QUANTILE TREATMENT EFFECTS IN THE PRESENCE OF COVARIATES

David Powell*

Abstract—This paper proposes a method to estimate unconditional quantile treatment effects (QTEs) given one or more treatment variables, which may be discrete or continuous, even when it is necessary to condition on covariates. The estimator, generalized quantile regression (GQR), is developed in an instrumental variable framework for generality to permit estimation of unconditional QTEs for endogenous policy variables, but it is also applicable in the conditionally exogenous case. The framework includes simultaneous equations models with nonadditive disturbances, which are functions of both unobserved and observed factors. Quantile regression and instrumental variable quantile regression are special cases of GQR and available in this framework.

I. Introduction

It is often important to understand the distributional impacts of policies. Mean estimates can mask critical heterogeneity, but quantile treatment effects (QTEs) characterize the effects of policy variables throughout the outcome distribution. Quantile estimators, such as the quantile regression (QR; Koenker & Bassett, 1978) and instrumental variable quantile regression (IVQR; Chernozhukov & Hansen, 2006) estimators, are useful for the estimation of conditional quantile treatment effects. However, researchers are often interested in the relationship between the treatment variables and the outcome distribution, unconditional on additional covariates. This paper introduces a framework and method to estimate unconditional quantile treatment effects even when it is necessary, or simply desirable, to condition on other control variables. The estimator permits joint estimation of QTEs for multiple treatment variables, which can be discrete or continuous. The estimator is developed in an instrumental variable framework for generality and allows for estimation of unconditional QTEs for endogenous or exogenous policy variables.

Due to the linearity of the expected value operator, unconditional and conditional average treatment effects have similar interpretations. However, this feature does not extend to quantile models since the mean of conditional quantile models fails to provide information about the unconditional quantile function. For example, we are likely interested in how job placement affects the lower part of the earnings distribution. Conditioning on education should be useful for identification and estimation but poses difficulties in quantile models. The 10th percentile of the distribution conditional on college education may be relatively high in the unconditional earnings distribution given that college education predicts higher earnings. The conditional and unconditional models have different interpretations. The estimator introduced in this paper provides unconditional QTEs. Conditioning on additional covariates using this approach will not affect the interpretation of the estimates beyond their effects on the plausibility of the identification assumptions, similar to the gains in controlling for covariates in mean regression.

Consider a latent (potential) outcome framework, where Y_d represents a continuous outcome given treatment variables, $D = d$.[1] The observed outcome is $Y \equiv Y_D$. We are interested in the τth quantile of Y_d, represented by $q(d, \tau)$. The QTEs are defined as the changes in the τth quantile of the outcome distribution given a shift in the policy variables from d_0 to d_1: $q(d_1, \tau) - q(d_0, \tau)$. For continuous policy variables, QTEs can be represented by $\frac{\partial q(d,\tau)}{\partial d}$.

In this paper's framework, additional covariates ($X = x$) are not included in $q(d, \tau)$, which distinguishes it from conditional quantile estimators. The covariates are used for identification purposes and variance reduction to control for varying propensities to have outcomes above or below the quantile function given those observable characteristics. For example, a person with a college degree is more likely to have labor earnings in the upper parts of the earnings distribution, and this conditional probability is jointly estimated.

Chernozhukov and Hansen (2013) note that the quantile index in their framework refers to the quantile of the potential outcome for fixed exogenous covariates $X = x$ and "not to the unconditional quantile of Y_d." Using similar assumptions, though, this framework can be extended to allow for more flexible estimation of QTEs. In a conditional quantile framework, all variables are considered treatment variables. The flexibility of this paper's framework is that it permits the researcher to use treatment and control variables differently. The estimator does not require including the covariates in $q(d, \tau)$ in order to condition on those covariates. When all variables are treatment variables, the framework and estimator of this paper are equivalent to conditional quantile models. In this manner, the estimator nests QR and IV-QR and, for this reason, I refer to the estimator as generalized quantile regression (GQR).

Recent work has developed techniques to estimate unconditional QTEs with similar motivations as GQR. Using a propensity score framework, Firpo (2007) introduces a

Received for publication November 13, 2017. Revision accepted for publication May 22, 2019. Editor: Bryan S. Graham.

*Powell: RAND.

I thank Whitney Newey and Jerry Hausman for their guidance when I began working on this topic. I gratefully acknowledge funding from the CDC (R01CE02999). I received helpful comments from seminar participants at the North American Summer Meeting of the Econometric Society, University of California–Irvine, the Center for Causal Inference, RAND, and the 2014 Stata Conference. I also had helpful discussions with Abby Alpert, David Autor, Matthew Baker, Marianne Bitler, Yingying Dong, Helen Hsi, Mireille Jacobson, Nicole Maestas, Erik Meijer, Kathleen Mullen, Amanda Pallais, Christopher Palmer, Michael Robbins, João M.C. Santos Silva, Hui Shan, and Travis Smith.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/rest_a_00858.

[1] In this paper, I follow the convention that capital letters denote random variables, and lowercase letters represent the potential values of those random variables.

The Review of Economics and Statistics, December 2020, 102(5): 994–1005

https://doi.org/10.1162/rest_a_00858

technique for estimation of unconditional QTEs with covariates and one exogenous binary treatment variable. Frölich and Melly (2013) extend this method to the case of one endogenous binary treatment variable. In contrast, GQR permits multiple treatment variables, which can be discrete or continuous. GQR relies on a different set of assumptions relative to these binary treatment variable estimators. These differences are evaluated below.

GQR is simple to implement with standard statistical software. I apply the estimator to study the effects of direct-hire and temporary-help job placements on the earnings distribution using data from Autor, Houseman, and Kerr (2017), which implements IVQR to estimate conditional quantile effects. They discuss the limitations of conditional quantiles in this context and the inapplicability of other unconditional quantile estimators given the inclusion of two endogenous variables in the quantile function.[2] The instruments they employ are only conditionally exogenous, so it is critical to condition on additional covariates. GQR is able to estimate unconditional QTEs for this application, while conditioning on the full set of covariates for identification purposes.

II. Model

A. Framework

This section builds on the framework developed in Chernozhukov and Hansen (CH, 2005), and I will highlight the relevant departures from their model. Each Y_d, assumed continuous and defined in section I, is a function of the policy variables, represented by d.[3] The main contribution of this paper is the introduction of a method to estimate unconditional QTEs given one or more, discrete or continuous, treatment variables. I develop the theoretical framework and estimator in an IV setting for generality, but the framework and estimator apply to the conditionally exogenous case. All conditions in this paper are assumed to hold jointly with probability 1:

- 1A Potential Outcomes: Y_d is the outcome given policy variables d; $q(d, \tau)$ represents the τth quantile of Y_d.
- 1B Conditional Independence: $Y_d|X, Z \sim Y_d|X$ for all d.
- 1C Selection: $D = \omega(Z, X, V)$ for some unknown function ω and random vector V.
- 1D Rank Similarity: $P(Y_d \leq q(d, \tau)|X, Z, V) = P(Y_{d'} \leq q(d', \tau)|X, Z, V)$ for all d, d'.
- 1E Observed random vector consists of $Y := Y_D, D, X, Z$.

These assumptions lead to the following result:

Theorem 1. *Suppose assumption 1 holds. Then for each* $\tau \in (0, 1)$,

$$P[Y \leq q(D, \tau)|X, Z] = P[Y \leq q(D, \tau)|X], \tag{1}$$

$$P[Y \leq q(D, \tau)] = \tau. \tag{2}$$

The proofs are in section A.1. Theorem 1 provides both a conditional and an unconditional quantile result. The conditional result, equation (1), states that once X is conditioned on, the instruments do not provide additional information about the probability that the outcome is less than (or equal to) the quantile function. The conditional probability varies based on the control variables. The unconditional result, equation (2), states that on average, the probability that the outcome variable is smaller than (or equal to) the quantile function is equal to τ. Alternatively, this condition can be written as $E\{P[Y \leq q(D, \tau)|X]\} = \tau$. The probability varies based on the covariates but, on average, it is equal to τ.

B. Discussion

Assumptions. Condition 1A defines the quantile function of interest as $q(d, \tau)$. Condition 1B is the primary departure from CH. Define $U_d^* \equiv F_d^{-1}(Y_d)$,[4] representing a rank variable determining placement in the outcome distribution for a given set of policy variables. Also define $U_d \equiv F_{d|X}^{-1}(Y_d|X)$, the conditional rank variable used in CH. Given this framework, it is helpful to model U_d^* as an arbitrary function of X and U_d: $U_d^* = \lambda_d(X, U_d)$.

CH assumes $U_d|Z, X \sim U(0, 1)$ for all values of Z and X. The framework of this paper uses X to provide information about the outcome distribution of Y_d, permitting different distributions of U_d^* for different values of X. Conditioning on a high education level should provide information that the conditional distribution is distributed differently than when conditioning on a low education level. In the CH framework, policy and control variables are treated in the same manner.

Assumption 1C models the function determining the treatment variables and is met trivially when $D = Z$. The flexibility of this assumption and the lack of an explicit "first-stage" specification when implementing the estimator (discussed below) distinguish this setup from alternative approaches using control functions that may require additional restrictions on the first-stage specification.

Condition 1D is a rank similarity assumption. This assumption is also different from the equivalent assumption in CH, which assumes rank similarity regarding U_d. As an example, note that when D is randomly assigned (or conditionally random) that $U_d \,|\, (X, D = d) \sim U_{d'} \,|\, \left(X, D = d'\right) \sim U(0, 1)$. However, the unconditional outcome ranks do not necessarily satisfy condition 1D in this case.

[2] See note 27 of their paper.

[3] In Chernozhukov and Hansen (2005), outcomes are a function of the endogenous variables d, exogenous variables x, and rank variable U. The notation in this paper does not distinguish between endogenous and exogenous variables, only treatment and control variables. The inconsistencies in notation across these papers make comparisons across CH and this paper slightly awkward. In the notation of this paper, CH does not have any X variables. Instead, all endogenous and exogenous variables in the CH framework are contained in D in this paper's framework.

[4] F_d is the CDF of Y_d.

In the binary treatment variable case, Abadie, Angrist, and Imbens (2002) and Frölich and Melly (2013) identify QTEs for "compliers." These models impose a local quantile treatment effect (LQTE) monotonicity condition on the effect of the instrument on the endogenous variable but relax the rank similarity assumption. While this paper creates a more general framework than CH, it does not attempt to nest the LQTE framework. Concerns about the monotonicity assumption (for local average treatment effects but applicable to LQTEs as well) are discussed in de Chaisemartin (2017). Tests of the monotonicity assumption have also been introduced (Mourifié & Wan, 2017). Tests for the rank similarity assumption have been developed in Dong and Shen (2018) and Frandsen and Lefgren (2018) and should apply here as well. Wüthrich (2020) discusses the relationship between the IVQR model and the LQTE model introduced in Abadie et al. (2002), finding that the two models are closely-connected.

