it is possible that $\mathbf{z} = \mathbf{d}$). Let $\mu_i \equiv y_i - \mathbf{d}_i' \beta(\tau)$. Some additional assumptions are necessary:

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

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

IV-A8 $\nu_i(\check{\beta})$ continuous in $\check{\beta}$.

IV-A9 $\|\mathbf{z}_i\|^{2+\delta} < \infty$ for some $\delta > 0$.

IV-A10 (a) $E \left[\mathbf{z}_i \frac{\partial \nu_i}{\partial \check{\beta}} \right] = 0$; (b) $G \equiv E \left\{ \mathbf{z}_i \left(\tilde{\mathbf{d}}_i' + \frac{\partial \tilde{\gamma}}{\partial \delta'} \right) [f_\mu(0|\mathbf{x}_i, \mathbf{z}_i) - f_\mu(0|\mathbf{x}_i)] \right\}$ exists with $G'AG$ nonsingular.

where δ represents the coefficients on $\tilde{\mathbf{d}}_i$. Assumption **IV-A10(b)** accounts for the proposed estimation technique which imposes a relationship between the coefficients on the treatment variables and the constant.

The other assumptions are standard with a couple of exceptions. Assumption **IV-A8** states that the difference between $P(y_i \leq \mathbf{d}_i' \check{\beta} | \mathbf{x}_i)$ and $\hat{P}(y_i \leq \mathbf{d}_i' \check{\beta} | \mathbf{x}_i)$ must be continuous in $\check{\beta}$. This rules out large jumps in the estimation error of $\tau_{\mathbf{x}_i}$. Assumption **IV-A10(a)** states that changes in the estimation error are uncorrelated with $\mathbf{z}_i$.

Theorem 5.1 (Consistency). *If **IV-A1** - **IV-A9** 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 [1994]).

The following conditions hold:

1. Theorem 4.2 proves identification.
2. Compactness of $\mathcal{B}$ holds by assumption (**IV-A7**).

3. $g_i(\check{\beta})$ is continuous at each $\check{\beta}$ with probability one under **IV-A5** and **IV-A8**.

4. $\left\|g_i(\check{\beta})\right\| \leq \|\mathbf{z}_i\| < \infty$

Under these conditions, consistency follows immediately from Theorem 2.6 of Newey and McFadden [1994].

5.1 Asymptotic Normality

The conditions for asymptotic normality are difficult with discontinuous sample moments. 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 5.2 (Asymptotic Normality). *If **IV-A1** - **IV-A10** 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 [1994] discuss asymptotic normality results for discontinuous moment conditions. Stochastic equicontinuity is an important condition for these results and follows here from the fact that the functional class $\{\mathbf{1}(y_i \leq \mathbf{d}'_i b) - P(y_i \leq \mathbf{d}'_i b | \mathbf{x}_i), b \in \mathbb{R}^k\}$ is Donsker and the Donsker property is preserved when the class is multiplied by a bounded random variable. Thus,

$$\{\mathbf{z}_i [\mathbf{1}(y_i \leq \mathbf{d}'_i b) - P(y_i \leq \mathbf{d}'_i b | \mathbf{x}_i)], b \in \mathbb{R}^k\}$$

is Donsker with envelope $2 \max_i |\mathbf{z}_i|$. Stochastic equicontinuity follows from Theorem 1 in Andrews [1986]. The appendix includes further discussion.

The advantage of conditioning on control variables for variance reduction should be clear from G .

5.2 Inference

With QR, is it common to make the assumption $f_{\mu_i}(0 | \mathbf{d}_i) = f_{\mu_i}(0)$ for variance estimation. The equivalent assumption here ($f_{\mu}(0 | \mathbf{x}_i, \mathbf{z}_i) = f_{\mu}(0)$) is difficult since a main motivation of this paper is that $\mathbf{x}_i$ and $\mathbf{z}_i$ provide information about the value of μ_i .

