A short note on quantile treatment effect model

Tong Zhou*

October 5, 2020

This note is mainly based on the following two papers:

1. Koenker2017 Quantile Regression: 40 Years On (2017), *Annual Review of Economics*, by Roger Koenker
2. CH2013 Quantile Models with Endogeneity (2016), *Annual Review of Economics*, by V. Chernozhukov and C. Hansen.

1 Framework

We use the framework owing to the work of Heckman and Robb. Let D denote the random variable that a treatment may take. The response or outcome variable under a treatment $d \in D$ is denoted as $Y_{D=d}$. Throughout this note, capital letters are used to represent random variables, while small letters mean realizations. Hence, sometimes Y_D is also called *potential* or *latent* outcomes.

The interest of our study is causal or structural analysis of the potential outcome Y_D. Let $q(d, x, \tau)$ denote the τ-th quantile of potential outcomes under the treatment $d \in D$, conditional on the observed characteristics $X = x$. The function $\tau \mapsto q(d, x, \tau)$ is referred to as the quantile treatment response function (QTR). It follows that one quantity that interests us is the quantile treatment effect (QTE), defined as

$$q(d_1, x, \tau) - q(d_0, x, \tau)$$

Endogeneity in this context means that the realized procedure of D is in relation to the potential outcomes, rendering inappropriate using the following restriction

$$\mathbf{P}\left(Y \leqslant \theta(D, X, \tau) | X = x, D = d\right) = \tau \text{ a.s.}$$

to conclude $q(d, x, \tau) = \theta(d, x, \tau)$ as a solution. One of the solution is through the use of an instrument variable Z which is correlated with D but is independent of the potential outcome Y.

*Johns Hopkins University, tzhou11@jhu.edu

The workhorse we shall use a lot is the relation between the latent outcome Y_d for each $d \in D$ and its quantile function $q(d, x, \tau)$, conditional on $X = x$, by

$$Y_d = q(d, x, U_d), \quad \text{where} \quad U_d \sim \mathsf{Unif(0,1)}$$

is the structural error term.

This relation is due to the famous Skorohod Representation.

$\boxed{\text{Skorohod Representation:}}$ given a collection of variables $\{\zeta_d\}$, each variable ζ_d can be represented as $\zeta_d = q(d, U_d)$, for some $U_d \sim \mathsf{Unif}(0, 1)$, where $q(d, \tau)$ denotes the τ-quantile of variable ζ_d.

The role of U_d is critical: it captures the heterogeneity of potential outcomes among individuals with the *same* observed characteristics x.

2 The IVQTE Model

The IVQTE model consists of five assumptions. Fix a common probability space $(\Omega, \mathscr{F}, \mathbf{P})$ and the set of potential outcome variables $\{Y_d\}_{d \in D}$, the covariate variable X and the IV Z. Then the following conditions hold jointly with probability 1:

1. **Potential outcomes:** Conditional on X and for each $d \in D$, $Y_d = q(d, X, U_d)$, where $\tau \mapsto q(d, X, \tau)$ is nondecreasing on $[0, 1]$ and left-continuous and $U_d \sim \mathsf{Unif}(0, 1)$.

2. **Independence:** Conditional on X and for each $d \in D$, U_d is independent of IV Z.

3. **Selection:** $D := \delta(Z, X, V)$ for some unknown function δ and random vector V.

4. **Rank similarity:** Conditional on (X, Z, V), $\{U_d\}$ are identically distributed.

5. **(Observables)**: The observed random vector consists of $Y := Y_D$, D, X, and Z.