Carneiro and Lee (2009) also study estimating distributions given a single binary treatment variable without imposing a rank similarity assumption. This approach requires flexible estimation of the probability of treatment to include as a control function in an equation with $\mathbf{1}(Y \leq y)$ as the outcome, assuming the instruments are exogenous in the selection equation.[5]

Theorem 1 result. Theorem 1 provides a way to identify unconditional QTEs while still conditioning on a separate set of covariates. The implication of conditioning on X is that the conditional probability $P[Y \leq q(D, \tau)|X, Z]$ is not necessarily constant (i.e., equal to τ) for all values of X. Higher education provides additional information about the probability that an individual has earnings below the quantile function.

With conditional quantile models, there are two options. First, one can assume that the conditional probability is the same for all values of the instruments and simply not use any information provided by the additional covariates. Second, one can condition on X, but the conditional framework requires including these variables in the quantile function and estimating conditional QTEs.

The generality of this framework stems from its ability to handle treatment variables differently from control variables. To illustrate the benefits of this generality, let us again consider the case in which the researcher considers all variables as treatment variables in the above framework (i.e., X is empty).[6] In this case, theorem 1 reduces to

$$P[Y \leq q(D, \tau)|Z] = P[Y \leq q(D, \tau)] = \tau.$$

This condition is equivalent to an IVQR condition. The flexibility of the above framework is that control variables are not included in the quantile function $q(D, \tau)$. The researcher can decide which variables to include in the quantile function and which variables to use to inform the conditional probability. This decision should be based on what the quantile function of interest is.

Relationship to literature. A large literature has considered models with nonseparable errors (e.g., Matzkin, 2003; Torgovitsky, 2015), often assuming scalar heterogeneity and monotonicity.[7] A growing literature considers linear random coefficients models without indexing the heterogeneity as in a quantile framework. Masten (2018) discusses conditions necessary for identification of the marginal distributions of coefficients on endogenous variables conditional on exogenous covariates in a system of two linear simultaneous equations. Counterfactual outcomes are not identified in this setup. Hoderlein, Holzmann, and Meister (2017) consider triangular models with random coefficients, placing restrictions on the first-stage coefficients for identification.[8] In contrast, the model of this paper permits estimation of counterfactual outcomes with sparse restrictions at the first stage.

There are techniques to estimate unconditional QTEs with similar motivations as discussed in this paper. Using a propensity score framework, Firpo (2007) introduces a technique for estimating unconditional QTEs with covariates and one exogenous binary treatment variable. Frölich and Melly (2013) extend this method to the case of one endogenous binary treatment variable, building on the approach introduced in Abadie (2003). The estimators in Firpo (2007) and Frölich and Melly (2013) estimate the τth quantile of the outcome distribution for a binary treatment.

The motivations for these estimators are similar to the motivation for GQR, though GQR permits multiple treatment variables, which can be discrete or continuous. In principle, these estimators can be applied to cases where the treatment variables and instruments are discrete but not binary by estimating the effect of each possible value of the treatment variable separately with respect to the baseline (and creating appropriate instruments for each pairwise comparison as well). This approach requires nonparametric estimation of the quantile function, even when the researcher is willing to assume a functional form.

In addition, Firpo, Fortin, and Lemieux (2009) introduce unconditional quantile regression (UQR) for exogenous explanatory variables. The motivation for the UQR estimator is similar to the one discussed in this paper, though the estimand is different.[9] Chernozhukov, Fernández-Val, and Melly

[5]The following section ("Relationship to Literature") discusses the downsides of approaches that estimate $\mathbf{1}(Y \leq y)$ as a function of policy variables and additional covariates. Even when the true equation is linear in the policy variables, these approaches can require nonparametric estimation.

[6]Remember that in the framework of this paper, CH does not permit X variables. See note 3 for an explanation.

[7]There are exceptions. For example, Hoderlein and Mammen (2007) discuss conditions under which marginal effects are identified without these assumptions.

[8]Kasy (2011) discusses assumptions necessary for random coefficient models in triangular systems using the control function approach in Imbens and Newey (2009), showing that such an approach requires restrictions on the heterogeneity in the first-stage equation.

[9]Firpo et al. (2009) discuss estimation of "unconditional quantile partial effects" and "policy effects." These parameters are similar in spirit to unconditional QTEs but are practically different.

(CFM, 2013) note that UQR is "a first-order approximation" of the effect on unconditional quantiles, which may "differ substantially" from the true effect.

CFM propose methods to estimate counterfactual distributions of the outcome variable given different distributions of the exogenous explanatory variables. The first method is similar to the Mata and Machado (2005) estimator, estimating conditional quantile models and then simulating the outcome distribution under different explanatory variable distributions. For the second method, CFM introduce distribution regression (DR). For several possible values of the outcome variable, DR estimates the conditional (on all explanatory variables) probability that the outcome variable is less than this threshold. By estimating this probability for different thresholds, this technique allows the slope coefficients to vary based on the threshold index. Recent work has extended this type of approach in the presence of a single, continuous endogenous treatment variable (Pereda Fernández, 2016).

While GQR reduces to QR when all variables are (exogenous) treatment variables, GQR reduces to DR when all variables are considered control variables.[10] Thus, quantile regression and distribution regression represent two special cases in the GQR framework. Assuming linearity, DR models conditional distributions for each y such that

$$P(Y_i < y|D_i = d, X_i = x) = \Lambda\left(d'\gamma(y) + x'\phi(y)\right).$$

Setting y in the above threshold to $d'\beta(\tau)$ for some d and $\tau \in (0, 1)$, it is not generally true that $\Lambda\left(d'\gamma(y) + x'\phi(y)\right)$ is equal to τ. Even if x is excluded from the estimation of the conditional probability, the estimated probability remains not generally equal to τ. CFM point out that QR and DR coincide in the nonparametric case (e.g., indicator variables contain the entire support of the explanatory variables). GQR, DR, and QR should provide different results in parametric cases. CFM develop the estimation of counterfactual distributions using flexible functions of the explanatory variables. Even flexible functions may not correctly specify the counterfactual distributions given a simple, linear quantile function (i.e., even when the quantile function is linear, DR requires nonparametric estimation of the conditional density function).

C. *Moment Conditions.*

Theorem 1 implies a set of moment conditions.

Corollary 1 *(Moment Conditions). Suppose assumption 1 holds. Then for each* $\tau \in (0, 1)$,

$$E\{m(Z,\tau)[\mathbf{1}(Y \leq q(D,\tau)) - P[Y \leq q(D,\tau)|X]]\} = 0, \tag{3}$$

$$E[\mathbf{1}(Y \leq q(D,\tau)) - \tau] = 0. \tag{4}$$

[10] The CFM technique can be applied to derive and interpret the relationship between the control variables and outcome distribution in the GQR context, though this approach will not be discussed in this paper.

$m(Z, \tau)$ is a function of Z, which can vary across quantiles. Section A.1 includes a brief discussion of conditions (3) and (4), which follow directly from theorem 1.

D. *Identification*

I initially discuss the case where the treatment variables and instruments are discrete, followed by a more general discussion.

Discrete D and Z. I assume that there are K possible (positive probability) values of the treatment variables, and I define the relationship between the instruments and the treatment variables with $1 \times K$ matrix:

$$\Pi(Z,X) \equiv [P(D = d^{(1)}|Z,X) \quad \cdots \quad P(D = d^{(K)}|Z,X)]. \tag{5}$$

Identification requires additional assumptions on this relationship:

Assumption 2.

2A *First Stage:* $E[m(Z,\tau)\Pi(Z,X)]$ *is rank* $K-1$.
2B *Continuity: Y continuously distributed conditional on* Z, X.

Assumption 2A is a first-stage assumption that states that the instruments have an impact on the policy variables. This assumption is stronger than an equivalent assumption in mean regression because the instruments must have a rich effect on the distribution of the policy variables. Moreover, condition 2A implies that the control variables do not perfectly predict the treatment variables. To discuss identification, I consider the alternative function $\tilde{q} \equiv \tilde{q}(D, \tau)$.

Theorem 2 *(Discrete Identification). If (i) assumptions 1 and 2 hold; (ii)* $E\{m(Z,\tau)[\mathbf{1}(Y \leq \tilde{q}) - P[Y \leq \tilde{q}|X]]\} = 0$*; (iii)* $E[\mathbf{1}(Y \leq \tilde{q})] = \tau$*, then* $\tilde{q} = q(D,\tau)$.

A proof is included in section A.1.

Identification for general D. I now consider the continuous case. Define $\epsilon \equiv Y - q(D,\tau)$; $\psi(D,Z,X) \equiv \int_0^1 f_\epsilon(\delta\Delta(D)|D,X,Z)d\delta$ and $\psi(D,X) \equiv \int_0^1 f_\epsilon(\delta\Delta(D)|D,X)\,d\delta$, where $\Delta(D) \equiv \tilde{q}(D,\tau) - q(D,\tau)$. It is necessary to impose a bounded completeness condition. The following condition is the analog to condition L2* in CH and implies that deviations from $q(D,\tau)$ are correlated with the instruments:

Bounded completeness condition:

For any bounded $\Delta(d)$, if

$$E\Big\{m(Z,\tau)\Big[E\big[\Delta(D)\cdot\psi(D,Z,X)|Z,X\big] - E\big[\Delta(D)\cdot\psi(D,X)|X\big]\Big]\Big\} = 0, \text{ then } \Delta(D) = 0, \text{ for } \psi(D,Z,X) > 0.$$

Theorem 3 *(Continuous Identification). Suppose (i) assumption 1 holds; (ii) Y, D have bounded support; (iii) $f_\epsilon(e|D, X)$ and $f_\epsilon(e|D, X, Z)$ continuous and bounded in e; (iv) the bounded completeness condition holds; (v) $E\{m(Z, \tau)[\mathbf{1}(Y \leq \tilde{q}) - P[Y \leq \tilde{q}|X]]\} = 0$; (vi) $E[\mathbf{1}(Y \leq \tilde{q})] = \tau$, then $\tilde{q}(D, \tau) = q(D, \tau)$.*

A discussion is included in section A.1.

III. Generalized Quantile Regression Estimator

This section discusses implementation of GQR. I focus on the case of linear quantiles, $q(d, \tau) = d'\beta(\tau)$ for all d, given its popularity in applied work and relative ease in implementing. I also set $m(Z, \tau) \equiv Z$ for all τ for the implementation of the estimator.

A. Sample Moment Conditions

I introduced moment conditions for the nonparametric function $q(D, \tau)$ in section IIC. The equivalent conditions for linear quantiles are

$$E\left\{Z_i\left[\mathbf{1}\left(Y_i \leq D_i'\beta(\tau)\right) - F(X_i'\delta(\tau))\right]\right\} = 0, \tag{6}$$

$$E[\mathbf{1}(Y_i \leq D_i'\beta(\tau)) - \tau] = 0. \tag{7}$$

I discuss joint estimation of the $\delta(\tau)$ parameters below. I replaced $P\left(Y_i \leq D_i'\beta(\tau)|X_i\right)$ with a more parametric form, $F\left(X_i'\delta(\tau)\right)$, and discuss conditions under which this replacement is appropriate (see assumption 2E') below.