It is possible to use a histogram estimation technique in the spirit of the one suggested in Powell [1986] to estimate G . I suggested a slight variation of this method because of concerns with estimating the $f_\mu(0|\mathbf{x}_i)$ term.

$$\hat{G} = \frac{1}{2Nh} \sum_{i=1}^N \mathbf{z}_i \left(\tilde{\mathbf{d}}'_i + \frac{\partial \hat{\gamma}}{\partial \delta'} \right) \left\{ \left[\mathbf{1} \left(y_i - \mathbf{d}'_i \widehat{\beta(\tau)} \leq h \right) - \mathbf{1} \left(y_i - \mathbf{d}'_i \widehat{\beta(\tau)} \leq -h \right) \right] - \left[\hat{P} \left(y_i - \mathbf{d}'_i \widehat{\beta(\tau)} \leq h \mid \mathbf{x}_i \right) - \hat{P} \left(y_i - \mathbf{d}'_i \widehat{\beta(\tau)} \leq -h \mid \mathbf{x}_i \right) \right] \right\} \quad \text{for small } h.$$

The $\left[\mathbf{1} \left(y_i - \mathbf{d}'_i \widehat{\beta(\tau)} \leq h \right) - \mathbf{1} \left(y_i - \mathbf{d}'_i \widehat{\beta(\tau)} \leq -h \right) \right]$ term is equivalent to $\mathbf{1} \left| y_i - \mathbf{d}'_i \widehat{\beta(\tau)} \right| \leq h$, found in Powell [1986]. The second part of $\hat{G}$ is the estimated probability that $y_i - \mathbf{d}'_i \widehat{\beta(\tau)}$ is close to 0 given only $\mathbf{x}_i$. I recommend separately estimating (a) $\hat{P} \left(y_i - \mathbf{d}'_i \widehat{\beta(\tau)} \leq h \mid \mathbf{x}_i \right)$ and (b) $\hat{P} \left(y_i - \mathbf{d}'_i \widehat{\beta(\tau)} \leq -h \mid \mathbf{x}_i \right)$ using the same procedure used in the Step 2 of the estimation procedure found in section 3.3.

Finally, given the proposed estimation procedure, it is necessary to estimate $\frac{\partial \hat{\gamma}}{\partial \delta'}$.

Note that a standard bootstrap with replacement is typically thought to have many advantages for quantile estimators (Hahn [1995]) and justified in this context (see Hahn [1996]).

6 Empirical Applications

6.1 Monte Carlo Simulations

This section tests the GQR estimation technique. I perform two separate tests. First, I generate data where the policy variable is exogenous, unconditional on covariates. The purpose of this test is to compare the precision of the GQR and QR estimates. Second, I generate data where conditioning on covariates is necessary for identification.

In the first case, d is randomly-assigned. The impact of d on y varies by observation. The impact is a function of observed proneness, x_i , and unobserved proneness, u_i . I let the observed proneness variable have a larger impact on rank to illustrate the importance of unconditional QTEs. I look at several quantiles. The coefficient of interest is $\beta(\tau) = \tau$. I

generate the following data for N=500:

$$\begin{aligned}
\text{Observed Skill: } & x_i \sim U(0, 1), \\
\text{Unobserved Skill: } & \epsilon_i \sim U(0, 0.1), \\
\text{Total Disturbance: } & u_i \equiv F_{x_i+\epsilon_i}(x_i + \epsilon_i) \Rightarrow u_i \sim U(0, 1), \\
\text{Policy Variable: } & d_i \sim U(0, 1), \\
\text{Outcome: } & y_i = u_i(1 + d_i).
\end{aligned}$$

I report four sets of results in Table 1. First, I perform quantile regressions of y on d and x to obtain conditional QTEs. Second, I perform quantile regressions of y on d to obtain unconditional QTEs. These results are shown in the top half of Table 1. The bottom half presents results from the GQR estimator. The third set of results uses logit regression to estimate $\tau_{\mathbf{x}_i}$, while the fourth set uses probit regression.

The conditional QRs, as expected, produced biased results (with the exception of the median, a product of the data-generating process). This is an especially important result given that d is randomly-assigned. As discussed previously, even in this situation, controlling for additional covariates biases the results. QR (without covariates) produces consistent estimates in this case. The GQR estimators perform very well. Note that GQR performs better than QR (without covariates) since it is using additional information. This is a major benefit of the GQR estimator.