For comparison, instrumental variables quantile regression relies on the moment conditions $E\left\{Z_i\left[\mathbf{1}\left(Y_i \leq D_i'\beta(\tau)\right) - \tau\right]\right\} = 0$.[11] GQR replaces τ in this condition with a function of X_i, which I denote τ_{X_i}. The probability that the outcome is less than or equal to the quantile function varies based on the control variables. On average, it is τ (equation [7]), but GQR does not require this probability to be equal to τ for every observation. Precise estimation of τ_{X_i} is advantageous, but estimation error is not necessarily problematic.

For estimation of $\delta(\tau)$, I assume a maximum likelihood framework:[12]

$$\hat{\delta}(b, \tau) = \underset{\delta(b,\tau)}{\operatorname{argmax}} \sum_{i=1}^{N} \mathbf{1}\left(Y_i \leq D_i'b\right) \ln F(X_i'\delta(b, \tau)) + \mathbf{1}\left(Y_i > D_i'b\right) \ln\left(1 - F(X_i'\delta(b, \tau))\right). \tag{8}$$

[11]Chernozhukov and Hansen (2006) introduce an inverse quantile regression method to simplify estimation that does not use this moment condition specifically. The above condition is more comparable to the approach taken in this paper.

[12]There are likely advantages to using nonlinear least squares estimation for this step using techniques discussed in Khan (2013) and Blevins and Khan (2013). I will rely on methods typically employed by users of standard statistical software to estimate binary choice models, but it is straightforward to replace this step with more flexible alternatives.

I index δ by the parameters associated with the treatment variables (b) and the quantile (τ). The estimation strategy below will require estimation of $\delta(b, \tau)$ for different values of b. Equation (8) implies a binary choice model with outcome $\mathbf{1}\left(Y_i \leq D_i'b\right)$. The framework represented in this equation includes probit and logit regression while also permitting semiparametric estimators (e.g., Klein & Spady, 1993). The linearity assumption of the index can also be relaxed but is imposed here for simplicity.

B. Discussion

The sample moments for GQR are the sample equivalents of equations (6) and (7). If all variables are treatment variables (i.e., there are no control variables), then $\tau_{X_i} = \tau$. Equation (6) reduces to an IVQR moment condition. Thus, IVQR (as well as QR) is a special case of GQR and still available in this framework.

Furthermore, consider the case where there is only a constant and control variables. This case reduces to estimation of $P(Y_i \leq y(\tau)|X_i)$, where $y(\tau)$ represents the τth quantile of the observed outcome distribution. DR requires the estimation of this probability at several different thresholds. Consequently, GQR resembles DR in the case where there are no treatment variables.

C. Estimation

I use a GMM framework for estimation. The moments comprise the vector

$$h_i(b, \delta) \equiv \begin{bmatrix} Z_i\left[\mathbf{1}\left(Y_i \leq D_i'b\right) - F(X_i'\delta)\right] \\ \mathbf{1}\left(Y_i \leq D_i'b\right) - \tau \\ X_i f(X_i'\delta)\left[\dfrac{\mathbf{1}\left(Y_i \leq D_i'b\right) - F(X_i'\delta)}{F(X_i'\delta)\left(1 - F(X_i'\delta)\right)}\right] \end{bmatrix}.$$

The last term is the score of the maximum likelihood function represented in equation (8), where $f(\cdot)$ represents the probability density function. Sample moments are defined by

$$\hat{h}(b, \delta) = \frac{1}{N}\sum_{i=1}^{N} h_i(b, \delta). \tag{9}$$

Estimation uses GMM: $(\widehat{\beta(\tau)}, \widehat{\delta(\tau)}) = \operatorname{argmin}_{b,\delta}\hat{h}(b, \delta)' \widehat{W}\hat{h}(b, \delta)$, for some weighting matrix $\widehat{W}$. Joint estimation of these parameters may be difficult. A contribution of this paper is to provide an estimation technique that is straightforward to implement using standard statistical software. I suggest a method to simplify estimation. Define

$$\mathcal{B} \equiv \left\{b \;\middle|\; \tau - \frac{1}{N} < \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left(Y_i \leq D_i'b\right) \leq \tau\right\}. \tag{10}$$

Constraining the parameters to $\mathcal{B}$ is a simple way to force $Y_i \le D_i'b$ to hold for (as close as possible to) 100τ% of the observations, pertaining to equation (7). To confine b to the set $\mathcal{B}$, I assume the inclusion of a constant in the quantile function. Define $D_i = (1, \tilde{D}_i)$, $b = (\gamma, \tilde{b})$. Let $\gamma(\tau, \tilde{b})$ represent the τth quantile of the distribution of $Y_i - \tilde{D}_i'\tilde{b}$ set:

$$\hat{\gamma}(\tau, \tilde{b}) \quad \text{such that}$$

$$\tau - \frac{1}{N} < \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(Y_i - \tilde{D}_i'\tilde{b} \le \hat{\gamma}(\tau, \tilde{b})) \le \tau. \tag{11}$$

For any $\tilde{b}$, there is a corresponding estimate of the constant, which confines b to $\mathcal{B}$.[13]

The next step is to estimate $\tau_{X_i}(b)$ using equation (8) which models the probability that Y_i is less than or equal to $D_i'b$ as a function of X_i. Certain variables predict that the outcome is above or below the given quantile function. A probit or logit regression is easy to implement, and assumption 2E' permits misspecification at this step as long as it is orthogonal to the instruments. Simulations in section V suggest that probit or logit estimation at this step works well, even when there is little reason to believe that these estimators make appropriate distributional assumptions. Estimation uses

$$g_i(b, \widehat{\delta}(b, \tau)) = Z_i\left[\mathbf{1}\left(Y_i \le D_i'b\right) - \hat{F}(X_i'\widehat{\delta}(b, \tau))\right],$$

with sample moments

$$\hat{g}(b, \widehat{\delta}(b, \tau)) = \frac{1}{N}\sum_{i=1}^{N} g_i(b, \widehat{\delta}(b, \tau)). \tag{12}$$

The estimated parameters minimize a quadratic form of these sample moments, constrained by the set defined in equation (10) and the maximization in equation (8),

$$\widehat{\beta(\tau)} = \arg\min_{b\in\mathcal{B}} \hat{g}(b, \widehat{\delta}(b, \tau))'\hat{A}\hat{g}(b, \widehat{\delta}(b, \tau)), \tag{13}$$

for some weighting matrix $\hat{A}$, where $\widehat{\delta}(b, \tau)$ is the estimate of the parameter vector from equation (8) given b and $\hat{F}(X_i'\widehat{\delta}(b, \tau))$ is the corresponding predicted probability given X_i. When overidentified, two-step GMM is recommended, where the identity matrix is used initially. Using two-step GMM, $\hat{A}$ includes the optimal relative weights for the moments included in $g_i(b, \widehat{\delta}(b, \tau))$. The other moments involve separate calculations or statistical techniques that set the moments close to 0. There is potentially a sacrifice in efficiency using this method, but the computational gains are substantial. Given the estimates $\widehat{\beta(\tau)}$, then $\widehat{\delta(\tau)} = \hat{\delta}(\widehat{\beta(\tau)}, \tau)$.

GQR estimation steps. In many contexts, it is standard to have only one or two treatment variables. Grid searching is practical in these circumstances. When the proposed estimation procedure is used, standard statistical programs are capable of conditioning on numerous covariates. The proposed method makes grid searching more practical by reducing the number of parameters that are estimated independently. The estimation steps are as follows. Define a grid of values for the parameters associated with the policy variables. For each $\tilde{b}$ in the grid:

1. Calculate $\widehat{\gamma}(\tau, \tilde{b})$ using equation (11).
2. Estimate $\hat{\delta}(b, \tau)$ and predict $\hat{\tau}_{X_i}(b)$ using equation (8).
3. Calculate $\hat{g}(b, \widehat{\delta}(b, \tau))'\hat{A}\hat{g}(b, \widehat{\delta}(b, \tau))$.

The b that minimizes $\hat{g}(b, \widehat{\delta}(b))'\hat{A}\hat{g}(b, \widehat{\delta}(b))$ is $\widehat{\beta(\tau)}$. This estimation procedure is straightforward to implement using standard statistical software and arguably easier than conditional quantile estimation techniques.[14] By focusing on the distributional impacts of the treatment variables, which are usually limited in number, simple grid searching coupled with procedures already available in standard statistical programs (relying on optimization methods simpler than those required for QR) are often adequate to implement the above estimation technique. If there are more than two treatment variables, then optimization techniques such as MCMC (see Chernozhukov & Hong, 2003) are necessary.

Identification for linear quantiles. Given the focus on linear quantiles in the context of estimation, I introduce assumptions specific to this case.

Assumption 2'

- 2A' $\begin{pmatrix}\beta(\tau)\\ \delta(\tau)\end{pmatrix}$ *is an interior point of* Θ, *which is compact.*
- 2B' $q(d, \tau) = d'\beta(\tau)$ *for all* d.
- 2C' *The function* $(\beta, \delta) \mapsto E\left[h_i(\beta, \delta)\right]$ *is one-to-one over* Θ *for all* τ.
- 2D' Y_i *continuously distributed conditional on* Z_i, X_i.
- 2E' $E\left\{Z_i\left[P\left(Y_i \le D_i'\beta(\tau)|X_i\right) - F\left(X_i'\delta(\tau)\right)\right]\right\} = 0.$

Assumption 2B' enforces linearity. Assumption 2C' parallels the identification assumption found in Chernozhukov and Hansen (2008; see assumption R6) and requires a rich relationship between the instruments and policy variables conditional on covariates. More primitive assumptions are also possible. Identification is discussed in section A.1 in the proof for theorem 4 below.[15]

[13] $\mathcal{B}$ is nonempty by construction. $\gamma(\tau, \tilde{b})$ is not necessarily unique in finite samples but bound tightly.

[14] The motivation for the proposed estimator is not to improve the computational speed relative to other quantile methods, but it is instructive 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 both recommend grid searching for some parameters and using techniques that are available in most statistical software to jointly estimate the other parameters. Chernozhukov and Hansen (2008) require use of quantile regression optimization for each element in the grid search, while the proposed estimator requires use of probit regression. Given the relative speed of the latter, the proposed estimator is an order of magnitude faster.

[15] The uniqueness of the δ parameters is less important in this context. For example, if the goal is to make comparisons between observations with

Assumption 2E' relates to the specification of $P(Y_i \leq D_i'\beta(\tau)|X_i)$ and is useful when thinking about estimation of this probability. Deviations from $P(Y_i \leq D_i'\beta(\tau)|X_i)$ are not necessarily problematic as long as the errors are orthogonal to the instruments. Choice of $F(\cdot)$ and the functional form of the covariates (linear in X as written in 2E') are important considerations to determine validity of 2E'. In practice, misspecification and poor distributional assumptions may result in $P(Y_i \leq D_i'\beta(\tau)|X_i) \neq F(X_i'\delta(\tau))$. However, 2E' holds if the instruments are uncorrelated with these errors. The advantage of assumption 2E' is that the estimator does not require consistent estimation of the conditional probability.

IV. Properties

This section briefly discusses uniform consistency and asymptotic normality of the GQR estimates as well as inference. I use $\delta(\tau)$ to denote the parameters associated with the control variables for quantile τ.[16] Let $\| \cdot \|$ represent the Euclidean norm.

Assumption 3

3A (Y_i, D_i, Z_i, X_i) *i.i.d.*
3B $F(\cdot)$ *continuous.*
3C $E\left\| \begin{matrix} Z_i \\ X_i \end{matrix} \right\|^{2+\varepsilon} < \infty$ *for some* $\varepsilon > 0$.
3D $G \equiv E\left[\nabla h_i\left(\beta(\tau), \delta(\tau)\right)\right]$ *exists with* $G'WG$ *nonsingular;* $\left\| \left(G'WG\right)^{-1} G'W \right\| < \infty$.
3E $\Sigma \equiv E\left[h_i\left(\beta(\tau), \delta(\tau)\right) h_i\left(\beta(\tau), \delta(\tau)\right)'\right]$ *has finite entries.*

Assumption 3B states that the function representing the probability that the outcome is smaller than the quantile function is continuous, ruling out large jumps in the conditional probability. The other assumptions are standard.

A. Uniform Consistency and Asymptotic Normality

Theorem 4 *(Uniform Consistency and Asymptotic Normality). If assumptions 1, 2', 3 hold and* $\hat{W} \xrightarrow{p} W$ *positive definite, then (i)* $\sup_\tau \left\| \begin{pmatrix} \widehat{\beta(\tau)} \\ \widehat{\delta(\tau)} \end{pmatrix} - \begin{pmatrix} \beta(\tau) \\ \delta(\tau) \end{pmatrix} \right\| \xrightarrow{p} 0$ *and (ii)* $\sqrt{N}\begin{pmatrix} \widehat{\beta(\tau)} - \beta(\tau) \\ \widehat{\delta(\tau)} - \delta(\tau) \end{pmatrix} \xrightarrow{d} N\left[0, (G'WG)^{-1}G'W\Sigma WG\, (G'WG)^{-1}\right]$.

Stochastic equicontinuity is an important condition for this result and follows from the fact that the functional class $\{\mathbf{1}(Y_i \leq D_i'b) - F(X_i'\delta), (b, \delta) \in \Theta\}$ is Donsker and the Donsker property is preserved when the class is multiplied by a bounded random variable.[17] Stochastic equicontinuity then follows from theorem 1 in Andrews (1986). Section A.2 includes further discussion of theorem 4.

B. Inference

For inference, it is possible to estimate the variance-covariance matrix $(G'WG)^{-1}G'W\Sigma WG(G'WG)^{-1}$ using standard methods. $\widehat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} h_i\left(\widehat{\beta(\tau)}, \widehat{\delta(\tau)}\right)h_i\left(\widehat{\beta(\tau)}, \widehat{\delta(\tau)}\right)'$ provides a consistent estimate of Σ, and G can be estimated using finite differences.[18]

Quantile regression inference often depends on estimating the reciprocal of the conditional density of the outcome variable, and it is common to use kernel estimation methods.[19] Broadly, estimating standard errors for quantile estimators can be problematic given the discontinuous nature of the moment conditions.[20]

I propose comparing the value of $h'\Sigma^{-1}h$ when the null hypothesis is imposed to the unrestricted value, where Σ is defined in 3E and h is defined in equation (9). The convergence of the distance metric statistic to a chi-squared distribution is established (Newey & West, 1987; Newey & McFadden, 1994) when a consistent estimate of the variance-covariance matrix is used in the minimization. Typically, this requires use of two-step GMM with an optimal weighting matrix. However, to simplify estimation, I recommend a procedure that constrains some moments to equal 0. In the overidentified case, this procedure does not necessarily use the optimal weighting matrix across all moment conditions, only the unconstrained moment conditions. However, a distance metric can still be used given $\hat{\Sigma}$. I represent the null hypothesis by $a(b) = 0$, where $a(b)$ is rank p. The steps are as follows:

1. Estimate $\widehat{\beta(\tau)}$ and $\widehat{\delta(\tau)}$ using equation (13) and calculate $\hat{h} \equiv \hat{h}(\widehat{\beta(\tau)}, \widehat{\delta(\tau)})$.
2. Estimate $\tilde{\beta}$ and $\tilde{\delta}$ using equation (13) while enforcing $a(b) = 0$ and calculate $\tilde{h} \equiv \hat{h}(\tilde{\beta}, \tilde{\delta})$.
3. Construct $\widehat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}\left[h_i(\widehat{\beta(\tau)}, \widehat{\delta(\tau)})h_i(\widehat{\beta(\tau)}, \widehat{\delta(\tau)})'\right]$ where h_i is defined in section IIIC.
4. $T_N \equiv N\left[\tilde{h}\widehat{\Sigma}^{-1}\tilde{h} - \hat{h}\widehat{\Sigma}^{-1}\hat{h}\right]$ converges in distribution to $\chi^2(p)$ under the null hypothesis.

similar predicted conditional probabilities, then a lack of independent variation in the covariates is not necessarily problematic since it is still possible to make these comparisons. However, condition 2C' nests identification of these parameters as well for simplicity.

[16] Relative to the notation used in equation (8), this notation suppresses the dependence of the estimate of δ on b. The suggested estimation method discussed in section IIIC ("GQR Estimation Steps") involves estimating $\delta(b, \tau)$ for several possible values of b. More generally, $\beta(\tau)$ and $\delta(\tau)$ are estimated jointly, and this dependence does not need to be made explicit.

[17] The other moment conditions are also Donsker under the given assumptions.

[18] Alternatively, a histogram estimation technique resembling the method suggested in Powell (1986) can be implemented. This technique is difficult in this circumstance because it is necessary to estimate the conditional (on X_i) probability that $Y_i - D_i'\widehat{\beta(\tau)}$ is equal to 0.

[19] Parente and Santos Silva (2016) adopt this approach and extend it to account for clustered data. Simulations suggest that this approach can work quite well.

[20] Hagemann (2016) discusses an alternative wild bootstrap approach.

TABLE 1.—SIMULATION RESULTS: RANDOM ASSIGNMENT

	QR (with Covariates)			QR (without Covariates)					
Quantile	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE			
5	0.4111	0.41	0.4177	0.0004	0.03	0.0492			
10	0.3739	0.37	0.3775	0.0020	0.05	0.0710			
15	0.3291	0.33	0.3317	0.0034	0.06	0.0855			
20	0.2821	0.28	0.2842	0.0018	0.07	0.0951			
25	0.2357	0.24	0.2373	0.0016	0.07	0.1017			
30	0.1885	0.19	0.1900	−0.0019	0.07	0.1054			
35	0.1412	0.14	0.1429	−0.0003	0.08	0.1071			
40	0.0942	0.09	0.0962	0.0013	0.08	0.1095			
45	0.0467	0.05	0.0506	0.0015	0.07	0.1096			
50	−0.0005	0.01	0.0194	0.0021	0.08	0.1102			
55	−0.0474	0.05	0.0515	0.0028	0.08	0.1096			
60	−0.0947	0.09	0.0970	0.0026	0.07	0.1062			
65	−0.1414	0.14	0.1431	0.0024	0.07	0.1031			
70	−0.1886	0.19	0.1902	0.0017	0.07	0.0991			
75	−0.2353	0.24	0.2371	0.0031	0.06	0.0943			
80	−0.2804	0.28	0.2826	0.0033	0.06	0.0880			
85	−0.3257	0.33	0.3286	0.0054	0.05	0.0782			
90	−0.3684	0.37	0.3723	0.0047	0.05	0.0663			
95	−0.4040	0.40	0.4110	0.0032	0.04	0.0488			
	DR (Logit)			Machado-Mata			GQR		
	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE
5	−0.0001	0.05	0.0666	0.4348	0.44	0.4354	−0.0083	0.02	0.0314
10	0.0007	0.09	0.1173	0.3902	0.39	0.3907	−0.0045	0.02	0.0303
15	0.0024	0.13	0.1676	0.3431	0.34	0.3436	−0.0046	0.02	0.0316
20	0.0039	0.18	0.2206	0.2952	0.30	0.2957	−0.0037	0.02	0.0319
25	−0.0023	0.22	0.2575	0.2468	0.25	0.2474	−0.0035	0.02	0.0309
30	−0.0227	0.25	0.2783	0.1980	0.20	0.1987	−0.0064	0.02	0.0312
35	−0.0477	0.28	0.2953	0.1490	0.15	0.1500	−0.0065	0.02	0.0312
40	−0.0788	0.29	0.3093	0.0996	0.10	0.1011	−0.0055	0.02	0.0315
45	−0.1121	0.29	0.3279	0.0500	0.05	0.0530	−0.0055	0.02	0.0315
50	−0.1451	0.29	0.3484	0.0005	0.01	0.0176	−0.0057	0.02	0.0313
55	−0.1796	0.29	0.3722	−0.0490	0.05	0.0521	−0.0060	0.02	0.0305
60	−0.2122	0.30	0.3959	−0.0985	0.10	0.1001	−0.0048	0.02	0.0316
65	−0.2460	0.32	0.4205	−0.1479	0.15	0.1490	−0.0057	0.02	0.0330
70	−0.2815	0.31	0.4420	−0.1971	0.20	0.1978	−0.0067	0.02	0.0307
75	−0.3171	0.33	0.4661	−0.2459	0.25	0.2466	−0.0042	0.02	0.0314
80	−0.3590	0.35	0.4957	−0.2942	0.29	0.2947	−0.0055	0.02	0.0316
85	−0.4053	0.38	0.5330	−0.3417	0.34	0.3422	−0.0064	0.02	0.0315
90	−0.4590	0.45	0.5825	−0.3884	0.39	0.3889	−0.0047	0.02	0.0306
95	−0.5343	0.61	0.6626	−0.4333	0.43	0.4339	−0.0040	0.02	0.0310

Results based on 1000 replications, $N = 500$. MAD = mean absolute deviation, RMSE = root-mean-squared-error. DR and Machado-Mata are implemented using the `counterfactual` Stata package.

Large differences (normalized by the variance) between the moment conditions for the constrained and unconstrained estimates suggest that the null hypothesis is wrong. This approach is simple to implement given the proposed estimation strategy. Using a grid search, $\tilde{\beta}$ is often estimated in the process of estimating $\beta(\tau)$. An MCMC approach can also be tailored to estimate both the restricted and unrestricted parameters during the same optimization. The null hypothesis is rejected at significance level α if $T_N > \chi^2_\alpha(p)$.

V. Empirical Applications

A. *Monte Carlo Simulations*

This section tests the performance of the GQR estimator in simulations with a continuous treatment variable. First, I generate data where the policy variable is randomly assigned. Second, I generate data where conditioning on covariates is necessary to obtain consistent estimates.