Next, I generate data where conditioning on x is necessary, such that QR does not provide consistent estimates with or without covariates. With these data, d is exogenous, conditional on x .

$$\begin{aligned}
\text{Observed Skill: } & x_i \sim U(0, 1), \\
\text{Unobserved Skill: } & \epsilon_i \sim U(0, 0.1), \\
\text{Total Disturbance: } & u_i \equiv F_{x_i+\epsilon_i}(x_i + \epsilon_i) \Rightarrow u_i \sim U(0, 1), \\
& \psi_i \sim U(0, 1), \\
\text{Policy Variable: } & d_i = x_i + \psi_i, \\
\text{Outcome: } & y_i = u_i(1 + d_i).
\end{aligned}$$

Table 2 presents the same statistics as before. Both sets of QR results are biased.

Table 1: Simulation Results: Policy Variable Randomly-Assigned

QR (conditional)				QR (unconditional)		
Quantile	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE
5	0.40555	0.40555	0.41231	0.00120	0.04159	0.05007
10	0.37051	0.37051	0.37440	0.00166	0.05665	0.06928
15	0.32667	0.32667	0.32933	0.00436	0.06725	0.08252
20	0.27998	0.27998	0.28215	0.00305	0.07552	0.09295
25	0.23430	0.23430	0.23605	0.00272	0.08028	0.09881
30	0.18773	0.18773	0.18938	0.00303	0.08514	0.10510
35	0.14101	0.14101	0.14283	0.00406	0.08675	0.10875
40	0.09373	0.09373	0.09606	0.00408	0.09021	0.11247
45	0.04702	0.04722	0.05121	0.00281	0.09162	0.11369
50	0.00005	0.01655	0.02059	0.00446	0.09243	0.11460
55	-0.04703	0.04722	0.05127	0.00364	0.09156	0.11297
60	-0.09393	0.09393	0.09614	0.00374	0.09027	0.11148
65	-0.14057	0.14057	0.14241	0.00400	0.08791	0.10953
70	-0.18723	0.18723	0.18893	0.00371	0.08493	0.10515
75	-0.23423	0.23423	0.23604	0.00035	0.07901	0.09851
80	-0.28087	0.28087	0.28305	-0.00055	0.07203	0.08940
85	-0.32529	0.32529	0.32809	-0.00029	0.06454	0.07939
90	-0.36743	0.36743	0.37142	0.00040	0.05400	0.06637
95	-0.40129	0.40129	0.40843	0.00044	0.04085	0.04906

GQR (logit)				GQR (probit)		
Quantile	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE
5	-0.00121	0.02397	0.02954	-0.00330	0.02396	0.02974
10	-0.00025	0.02491	0.03037	-0.00006	0.02528	0.03110
15	0.00056	0.02582	0.03144	0.00038	0.02556	0.03094
20	0.00083	0.02599	0.03175	0.00088	0.02602	0.03157
25	-0.00015	0.02517	0.03057	0.00053	0.02523	0.03068
30	0.00061	0.02455	0.03013	0.00006	0.02540	0.03125
35	0.00098	0.02540	0.03129	0.00003	0.02609	0.03199
40	0.00062	0.02560	0.03151	0.00013	0.02547	0.03135
45	-0.00016	0.02508	0.03052	-0.00104	0.02512	0.03086
50	0.00103	0.02437	0.03006	0.00073	0.02539	0.03135
55	0.00033	0.02561	0.03077	0.00030	0.02542	0.03068
60	-0.00010	0.02588	0.03144	-0.00067	0.02581	0.03133
65	-0.00033	0.02515	0.03054	-0.00022	0.02502	0.03012
70	0.00117	0.02521	0.03125	0.00083	0.02509	0.03126
75	-0.00004	0.02374	0.02941	-0.00011	0.02435	0.02999
80	-0.00037	0.02469	0.03000	0.00039	0.02515	0.03080
85	0.00066	0.02580	0.03136	0.00042	0.02564	0.03103
90	-0.00015	0.02475	0.03081	0.00050	0.02454	0.03056
95	-0.00304	0.02520	0.03081	-0.00128	0.02460	0.03012

Results based on 1000 replications, N=500. MAD is Mean Absolute Deviation, RMSE is Root Mean Squared Error.