Random assignment. In the first set of simulations, D is randomly assigned. The impact of D on Y varies by observation and is a function of observed, X_i, and unobserved, U_i, factors. The observed factors have a larger impact on rank. I generate the following data for $N = 500$,

$$Y_i = U_i^*(1 + D_i),$$

where $D_i, X_i \sim U(0, 1)$, $U_i \sim U(0, 0.1)$, and $U_i^* = F_{X+U}(X_i + U_i)$ where $F_{X+U}(\cdot)$ is the CDF of $X_i + U_i$ such that $U_i^* \sim U(0, 1)$. The parameters of interest are $\beta(\tau) = \tau$. I report five 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 under the assumption that D is randomly assigned. Third, I use distribution regression (DR), which relies on a series of logit regressions.[21] Fourth, I use the Mata and Machado (2005) method (MM), which estimates a series of quantile regressions and then integrates out the control variables.[22] Fifth, I use GQR with a probit regression to estimate τ_{X_i}. Results (not shown) are nearly identical if logit regression is used. I present three metrics for each estimator and set of simulations: mean bias, median absolute deviation (MAD), and root-mean-square error (RMSE).

As table 1 shows, QR with covariates is not estimating the quantile function of interest since controlling for additional covariates alters the quantile function. QR (without covariates) produces consistent estimates in this case given that D is randomly assigned. The mean bias is close to 0 throughout the distribution when X is excluded from the QR analysis. DR exhibits significant bias, especially in the top part of the distribution. MM performs poorly as well.

The GQR estimator performs well throughout the distribution. Focusing on the MAD and RMSE metrics, GQR performs better than QR (without covariates) since it is using additional information. This is a major benefit of the GQR estimator even when treatment is unconditionally random.

Conditional random assignment. Next, I generate data where conditioning on X is necessary to obtain consistent estimates. D is conditionally exogenous,

$$Y_i = U_i^*(1 + D_i) \quad \text{where} \quad D_i = X_i + \psi_i,$$

and $\psi_i, X_i \sim U(0, 1)$, $U_i \sim U(0, 0.1)$, and U_i^* is defined as before.

Table 2 presents the same statistics as before. QR (without covariates) now performs poorly given that it is necessary to condition on X. The GQR estimator performs well relative to other methods. The data-generating process is relatively straightforward, but existing quantile and distribution methods are inappropriate to analyze data with a nonseparable disturbance term, which is a function of both unobserved terms and observed variables.

Section B.1 tests the inference procedure discussed in section IVB. The rejection rates are close to the expected rates at 5% and 10% significance levels. Finally, section B.2 studies how GQR performs in predicting counterfactual distributions given two treatment variables when one of the treatment variables is used as a control variable. Even in this "misspecified" case, GQR performs quite well.[23]

[21] I use the Stata package `counterfactual` found at http://www.econ.brown.edu/fac/Blaise_Melly/code_counter.html (accessed September 23, 2014) to implement the DR estimator.

[22] I use the Stata package `counterfactual` to implement this estimator as well.

[23] Note that GQR does not require that all treatment and control variables be categorized correctly so the quantile function is not technically "misspecified." Even when a treatment variable is used as a control variable, condition 1B still holds, as discussed in more detail in section B.2.

B. *Empirical Application: Job Placement*

Autor and Houseman (2010) and Autor et al. (2017) study the effect of job placement into direct-hire jobs and temporary help on future labor earnings. They examine a job placement service in which contractors have varying propensities to place participants in any job at all and, conditional on placement in a job, different probabilities of temporary help versus direct-hire jobs. These varying probabilities act as instruments for the two endogenous variables. Autor and Houseman (2010) estimate mean effects, and Autor et al. (2017) estimate conditional QTEs while discussing that unconditional QTEs are likely of more interest. However, the identification strategy necessitates conditioning on area-time fixed effects, and it is also helpful to condition on the rich set of information known for the individuals in the data. Using IVQR, it is only possible to estimate conditional QTEs. Using GQR, I estimate the quantile function,

$$S_Y(\tau|\text{Temp, Direct}) = \alpha(\tau) + \beta_1(\tau)\text{Temp} + \beta_2(\tau)\text{Direct}, \tag{14}$$

and report the estimates for $\beta_1(\tau)$ and $\beta_2(\tau)$. I use the contractors' probabilities as instruments for both GQR and IVQR. Confidence intervals are generated using the procedure discussed in section IVB. The confidence intervals for IVQR are generated using an equivalent procedure discussed in Chernozhukov and Hansen (2008).[24] The IVQR and GQR results are presented in table 3.[25] I also report the τth quantile of the "untreated" earnings distribution, which is simply equal to the τth quantile of the outcome variable setting Temp = Direct = 0 in equation (14). This metric helps benchmark the quantiles to actual dollar values. An equivalent calculation is more difficult for IVQR given that the quantile function includes more than the treatment variables.[26]

Autor et al. (2017) find strong gradients for both policy variables. As the quantiles increase, the point estimates for temporary placements generally become more negative; the point estimates for direct-hire placements generally become increasingly positive (see figure 3 in their paper). Table 3 suggests a similar pattern replicating the conditional QTE estimates. We observe less evidence of this pattern when estimating unconditional QTEs using GQR. In general, the unconditional effects are larger at the bottom of the distribution than the conditional effects. For example, IVQR

[24] Autor et al. (2017) use standard errors generated by the formula in Chernozhukov and Hansen (2006), which appears to generate much smaller confidence intervals in this application.

[25] Autor et al. (2017) used IVQR code found at http://faculty.chicagobooth.edu/christian.hansen/research/iqrmat.zip, which generates instruments by predicting the endogenous variables (using OLS) based on the exogenous covariates and the excluded instruments (the contractors' probabilities). In contrast, I simply used the two probabilities as the instruments. I used personal code to implement IVQR and searched over a different grid of possible parameter values.

[26] While Autor et al. (2017) include results for the 15th quantile, I find that the (unconditional) quantile function is censored at 0 below quantile 30.

TABLE 2.—SIMULATION RESULTS: CONDITIONAL RANDOM ASSIGNMENT

	QR (with Covariates)			QR (without Covariates)		
Quantile	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE
5	0.4367	0.44	0.4381	1.1401	1.14	1.1517
10	0.3947	0.40	0.3957	1.1407	1.14	1.1469
15	0.3504	0.35	0.3514	1.1379	1.14	1.1421
20	0.3041	0.30	0.3053	1.1275	1.13	1.1307
25	0.2572	0.26	0.2585	1.1134	1.11	1.1158
30	0.2089	0.21	0.2104	1.0937	1.09	1.0956
35	0.1605	0.16	0.1625	1.0724	1.07	1.0739
40	0.1112	0.11	0.1141	1.0494	1.05	1.0507
45	0.0633	0.06	0.0687	1.0243	1.02	1.0253
50	0.0157	0.02	0.0321	0.9970	1.00	0.9979
55	−0.0335	0.03	0.0449	0.9697	0.97	0.9704
60	−0.0814	0.08	0.0879	0.9415	0.94	0.9421
65	−0.1293	0.13	0.1348	0.9102	0.91	0.9107
70	−0.1760	0.18	0.1815	0.8794	0.88	0.8798
75	−0.2226	0.22	0.2281	0.8472	0.85	0.8476
80	−0.2686	0.27	0.2743	0.8208	0.82	0.8212
85	−0.3137	0.31	0.3202	0.8093	0.81	0.8099
90	−0.3519	0.35	0.3601	0.8125	0.81	0.8133
95	−0.3704	0.37	0.3855	0.8289	0.84	0.8309

	DR (Logit)			Machado-Mata			GQR		
	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE
5	0.0083	0.04	0.0634	0.4584	0.46	0.4589	0.0056	0.02	0.0324
10	−0.0129	0.08	0.0912	0.4125	0.41	0.4132	0.0100	0.02	0.0352
15	−0.0376	0.10	0.1210	0.3656	0.36	0.3664	0.0102	0.03	0.0387
20	−0.0664	0.13	0.1542	0.3180	0.32	0.3191	0.0117	0.03	0.0393
25	−0.0977	0.16	0.1889	0.2695	0.27	0.2708	0.0113	0.03	0.0392
30	−0.1345	0.19	0.2269	0.2206	0.22	0.2222	0.0063	0.03	0.0408
35	−0.1676	0.22	0.2696	0.1715	0.17	0.1736	0.0080	0.03	0.0408
40	−0.2066	0.25	0.3113	0.1221	0.12	0.1250	0.0097	0.03	0.0438
45	−0.2460	0.28	0.3550	0.0723	0.07	0.0772	0.0103	0.03	0.0450
50	−0.2865	0.33	0.3998	0.0226	0.02	0.0353	0.0103	0.03	0.0454
55	−0.3286	0.37	0.4480	−0.0273	0.03	0.0387	0.0108	0.03	0.0472
60	−0.3705	0.42	0.4971	−0.0773	0.08	0.0820	0.0115	0.04	0.0495
65	−0.4106	0.46	0.5473	−0.1270	0.13	0.1299	0.0090	0.04	0.0523
70	−0.4432	0.51	0.5940	−0.1771	0.18	0.1792	0.0106	0.04	0.0512
75	−0.4783	0.55	0.6406	−0.2268	0.23	0.2285	0.0148	0.04	0.0528
80	−0.5198	0.59	0.6850	−0.2765	0.28	0.2779	0.0108	0.04	0.0555
85	−0.5677	0.62	0.7279	−0.3256	0.33	0.3268	0.0093	0.04	0.0566
90	−0.6266	0.65	0.7733	−0.3725	0.37	0.3737	0.0125	0.04	0.0571
95	−0.7144	0.67	0.8391	−0.4152	0.42	0.4167	0.0200	0.04	0.0610

Results based on 1,000 replications, $N = 500$. MAD = mean absolute deviation, RMSE = root-mean-squared-error. DR and Machado-Mata are implemented using the `counterfactual` Stata package.

TABLE 3.—EFFECT OF WORK FIRST JOB PLACEMENTS ON EARNINGS QUARTERS 2–8 FOLLOWING ASSIGNMENT

	Mean Effect	IVQR: Conditional QTEs at Quantile						
	2SLS	30	40	50	60	70	80	90
Temporary placement	−57	−37	32	−22	−79	−129	−111	−761
	[−451, 337]	[−305, −10]	[−219, 779]	[−225, 839]	[−370, 679]	[−575, 1355]	[−970, 2009]	[−1450, 2009]
Direct placement	503	166	238	417	679	502	585	1480
	[191, 815]	[40, 320]	[6, 588]	[16, 750]	[30, 1204]	[−114, 1295]	[−127, 2189]	[−535, 2500]
		GQR: Unconditional QTEs at Quantile						
		30	40	50	60	70	80	90
Temporary placement		56	93	−7	133	−344	−223	110
		[−49, 369]	[−163, 460]	[−287, 527]	[−419, 1114]	[−788, 949]	[−1079, 1752]	[−1412, 3500]
Direct placement		250	416	521	519	807	973	570
		[95, 430]	[160, 706]	[259, 866]	[173, 1330]	[79, 1499]	[−232, 1884]	[−1265, 3500]
Untreated earnings		33	150	380	725	1182	1839	3234

$N = 30{,}522$. IVQR refers to estimator in Chernozhukov and Hansen (2008). Confidence intervals in brackets estimated by inverting test statistics as discussed in Chernozhukov and Hansen (2008) for IVQR and section IVB for GQR. All models include indicators for quarter of assignment and district-year interactions. They also include controls for age, age-squared, gender, race (white and Hispanic), total UI earnings, and quarters of employment in eight quarters prior to Work First assignment. The instruments are the contractors' probabilities of temporary and direct-hire placements. Earnings are in 2003 dollars. "Untreated" earnings are the τth quantile of earnings setting both treatment variables to 0 using the GQR estimates. IVQR confidence intervals truncated at 2,500; GQR confidence intervals truncated at 3,500.

estimates that direct-hire placements causally improve earnings by \$166 at the 30th percentile. GQR estimates an increase of \$250 in earnings on a base (untreated) of \$33. We also observe relatively large differences for the direct-hire estimates at quantiles 40 and 50. At quantile 40, IVQR estimates that direct-hire placement increases earnings by \$238, but the GQR estimates that it increases earnings by \$416. At the top of the distribution, we find less evidence of large effects of direct-hire placement. At quantile 90, IVQR estimates an effect of \$1,480, while GQR estimates an effect of \$570, though the confidence intervals for both estimates are large.