Table 2: Simulation Results: Policy Variable Randomly-Assigned, Conditional on Covariate

QR (conditional)				QR (unconditional)		
Quantile	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE
5	0.40938	0.40938	0.41082	1.09017	1.09017	1.10298
10	0.36755	0.36755	0.36862	1.10850	1.10850	1.11502
15	0.32340	0.32340	0.32447	1.10497	1.10497	1.10947
20	0.27840	0.27840	0.27968	1.09717	1.09717	1.10016
25	0.23338	0.23338	0.23488	1.08256	1.08256	1.08475
30	0.18708	0.18708	0.18894	1.06411	1.06411	1.06602
35	0.14144	0.14144	0.14384	1.04266	1.04266	1.04421
40	0.09507	0.09509	0.09859	1.02061	1.02061	1.02188
45	0.04898	0.04976	0.05602	0.99611	0.99611	0.99717
50	0.00160	0.02305	0.02887	0.96943	0.96943	0.97033
55	-0.04627	0.04854	0.05625	0.94097	0.94097	0.94172
60	-0.09483	0.09488	0.10114	0.91168	0.91168	0.91231
65	-0.14223	0.14223	0.14768	0.88244	0.88244	0.88296
70	-0.19153	0.19153	0.19658	0.85331	0.85331	0.85374
75	-0.24021	0.24021	0.24560	0.82262	0.82262	0.82302
80	-0.28817	0.28817	0.29378	0.79468	0.79468	0.79509
85	-0.33448	0.33448	0.34085	0.77663	0.77663	0.77720
90	-0.38074	0.38074	0.38907	0.77256	0.77256	0.77332
95	-0.41806	0.41806	0.43170	0.78999	0.78999	0.79197

GQR (logit)				GQR (probit)		
Quantile	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE
5	-0.00252	0.02536	0.03149	-0.00497	0.02515	0.03100
10	-0.00076	0.02654	0.03254	-0.00146	0.02692	0.03319
15	-0.00004	0.02786	0.03381	-0.00024	0.02804	0.03394
20	-0.00024	0.02952	0.03567	-0.00120	0.03078	0.03743
25	-0.00176	0.02968	0.03593	-0.00176	0.03052	0.03673
30	-0.00133	0.03011	0.03704	-0.00213	0.03153	0.03850
35	-0.00066	0.03214	0.03936	-0.00249	0.03353	0.04061
40	-0.00121	0.03373	0.04080	-0.00191	0.03441	0.04192
45	-0.00165	0.03315	0.03993	-0.00328	0.03530	0.04291
50	-0.00106	0.03364	0.04128	-0.00173	0.03491	0.04311
55	-0.00187	0.03605	0.04326	-0.00334	0.03758	0.04585
60	-0.00172	0.03692	0.04474	-0.00385	0.04015	0.04838
65	-0.00222	0.03786	0.04525	-0.00405	0.03909	0.04759
70	-0.00106	0.03842	0.04694	-0.00158	0.04120	0.05038
75	-0.00323	0.03711	0.04487	-0.00415	0.04105	0.05008
80	-0.00243	0.03957	0.04822	-0.00267	0.04241	0.05250
85	-0.00054	0.04130	0.04976	-0.00294	0.04458	0.05452
90	-0.00483	0.04063	0.05009	-0.00290	0.04482	0.05560
95	-0.01141	0.04293	0.05195	-0.00331	0.04461	0.05429

Results based on 1000 replications, N=500. MAD is Mean Absolute Deviation, RMSE is Root Mean Squared Error.

The GQR estimator performs very well. Note that these simulations were a difficult test given that the impact of d was not static, but constantly changing throughout the outcome distribution.

6.2 Empirical Applications

6.2.1 Job Training

I use data analyzed in Abadie et al. [2002] to estimate the effect of job training on the distribution of earnings. Access to job training is randomly-assigned so an IV strategy is used. The dependent variable is 30-month earnings. Abadie et al. [2002] find significant heterogeneity in the impact of training throughout the conditional distribution of earnings. Similar results are found in Chernozhukov and Hansen [2008]. I report conditional QTEs using IV-QR from Chernozhukov and Hansen [2008]. I report unconditional QTEs as well using IV-GQR - one column reports results where a probit regression was used for the estimation of $\tau_{\mathbf{x}}$ while another reports results where a logit regression was used. For the IV-QR and IV-GQR estimations, I grid-search at an interval of \$25.¹

Table 3 reports the results for men. The conditional and unconditional QTEs are correlated, but the conditional QTEs do not find a large effect of training at the top of the distribution for me. The IV-GQR estimates suggest that men with high earnings potential are most helped by job training. Table 4 reports the corresponding results for women. Here, the unconditional QTEs suggest much larger effects at the bottom of the distribution for women than the conditional QTEs.