Neither IVQR nor GQR finds much evidence of large impacts of temporary placements, though IVQR provides suggestive evidence of large, negative effects at the top of the distribution. There is little evidence of this pattern using GQR. Overall, the unconditional and conditional QTE estimates are quite different. The unconditional QTEs suggest much larger gains at the bottom of the distribution for direct-hire placement. The unconditional QTE estimates for temporary placements are larger at the bottom of the distribution compared to the conditional QTE estimates. The conditional QTE point estimates refer to placement in the distribution conditional on preintervention earnings and many other factors that independently predict earnings. The GQR estimates provide evidence about the impact on the unconditional distribution.[27]

VI. Conclusion

This paper introduces a new, flexible framework for estimating unconditional quantile treatment effects and a corresponding generalized quantile regression estimator. The estimator provides consistent estimates of quantile treatment effects, even in the presence of covariates, for one or more treatment variables, which may be discrete or continuous. These properties distinguish the estimator from alternatives found in the literature. Conditional quantile estimators require altering the quantile function of interest to include additional covariates. The GQR estimator allows one to condition on a separate set of covariates without altering the quantile function. Conditional quantile models assume that the relationship between the treatment variables and the outcome varies based only on unobserved factors; consequently, the interpretation of the parameters changes as some of these factors become observed (i.e., covariates are added to the quantile function). Similar to mean regression, adding covariates when using GQR does not alter the interpretation of the estimates (beyond their effect on the plausibility of the identification assumptions).

[27] The differences between the two sets of estimates suggest that direct-hire placements increase earnings substantially for those with earnings much higher than their previous earnings (and other covariates that predict high earnings) given that IVQR includes prior earnings in the quantile function. However, this does not imply that they create such huge earnings increases at the top of the earnings distribution.

Typically, researchers include control variables for the purposes of identification and do not necessarily want the interpretation of the estimates to change. In fact, much empirical work interprets conditional QTEs as the impact of the treatment variables on the unconditional outcome distribution. GQR provides a straightforward method to estimate unconditional QTEs when the treatments or instruments are conditionally exogenous. QR and IVQR are special cases of the estimator introduced in this paper. Furthermore, distribution regression can also be nested in the framework.

Simulation results illustrate the usefulness of the GQR estimator given simple data-generating processes that likely resonate with researchers. I apply the estimator to study the effect of temporary and direct-hire job placement on labor earnings. Given that the quantile function includes two endogenous variables, existing methods estimating unconditional QTEs for a single binary treatment are not applicable or are potentially difficult to apply.

Many economic models imply heterogeneous effects, motivating analysis that permits treatment effects to vary throughout the outcome distribution. GQR provides an appropriate method to estimate quantile treatment effects and counterfactual distributions.

REFERENCES

Abadie, Alberto, "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics* 113 (2003), 231–263.
Abadie, Alberto, Joshua Angrist, and Guido Imbens, "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica* 70:1 (2002), 91–117.
Andrews, Donald W. K., "Empirical Process Methods in Econometrics" (pp. 2247–2294), in R. F. Engle and D. McFadden, eds., *Handbook of Econometrics* (Amsterdam: Elsevier, 1986).
Autor, David H., and Susan N. Houseman, "Do Temporary-Help Jobs Improve Labor Market Outcomes for Low-Skilled Workers? Evidence from 'Work First'," *American Economic Journal: Applied Economics* 2 (2010), 96–128.
Autor, David H., Susan N. Houseman, and Sari Pekkala Kerr, "The Effect of Work First Job Placements on the Distribution of Earnings: An Instrumental Variable Quantile Regression Approach," *Journal of Labor Economics* 35 (2017), 149–190.
Blevins, Jason R., and Shakeeb Khan, "Local NLLS Estimation of Semi-Parametric Binary Choice Models," *Econometrics Journal* 16 (2013), 135–160.
Carneiro, Pedro, and Sokbae Lee, "Estimating Distributions of Potential Outcomes Using Local Instrumental Variables with an Application to Changes in College Enrollment and Wage Inequality," *Journal of Econometrics* 149 (2009), 191–208.
Chernozhukov, Victor, and Christian Hansen, "An IV Model of Quantile Treatment Effects," *Econometrica* 73 (2005), 245–261.
——— "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics* 132 (2006), 491–525.
——— "Instrumental Variable Quantile Regression: A Robust Inference Approach," *Journal of Econometrics* 142 (2008), 379–398.
——— "Quantile Models with Endogeneity," *Annual Review of Economics* 5 (2013), 57–81.
Chernozhukov, Victor, and Han Hong, "An MCMC Approach to Classical Estimation," *Journal of Econometrics* 115 (2003), 293–346.
Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly, "Inference on Counterfactual Distributions," *Econometrica* 81 (2013), 2205–2268.
de Chaisemartin, Clément, "Tolerating Defiance? Local Average Treatment Effects without Monotonicity," *Quantitative Economics* 8 (2017), 367–396.

Dong, Yingying, and Shu Shen, "Testing for Rank Invariance or Similarity in Program Evaluation," this REVIEW 100 (2018), 78–85.

Firpo, Sergio, "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica* 75 (2007), 259–276.

Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux, "Unconditional Quantile Regressions," *Econometrica* 77 (2009), 953–973.

Frandsen, Brigham R., and Lars J. Lefgren, "Testing Rank Similarity," this REVIEW 100 (2018), 86–91.

Frölich, Markus, and Blaise Melly, "Unconditional Quantile Treatment Effects under Endogeneity," *Journal of Business and Economic Statistics* 31 (2013), 346–357.

Hagemann, Andrea, "Cluster-Robust Bootstrap Inference in Quantile Regression Models," *Journal of the American Statistical Association* 112 (2016), 446–456.

Hoderlein, Stefan, Hajo Holzmann, and Alexander Meister, "The Triangular Model with Random Coeficients," *Journal of Econometrics* 201 (2017), 144–169.

Hoderlein, Stefan, and Enno Mammen, "Identification of Marginal Effects in Nonseparable Models without Monotonicity," *Econometrica* 75 (2007), 1513–1518.

Imbens, Guido W., and Whitney K. Newey, "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," *Econometrica* 77 (2009), 1481–1512.

Kasy, Maximilian, "Identification in Triangular Systems Using Control Functions," *Econometric Theory* 27 (2011), 663–671.

Khan, Shakeeb, "Distribution Free Estimation of Heteroskedastic Binary Response Models Using Probit/Logit Criterion Functions," *Journal of Econometrics* 172 (2013), 168–182.

Klein, Roger W., and Richard H. Spady, "An Effcient Semiparametric Estimator for Binary Response Models," *Econometrica: Journal of the Econometric Society* 61 (1993), 387–421.

Koenker, Roger W., and Gilbert Bassett, "Regression Quantiles," *Econometrica* 46 (1978), 33–50.

Masten, Matthew A., "Random Coefficients on Endogenous Variables in Simultaneous Equations Models," *Review of Economic Studies* 85 (2018), 1193–1250.

Mata, José, and José A. F. Machado, "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression," *Journal of Applied Econometrics* 20 (2005), 445–465.

Matzkin, Rosa L., "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica* 71 (2003), 1339–1375.

Mourifié, Ismael, and Yuanyuan Wan, "Testing Local Average Treatment Effect Assumptions," this REVIEW 99 (2017), 305–313.

Newey, Whitney K., and Daniel McFadden, "Large Sample Estimation and Hypothesis Testing" (pp. 2111–2245), in R. F. Engle and D. L. McFadden, eds., *Handbook of Econometrics*, vol. 4 (Amsterdam: Elsevier, 1994).

Newey, Whitney K., and Kenneth D. West, "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review* 28 (1987), 777–787.

Parente, Paulo M. D. C., and João Santos Silva, "Quantile Regression with Clustered Data," *Journal of Econometric Methods* 5:1 (2016), 1–15.

Pereda Fernández, Santiago, "Estimation of Counterfactual Distributions with a Continuous Endogenous Treatment," Bank of Italy, Economic Research and International Relations Area technical report (2016).

Powell, James L., "Censored Regression Quantiles," *Journal of Econometrics* 32 (1986), 143–1556.

Torgovitsky, Alexander, "Identification of Nonseparable Models Using Instruments with Small Support," *Econometrica* 83 (2015), 1185–1197.

Wüthrich, Kaspar, "A Comparison of Two Quantile Models with Endogeneity," *Journal of Business and Economic Statistics* 38 (2020), 443–456.