6.2.2 Teacher Attendance

Duflo et al. [2012] study the effect of teacher incentives on teacher attendance and student achievement. They use randomized incentives to instrument the number of days that a school was open. They estimate the impact of additional school days on test scores, conditioning on three different pre-intervention test score results. I use the IV-GQR, conditioning on these test score measures, to study the differential impact the the number of “open school” days can have on the distribution of achievement. Duflo et al. [2012] find a mean effect of 0.97 standard deviations, implying that a school that was open 100% of the time would increase

¹The IV-QR results reported are similar to those found in Chernozhukov and Hansen [2008]. The major difference in estimates is due to differences in the interval of the grid-search - I use \$25 while Chernozhukov and Hansen [2008] use \$100.

test scores, on average, by 0.485 standard deviations more than school that was open only 50% of the time.

I grid-search for both the IV-QR and IV-GQR estimators at an interval of 0.1. Table 5 presents the results. Though the IV-GQR estimates cannot reject the IV-QR estimates, the IV-GQR results do suggest much large effects above the median.

7 Conclusion

This paper introduces a generalized quantile regression estimator. The GQR and IV-GQR estimators allow one to condition on a separate set of covariates without altering the interpretation of the estimates. QR and IV-QR may produce biased estimates in these cases. The benefit of the GQR estimator is that the interpretation of the estimates does not change as one adds control variables, paralleling the mean regression case. Typically, researchers include control variables for the purposes of identification (or variance reduction). They do not necessarily want the interpretation of the estimates change. In fact, much empirical work interprets conditional QTEs as unconditional QTEs. The GQR and IV-GQR estimators provide a simple way to estimate unconditional QTEs in the presence of covariates. QR and IV-QR are special cases of the estimator introduced in this paper. The GQR and IV-GQR estimators are easy to implement with standard statistical software. Simulation results illustrate the usefulness of these estimators.

Table 3: Impact of Job Training on Earnings – Men

Quantile	IVQR	GQR (probit)	GQR (logit)
15	-150 (321)	-400 (423)	-400 (426)
20	-50 (385)	175 (458)	150 (454)
25	525 (476)	100 (542)	100 (544)
30	675 (676)	325 (325)	325 (325)
35	700 (838)	325 (891)	300 (882)
40	975 (807)	550 (918)	525 (913)
45	1075 (884)	575 (1196)	575 (1200)
50	325 (985)	800 (1076)	800 (1068)
55	1400 (1124)	1250 (1053)	1250 (1028)
60	1625 (1234)	2350* (1218)	2350* (1224)
65	1975* (1158)	1925 (1181)	1925 (1175)
70	2100* (1253)	2075 (1499)	1825 (1492)
75	2650* (1492)	1500 (1435)	1500 (1423)
80	2300 (1629)	1950 (1592)	1950 (1570)
85	3200*** (1174)	3425*** (1475)	3425*** (1498)
90	1875 (1878)	4375*** (1584)	4375*** (1583)
95	2400 (2354)	4350*** (1574)	4350*** (1549)

IVQR refers to Chernozhukov and Hansen [2008]. Controls include dummy variables for age groups, high school or GED, married, black, Hispanic, and worked less than 13 weeks in past year.

Table 4: Impact of Job Training on Earnings – Women

Quantile	IVQR	GQR (probit)	GQR (logit)
15	25 (77)	200 (158)	200 (156)
20	250 (157)	450* (233)	450* (231)
25	400 (288)	900* (529)	900* (527)
30	675** (311)	1275** (528)	1275** (536)
35	900*** (337)	1575*** (442)	1575*** (443)
40	1100** (506)	1700*** (594)	1700*** (596)
45	1825*** (473)	2100*** (638)	2100*** (657)
50	1575** (707)	2325*** (823)	2325*** (813)
55	1825** (745)	2175** (885)	2175** (881)
60	2025** (897)	2475*** (951)	2475*** (944)
65	2575*** (856)	2725*** (824)	2725*** (815)
70	2725*** (802)	2625*** (992)	2625*** (989)
75	2500*** (784)	1975*** (990)	1975*** (975)
80	2575*** (938)	2250*** (934)	2250*** (941)
85	1850 (1193)	1800** (889)	1800* (896)
90	1350 (1156)	2125 (1357)	2125 (1358)
95	2925* (1637)	3250** (1472)	3250** (1474)

IVQR refers to Chernozhukov and Hansen [2008]. Controls include dummy variables for age groups, high school or GED, married, black, Hispanic, worked less than 13 weeks in past year, and a dummy variables for AFDC status.