A Appendix: For Online Publication

A.1 Moment Conditions and Identification

Theorem 2.1. *Suppose* ***Assumption 1*** *holds. Then for each* $\tau \in (0,1)$,

$$P\left[Y \leq q(D,\tau)|X,Z\right] = P\left[Y \leq q(D,\tau)|X\right], \tag{1}$$

$$P\left[Y \leq q(D,\tau)\right] = \tau. \tag{2}$$

Proof. **Equation** (1):

$$\begin{aligned}
P\left[Y \leq q(D,\tau)|X=x,Z=z\right] &= \int P\left[Y_D \leq q(D,\tau)|X=x,Z=z,V=v\right] dP\left[V=v|X=x,Z=z\right] \quad \text{by } \mathbf{1A,\ 1E} \\
&= \int P\left[Y_{\omega(Z,X,V)} \leq q(\omega(Z,X,V),\tau)|X=x,Z=z,V=v\right] dP\left[V=v|X=x,Z=z\right] \text{ by } \mathbf{1C} \\
&= \int P\left[Y_d \leq q(d,\tau)|X=x,Z=z,V=v\right] dP\left[V=v|X=x,Z=z\right] \quad \text{for any } d \text{, by } \mathbf{1D} \\
&= P\left[Y_d \leq q(d,\tau)|X=x,Z=z\right] \quad \text{by } \mathbf{1D} \\
&= P\left[Y_d \leq q(d,\tau)|X=x\right] \quad \text{by } \mathbf{1B}
\end{aligned}$$

$$\begin{aligned}
P\left[Y \leq q(D,\tau)|X=x\right] &= \int P\left[Y_D \leq q(D,\tau)|X=x,Z=z,V=v\right] dP\left[V=v,Z=z|X=x\right] \quad \text{by } \mathbf{1A,\ 1E} \\
&= \int P\left[Y_{\omega(Z,X,V)} \leq q(\omega(Z,X,V),\tau)|X=x,Z=z,V=v\right] dP\left[V=v,Z=z|X=x\right] \quad \text{by } \mathbf{1C} \\
&= \int P\left[Y_d \leq q(d,\tau)|X=x,Z=z,V=v\right] dP\left[V=v,Z=z|X=x\right] \quad \text{for any } d \text{, by } \mathbf{1D} \\
&= P\left[Y_d \leq q(d,\tau)|X=x\right] \quad \text{for any } d, \quad \text{by } \mathbf{1D}
\end{aligned}$$

$\square$

Proof. **Equation** (2):

$$
\begin{aligned}
P\left[Y \leq q(D,\tau)\right] &= \int P\left[Y_D \leq q(D,\tau) | X=x, Z=z, V=v\right] dP\left[X=x, V=v, Z=z\right] \quad \text{by } \textbf{1A, 1E} \\
&= \int P\left[Y_{\omega(Z,X,V)} \leq q(\omega(Z,X,V),\tau) | X=x, Z=z, V=v\right] dP\left[X=x, V=v, Z=z\right] \quad \text{by } \textbf{1C} \\
&= \int P\left[Y_d \leq q(d,\tau) | X=x, Z=z, V=v\right] dP\left[X=x, V=v, Z=z\right] \quad \text{for any } d\text{, by } \textbf{1D} \\
&= P\left[Y_d \leq q(d,\tau)\right] \quad \text{for any } d, \quad \text{by } \textbf{1D} \\
&= \tau \quad \text{by } \textbf{1A}
\end{aligned}
$$

□

Corollary 2.2 (Moment Conditions)**.** *Suppose* ***Assumption 1*** *holds. Then for each* $\tau \in (0,1)$*,*

$$E\Big\{m(Z,\tau)\Big[\mathbf{1}(Y \le q(D,\tau)) - P\left[Y \le q(D,\tau)|X\right]\Big]\Big\} = 0, \tag{3}$$

$$E[\mathbf{1}(Y \le q(D,\tau)) - \tau] = 0. \tag{4}$$

Equation (3) holds by equation (1) in Theorem 2.1 and the Law of Iterated Expectation. Equation (4) repeats equation (2) in Theorem 2.1.

Theorem 2.3 (Discrete Identification)**.** *If (i)* ***Assumptions 1-2*** *hold; (ii)* $E\Big\{m(Z,\tau)\Big[\mathbf{1}(Y \le \tilde{q}) - P\left[Y \le \tilde{q}|X\right]\Big]\Big\} = 0$*; (iii)* $E\left[\mathbf{1}(Y \le \tilde{q})\right] = \tau$*, then* $\tilde{q}(D,\tau) = q(D,\tau)$*.*

Proof. I define a matrix with elements representing the probability that the outcome is smaller than the quantile function for each possible value of the treatment variables, relative to $P\left[Y \le q(D,\tau)|X\right]$. I define this matrix $\Gamma(Z,X,q)$ which is a function of Z, X, and the function $q \equiv q(D,\tau)$:

$$\Gamma(Z,X,q) \equiv \begin{bmatrix} P(Y_{d^{(1)}} \le q(d^{(1)},\tau)|Z,X) - P\left[Y_{d^{(1)}} \le q(d^{(1)},\tau)|X\right] \\ \vdots \\ P(Y_{d^{(K)}} \le q(d^{(K)},\tau)|Z,X) - P\left[Y_{d^{(K)}} \le q(d^{(K)},\tau)|X\right] \end{bmatrix}.$$

We know from Corollary 2.2 that conditions (ii) and (iii) hold for $q(D,\tau)$. Condition (ii) implies $E[m(Z,\tau)\Pi(Z,X)\Gamma(Z,X,\tilde{q})] = 0$. By Assumption **2A**, it follows that $\Gamma(Z,X,\tilde{q}) = 0$. Without loss of generality, assume that $P(Y_{d^{(1)}} \le \tilde{q}(d^{(1)},\tau)|Z,X) = P(Y_{d^{(1)}} \le q(d^{(1)},\tilde{\tau})|Z,X)$ for some $\tilde{\tau} \in (0,1)$.

Then, $P(Y_{d^{(m)}} \leq \tilde{q}(d^{(m)}, \tau)|Z, X) = P(Y_{d^{(m)}} \leq q(d^{(m)}, \tilde{\tau})|Z, X)$ for all $m = 1, \ldots, K$ since $P[Y \leq \tilde{q}(D, \tau)|X]$ is constant given X. By Assumption **2B**, $\tilde{q}(D, \tau) = q(D, \tilde{\tau})$. Condition (iii) combined with Assumption **2B** then implies that $\tilde{\tau} = \tau$ such that $\tilde{q}(D, \tilde{\tau}) = q(D, \tau)$. $\square$

Continuous Treatment Variables: The conditions and proof for continuous treatment variables are similar to the analysis included in CH concerning identification of conditional QTEs given continuous endogenous variables and instruments.

Theorem 2.4 (Continuous Identification)**.** *Suppose (i)* ***Assumption 1*** *holds; (ii)* Y, D *have bounded support; (iii)* $f_\epsilon(e|D, X)$ *and* $f_\epsilon(e|D, X, Z)$ *continuous and bounded in* e*; (iv) the Bounded Completeness Condition holds; (v)* $E\left\{m(Z,\tau)\Big[\mathbf{1}(Y \leq \tilde{q}) - P\left[Y \leq \tilde{q}|X\right]\Big]\right\} = 0$*; (vi)* $E\left[\mathbf{1}(Y \leq \tilde{q})\right] = \tau$*, then* $\tilde{q}(D,\tau) = q(D,\tau)$*.*

Proof. Assumption (v) implies

$$E\left[m(Z,\tau)\Big[\underbrace{P\left[Y \leq \tilde{q}(D,\tau)|X,Z\right] - P\left[Y \leq q(D,\tau)|X,Z\right]}_{\text{(a)}} - \Big(\underbrace{P\left[Y \leq \tilde{q}(D,\tau)|X\right] - P\left[Y \leq q(D,\tau)|X\right]}_{\text{(b)}}\Big)\Big]\right]$$

Focusing initially on (a):

$$\begin{aligned} P\left[Y \leq \tilde{q}(D,\tau)|X,Z\right] - P\left[Y \leq q(D,\tau)|X,Z\right] &= E\left[E\left[\int_0^1 f_\epsilon\left(\delta\Delta(D)|D,X,Z\right)\Delta(D)d\delta|D,X,Z\right]|X,Z\right] \\ &= E\left[\int_0^1 f_\epsilon\left(\delta\Delta(D)|D,X,Z\right)\Delta(D)d\delta|X,Z\right] \\ &= E\left[\Delta(D)\cdot\omega(D,Z,X)|X,Z\right] \end{aligned}$$

Similarly, for (b):

$$\begin{aligned} P\left[Y \leq \tilde{q}(D,\tau)|X\right] - P\left[Y \leq q(D,\tau)|X\right] &= E\left[E\left[\int_0^1 f_\epsilon\left(\delta\Delta(D)|D,X\right)\Delta(D)d\delta|D,X\right]|X\right] \\ &= E\left[\int_0^1 f_\epsilon\left(\delta\Delta(D)|D,X\right)\Delta(D)d\delta|X\right] \\ &= E\left[\Delta(D)\cdot\omega(D,X)|X\right] \end{aligned}$$

By conditions (iii) and (iv), $\Delta(D) = 0$, implying that $\tilde{q}(D, \tau) = q(D, \tilde{\tau})$ for some $\tilde{\tau} \in (0, 1)$. By condition (vi) and (iii), $\tilde{\tau} = \tau$. $\square$

A.2 Properties

Theorem 4.1 (Uniform Consistency and Asymptotic Normality)**.** *If **Assumptions 1, 2', 3** hold and $\hat{W} \xrightarrow{p} W$ positive definite, then (i)* $\sup_\tau \left\| \begin{pmatrix} \widehat{\beta(\tau)} \\ \widehat{\delta(\tau)} \end{pmatrix} - \begin{pmatrix} \beta(\tau) \\ \delta(\tau) \end{pmatrix} \right\| \xrightarrow{p} 0$*; (ii)* $\sqrt{N} \begin{pmatrix} \widehat{\beta(\tau)} - \beta(\tau) \\ \widehat{\delta(\tau)} - \delta(\tau) \end{pmatrix} \xrightarrow{d} N\left[0, (G'WG)^{-1}G'W\Sigma WG(G'WG)^{-1}\right]$.

Proof.

Identification

We know that $E\left[h_i\left(\beta,\delta\right)\right]$ is equal to 0 at $\beta \equiv \beta(\tau)$ and, given **2E'**, $\delta \equiv \delta(\tau)$. Given **2C'**, $(\beta(\tau), \delta(\tau))$ uniquely solve equations (6), (7), and (8).

Consistency

Next, the proof of this theorem establishes $\begin{pmatrix} \widehat{\beta(\tau)} \\ \widehat{\delta(\tau)} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \beta(\tau) \\ \delta(\tau) \end{pmatrix}$. The conditions necessary for Theorem 2.6 of Newey and McFadden [1994] are met under the following:

1. Identification holds by **2C'** and **2E'** (see above).
2. Compactness of Θ (which is non-empty by construction) holds by assumption **2A'**.
3. $h_i(\breve{\beta},\breve{\delta})$ is continuous at each $(\breve{\beta},\breve{\delta})$ with probability one under **2D'** and **3B**.
4. $E\left|\left|h_i(\breve{\beta},\breve{\delta})\right|\right| \leq E\left\| \begin{matrix} Z_i \\ 1 \\ 4X_i \end{matrix} \right\| < \infty$ for all $\left(\breve{\beta},\breve{\delta}\right) \in \Theta$, implied by **3C**.

Consistency follows from this result.

Uniform Consistency

Next, the proof establishes uniform convergence in $\left(\left(\beta,\delta\right),\tau\right) \in \Theta \times \mathcal{T}$.

Define empirical process

$$v_N(b,\delta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left\{ Z_i \left[\mathbf{1}\left(Y_i \leq D_i'b\right) - F(X_i'\delta)\right] - E\big[Z_i\left[\mathbf{1}\left(Y_i \leq D_i'b\right) - F(X_i'\delta)\right]\big] \right\}.$$

The functional class $\{\mathbf{1}(Y_i \leq D_i'b) - F(X_i'\delta), (b,\delta) \in \Theta\}$ 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,

$$\{Z_i\left[\mathbf{1}(Y_i \leq D_i'b) - F(X_i'\delta)\right], (b,\delta) \in \Theta\}$$

is Donsker. Stochastic equicontinuity of $v_N(b,\delta) - v_N(\beta(\tau),\delta(\tau))$ then follows from **3C** and Theorem 1 in Andrews [1986].