Table 5: Impact of “Open School” Days on Student Achievement

Quantile	IV-QR	GQR	Quantile	IV-QR	GQR
5	0.1 (0.6)	0.6 (0.6)	55	0.8* (0.4)	1.6*** (0.5)
10	0.3 (0.6)	0.3 (0.7)	60	0.9** (0.5)	1.4*** (0.5)
15	0.7 (0.6)	0.1 (0.5)	65	0.9** (0.5)	1.4** (0.7)
20	0.8 (0.5)	0.5 (0.3)	70	0.9** (0.4)	1.8*** (0.6)
25	0.6 (0.5)	0.7** (0.4)	75	1.0** (0.4)	1.8*** (0.7)
30	0.7* (0.4)	0.5 (0.5)	80	1.0** (0.5)	1.5*** (0.5)
35	0.9** (0.4)	1.0 (0.6)	85	1.1** (0.5)	1.5*** (0.6)
40	0.8* (0.5)	1.0 (0.7)	90	0.9 (0.6)	1.0** (0.5)
45	0.8* (0.4)	1.2 (0.7)	95	0.5 (0.8)	0.3 (0.5)
50	0.8* (0.5)	1.4** (0.6)			

IV-QR refers to Chernozhukov and Hansen [2008]. GQR uses a probit regression. Controls include three different pre-intervention test score measures. All standard errors are adjusted for clustering at the school level.

Appendix

Proof of equation (8) in **Theorem 4.1**:

$$\begin{aligned}
& E \left\{ \mathbf{d}_i \left[\mathbf{1}(y_i \leq \gamma(\tau) + \mathbf{d}_i' \beta(\tau)) - \hat{\tau}_{\mathbf{x}_i} \right] \right\} = E \left\{ \mathbf{d}_i \left[\mathbf{1}(y_i \leq \gamma(\tau) + \mathbf{d}_i' \beta(\tau)) - \tau_{\mathbf{x}_i} - \nu_i \right] \right\} \\
& = E \left\{ \mathbf{d}_i \left[\mathbf{1}(y_i \leq \gamma(\tau) + \mathbf{d}_i' \beta(\tau)) - \tau_{\mathbf{x}_i} \right] \right\} \quad \text{by A2(b)} \\
& = E \left\{ E \left[\mathbf{d}_i \left[\mathbf{1}(y_i \leq \gamma(\tau) + \mathbf{d}_i' \beta(\tau)) - \tau_{\mathbf{x}_i} \right] \middle| \mathbf{d}_i, \mathbf{x}_i \right] \right\} \\
& = E \left\{ \mathbf{d}_i \left[P \left(\gamma(u_i^*) + \mathbf{d}_i' \beta(u_i^*) \leq \gamma(\tau) + \mathbf{d}_i' \beta(\tau) \middle| \mathbf{d}_i, \mathbf{x}_i \right) - P \left(\gamma(u_i^*) + \mathbf{d}_i' \beta(u_i^*) \leq \gamma(\tau) + \mathbf{d}_i' \beta(\tau) \middle| \mathbf{x}_i \right) \right] \right\} \quad \text{by A1} \\
& = E \left\{ \mathbf{d}_i \left[P(u_i^* \leq \tau \middle| \mathbf{d}_i, \mathbf{x}_i) - P(u_i^* \leq \tau \middle| \mathbf{x}_i) \right] \right\} \quad \text{by A1} \\
& = E \left\{ \mathbf{d}_i \left[P(u_i^* \leq \tau \middle| \mathbf{x}_i) - P(u_i^* \leq \tau \middle| \mathbf{x}_i) \right] \right\} \quad \text{by A2(a)} \\
& = 0
\end{aligned}$$

Theorem 5.2 [Asymptotic Normality] If **IV-A1** - **IV-A10** hold and $\hat{A} \xrightarrow{p} A$ positive definite, then $\sqrt{N}(\widehat{\beta(\tau)} - \beta(\tau)) \xrightarrow{d} N[0, (G'AG)^{-1}G'ASAG(G'AG)^{-1}]$.

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

Also, $\overline{G}(\beta) \equiv E \left\{ \mathbf{z}_i \left(\tilde{\mathbf{d}}_i' + \frac{\partial \hat{\gamma}}{\partial \beta'} \right) [f_y(\mathbf{d}_i' \beta | \mathbf{x}_i, \mathbf{z}_i) - f_y(\mathbf{d}_i' \beta | \mathbf{x}_i)] \right\}$.

$\overline{g}(\beta) \equiv E \left[\mathbf{z}_i \left(\mathbf{1}(y_i \leq \mathbf{d}_i' \beta) - P(y_i \leq \mathbf{d}_i' \beta | \mathbf{x}_i) \right) \right]$.

Proof:

$\overline{g}(\beta)'A\overline{g}(\beta)$ is minimized at β_0 implying that

$$\overline{G}(\beta_0)'A\overline{g}(\beta_0) = 0.$$

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

$$\sqrt{N}\overline{g}(\beta_0) = \sqrt{N}\overline{g}(\hat{\beta}) - \overline{G}(\hat{\beta})\sqrt{N}(\hat{\beta} - \beta_0) \quad (18)$$

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

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

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

$$\begin{aligned} -\sqrt{N}\bar{g}(\hat{\beta}) &= \left[\sqrt{N}g_N(\hat{\beta}) - \sqrt{N}\bar{g}(\hat{\beta}) \right] - \sqrt{N}g_N(\hat{\beta}) \\ &= \underbrace{\sqrt{N} \left[g_N(\hat{\beta}) - \bar{g}(\hat{\beta}) - g_N(\beta_0) \right]}_{(1)} + \underbrace{\sqrt{N}g_N(\beta_0)}_{(2)} - \underbrace{\sqrt{N}g_N(\hat{\beta})}_{(3)}. \end{aligned}$$

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

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

$$\{\mathbf{z}_i [\mathbf{1}(y_i \leq \mathbf{d}_i' b) - P(y_i \leq \mathbf{d}_i' b | \mathbf{x}_i)], b \in \mathbb{R}^k\}$$

is Donsker with envelope $2 \max_i |\mathbf{z}_i|$. Stochastic equicontinuity of $v_N(\cdot)$ follows from **IV-A9** and Theorem 1 in Andrews [1986]. Stochastic equicontinuity and consistency of $\hat{\beta}$ implies that part (1) is $o_p(1)$.

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

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

Plugging into (18) and using the assumption that $G'AG$ nonsingular

$$\sqrt{N}(\hat{\beta} - \beta(\tau)) \xrightarrow{d} N[0, (G'AG)^{-1}G'ASAG(G'AG)^{-1}]$$

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.
- 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 and Han Hong. An mcmc approach to classical estimation. *Journal of Econometrics*, 115(2):293–346, August 2003.
- Kjell Doksum. Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.*, 2(2):267–277, 1974.
- E. Duflo, R. Hanna, and S.P. Ryan. Incentives work: Getting teachers to come to school. *The American Economic Review*, 102(4):1241–1278, 2012.
- 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 Jr. Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics*, 164(1):142–157, September 2011.
- Jinyong Hahn. Bootstrapping quantile regression estimators. *Econometric Theory*, 11(1), 1995.

- Jinyong Hahn. A note on bootstrapping generalized method of moments estimators. *Econometric Theory*, 12(01):187–197, March 1996.
- 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.
- José Mata and José A. F. Machado. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20(4):445–465, 2005.
- Blaise Melly. Decomposition of differences in distribution using quantile regression. *Labour Economics*, 12(4):577–590, August 2005.
- Whitney K. Newey and Daniel McFadden. Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245, 1994.
- 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.
- David Powell. Unconditional quantile regression for panel data with exogenous or endogenous regressors. Working papers, RAND Corporation Publications Department, 2012.
- James L. Powell. Censored regression quantiles. *Journal of Econometrics*, 32(1):143–155, June 1986.
- A.W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996.