We also need equivalent stochastic equicontinuity conditions for the other moments: $\mathbf{1}(Y_i \leq D_i'b) - \tau$ and $X_i f(X_i'\delta)\left[\frac{\mathbf{1}\left(Y_i \leq D_i'b\right) - F(X_i'\delta)}{F(X_i'\delta)\left(1-F(X_i'\delta)\right)}\right]$. Stochastic equicontinuity follows by the same logic and assumptions. Stochastic equicontinuity implies uniform convergence in $\Big((\beta,\delta),\tau\Big)$.

Given $\begin{pmatrix} \widehat{\beta(\tau)} \\ \widehat{\delta(\tau)} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \beta(\tau) \\ \delta(\tau) \end{pmatrix}$ and **2A'**, with probability approaching one,

$$\hat{G}'\hat{W}\hat{h}(\widehat{\beta(\tau)}, \widehat{\delta(\tau)}) = 0.$$

Expanding $\hat{h}(\widehat{\beta(\tau)}, \widehat{\delta(\tau)})$ around $(\beta(\tau),\delta(\tau))$,

$$\hat{G}'\hat{W}\hat{h}\left(\beta(\tau),\delta(\tau)\right) + \hat{G}'\hat{W}\overline{G}\begin{pmatrix} \widehat{\beta(\tau)} - \beta(\tau) \\ \widehat{\delta(\tau)} - \delta(\tau) \end{pmatrix} = 0,$$

for $\overline{G} \equiv E\left[\nabla h_i\left(\overline{\beta},\overline{\delta}\right)\right]$, where $\left(\overline{\beta},\overline{\delta}\right)$ lies on the line joining $\begin{pmatrix} \widehat{\beta(\tau)} \\ \widehat{\delta(\tau)} \end{pmatrix}$ and $\begin{pmatrix} \beta(\tau) \\ \delta(\tau) \end{pmatrix}$.

$\left(\hat{G}'\hat{W}\overline{G}\right)$ is nonsingular with probability approaching one (given **3D**) such that

$$\begin{pmatrix} \widehat{\beta(\tau)} - \beta(\tau) \\ \widehat{\delta(\tau)} - \delta(\tau) \end{pmatrix} = -\left(\hat{G}'\hat{W}\overline{G}\right)^{-1} \hat{G}'\hat{W}\hat{h}\left(\beta(\tau), \delta(\tau)\right),$$

$$\begin{aligned} \sup_{\tau\in(0,1)} \left\| \begin{pmatrix} \widehat{\beta(\tau)} - \beta(\tau) \\ \widehat{\delta(\tau)} - \delta(\tau) \end{pmatrix} \right\| &= \sup_{\tau\in(0,1)} \left\| \left(\hat{G}'\hat{W}\overline{G}\right)^{-1} \hat{G}'\hat{W}\hat{h}\left(\beta(\tau), \delta(\tau)\right) \right\|, \\ &\leq \sup_{\tau\in(0,1)} \left\| \left(\hat{G}'\hat{W}\overline{G}\right)^{-1} \hat{G}'\hat{W} \right\| \times \sup_{\tau\in(0,1)} \left\| \hat{h}\left(\beta(\tau), \delta(\tau)\right) \right\| \end{aligned}$$

Given uniform convergence (established above) and **3D**, $\sup_{\tau\in(0,1)} \left\| \begin{pmatrix} \widehat{\beta(\tau)} - \beta(\tau) \\ \widehat{\delta(\tau)} - \delta(\tau) \end{pmatrix} \right\| \xrightarrow{p} 0.$

Asymptotic Normality

The next part of the proof establishes condition (ii). This result follows from Theorem 7.2 in Newey and McFadden [1994]. The following conditions hold:

1. Consistency of the parameter estimates was shown above.
2. $G \equiv E\left[\nabla h_i\left(\beta(\tau), \delta(\tau)\right)\right]$ exists with $G'WG$ nonsingular by **3D**.
3. Compactness of Θ holds by assumption **2A'**.
4. $\frac{1}{\sqrt{N}}\sum_i h_i(\beta(\tau), \delta(\tau)) \xrightarrow{d} N(0, \Sigma)$. This condition holds by the Central Limit Theorem under assumptions **3A** and **3E**.
5. Stochastic equicontinuity was established above.

Consequently, all conditions for Theorem 7.2 in Newey and McFadden [1994] hold and the result follows.

□

B Additional Simulation Results

B.1 Simulations: Inference Procedure

In this section, I study the inference procedure proposed in Section 4.2. I test the inference procedure in an instrumental variable setting.

$$Y_i = U_i^*(1 + D_i),$$
$$D_i = Z_i^{(1)} + U_i,$$
$$Z_i^{(1)} = X_i + \psi_i,$$

where $\psi_i, U_i, Z_i^{(2)} \sim U(0,1)$; $X_i \sim U(0, 0.1)$; and U_i^* is defined as before. I generate two instruments, though $Z_i^{(2)}$ is not correlated with the endogenous variable. D_i is a function of U_i, necessitating the use of instruments. $Z_i^{(1)}$ is only an appropriate instrument conditional on X_i. Because there are more instruments than endogenous variables, I use the proposed two-step GMM procedure which generates initial estimates (using the identity matrix as the weighting matrix) and constructs a weighting matrix to use in the second step.

Using these data, I test the null hypothesis $H_0 : \beta(\tau) = \tau$. I use 5% and 10% significance levels. The rejection rates are presented in Table B1. The inference procedure works well throughout the distribution. On average, the rejection rates are close to the expected rates.

B.2 Simulations: Counterfactual Outcomes

The goal of this section is to evaluate the performance of GQR when one of the treatment variables is used as a control variable in estimation. We can think of this as a case when the quantile function is "misspecified," though a motivation of GQR is that the researcher can select the quantile function of interest. With other methods, this choice is often

Table B1: GQR Rejection Rates

Quantile	5% Significance Level	10% Significance Level
5	0.077	0.119
10	0.058	0.106
15	0.054	0.113
20	0.044	0.093
25	0.045	0.091
30	0.042	0.088
35	0.042	0.086
40	0.049	0.084
45	0.046	0.097
50	0.05	0.091
55	0.045	0.092
60	0.041	0.085
65	0.048	0.087
70	0.05	0.09
75	0.043	0.08
80	0.039	0.074
85	0.05	0.095
90	0.053	0.093
95	0.068	0.118
Overall	**0.050**	**0.094**

Results based on 1000 replications, N=500. Inference procedure in Section 4.2 used. Null hypothesis is $\beta(\tau) = \tau$.

unavailable. The model is

$$Y_i = U_i(1 + D_i^1 + D_i^2),$$

where $U_i \sim U(0,1)$. The two treatment variables are also uniformly-distributed and generated to have a correlation of about 0.5. In these simulations, I use different approaches to predict the distribution of $Y \,|\, (D^1 = 1)$.[1] Using GQR, I estimate the quantile function $\gamma(\tau) + \beta^1(\tau)$, where $\gamma(\tau)$ represents the constant and $\beta^1(\tau)$ represent the coefficient on D^1. GQR, here, estimates the quantile function using D^2 as a control variable.

I will also show results using DR and the Mata and Machado [2005] approach, both as implemented by the Stata package `counterfactual`. In this case, note that the Mata and Machado [2005] (MM) approach estimates the correct model. Quantile regression correctly assumes that both D^1 and D^2 are treatment variables. Alternatively, GQR could also be used to estimate $\gamma(\tau)+\beta^1(\tau)+\beta^2(\tau)d_i^2$ for any value of d_i^2 and then simulate the distribution for different values of D_i^2. This is the MM method, and it is still available with GQR (since QR is a special case of GQR). The point of this exercise is to test what happens when GQR is used but one of the treatment variables is used as a control variable. The results are presented in Table B2. MM performs very well. Again, this is not surprising because, in this case, it is assuming the correct model. DR does not perform as well.

The mean bias when using GQR is also small. GQR does not perform as well as MM given that MM is imposing the correct model. However, GQR still does quite well. GQR would likely perform even better if the conditional probability were estimated more flexibly (or the correct functional form for the conditional probability were imposed). Instead, in these simulations, it is assuming a linear index and a normal distribution. As before, there

[1] I generate this counterfactual distribution in the simulated data and use the τ^{th} quantile of the counterfactual outcomes as the "true" value for that simulation.

is little reason to believe that these assumptions are appropriate here and, yet, the estimator performs well (see assumption **2E'**).

Table B2: Counterfactual Outcomes

	DR (Logit)			**MM**			**GQR**		
Quantile	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE	Mean Bias	MAD	RMSE
5	0.0008	0.01	0.0157	0.0006	0.01	0.0081	-0.0012	0.01	0.0087
10	-0.0034	0.01	0.0275	0.0012	0.01	0.0103	-0.0023	0.01	0.0115
15	-0.0069	0.01	0.0396	0.0007	0.01	0.0116	-0.0030	0.01	0.0133
20	-0.0127	0.01	0.0518	0.0006	0.01	0.0131	-0.0046	0.01	0.0156
25	-0.0172	0.01	0.0639	-0.0002	0.01	0.0131	-0.0062	0.01	0.0162
30	-0.0222	0.02	0.0767	0.0005	0.01	0.0146	-0.0070	0.01	0.0178
35	-0.0284	0.02	0.0901	-0.0002	0.01	0.0155	-0.0083	0.01	0.0196
40	-0.0347	0.03	0.1033	-0.0001	0.01	0.0160	-0.0105	0.01	0.0216
45	-0.0437	0.03	0.1172	-0.0002	0.01	0.0163	-0.0116	0.02	0.0230
50	-0.0537	0.04	0.1312	-0.0003	0.01	0.0163	-0.0124	0.02	0.0234
55	-0.0650	0.05	0.1459	-0.0005	0.01	0.0162	-0.0142	0.02	0.0248
60	-0.0763	0.06	0.1608	-0.0004	0.01	0.0163	-0.0160	0.02	0.0262
65	-0.0857	0.07	0.1754	-0.0009	0.01	0.0165	-0.0175	0.02	0.0271
70	-0.0903	0.07	0.1877	-0.0008	0.01	0.0168	-0.0181	0.02	0.0279
75	-0.0894	0.07	0.1982	0.0000	0.01	0.0167	-0.0175	0.02	0.0271
80	-0.0821	0.06	0.2076	-0.0004	0.01	0.0182	-0.0129	0.02	0.0271
85	-0.0783	0.06	0.2203	-0.0011	0.02	0.0215	-0.0085	0.02	0.0287
90	-0.0739	0.05	0.2356	-0.0020	0.02	0.0253	-0.0035	0.02	0.0334
95	-0.0600	0.04	0.2565	-0.0007	0.02	0.0290	0.0031	0.03	0.0412

Results based on 1000 replications, N=500. MAD = Mean Absolute Deviation, RMSE = Root Mean Squared Error. DR and Machado-Mata are implemented using the `counterfactual` Stata package. All estimates are compared to the τ^{th} quantile of the counterfactual outcome distribution given $D^1 = 1$.