Interpreting Unconditional Quantile Regression with Conditional Independence

David M. Kaplan*

October 8, 2020

Abstract

This note provides additional interpretation for the counterfactual outcome distribution and corresponding unconditional quantile "effects" defined and estimated by Firpo, Fortin, and Lemieux (2009) and Chernozhukov, Fernández-Val, and Melly (2013). With conditional independence of the policy variable of interest, these methods estimate the policy effect for certain types of policies, but not others. In particular, they estimate the effect of a policy change that itself satisfies conditional independence.

JEL classification: C21

Keywords: counterfactual, policy, unconfoundedness

1 Introduction

Firpo, Fortin, and Lemieux (2009) and Chernozhukov, Fernández-Val, and Melly (2013), among others, consider a counterfactual distribution of an outcome variable (scalar Y) constructed by replacing the marginal distribution of covariates (vector $\mathbf{X}$) with a new distribution (CDF $G_{\mathbf{X}}(\cdot)$) while maintaining the same conditional distribution (conditional CDF $F_{Y|\mathbf{X}}(\cdot)$). Using the notation from equations (1) and (2) of Firpo, Fortin, and Lemieux (2009), the actual and counterfactual distributions are, respectively,

$$\text{actual: } F_Y(y) = \int F_{Y|\mathbf{X}}(y \mid \mathbf{X} = \mathbf{x}) \, dF_{\mathbf{X}}(\mathbf{x}), \tag{1}$$

$$\text{counterfactual: } G_Y^*(y) = \int F_{Y|\mathbf{X}}(y \mid \mathbf{X} = \mathbf{x}) \, dG_{\mathbf{X}}(\mathbf{x}). \tag{2}$$

Both papers use the phrase "unconditional quantile regression" to mean the change in the quantiles of the unconditional distribution of Y associated with a change in the distribution of $\mathbf{X}$. Firpo, Fortin, and Lemieux (2009) consider an infinitessimal change in the direction

*Department of Economics, University of Missouri. Email: `kaplandm@missouri.edu`

of $G_{\mathbf{X}}$, starting at the actual $F_{\mathbf{X}}$. Chernozhukov, Fernández-Val, and Melly (2013) consider the full change to $G_{\mathbf{X}}$, and thus the difference between quantiles of the distributions in (1) and (2); see their discussion on page 2213 (including footnote 8). Chernozhukov, Fernández-Val, and Melly (2013, §2.2) call the difference between quantiles of the distributions in (1) and (2) a "type 2 counterfactual effect," but they do not mean "effect" in the causal sense, clarifying, "It is important to note that these effects do not necessarily have a causal interpretation without additional conditions" (p. 2210, §2.1).

For description, these methods provide a useful addition to the vast decomposition literature in economics. For example, consider the wage difference between male and female workers. This can be decomposed into two components: the difference in observable characteristics, and the difference in wage distributions conditional on observables. If (1) represents the actual female wage distribution, then (2) represents the counterfactual wage distribution from changing the observable characteristics to the male distribution, but holding constant the female conditional wage distributions.

For policy effects, however, generally (2) is not a good guess of the outcome of a policy that changes the distribution of $\mathbf{X}$ from $F_{\mathbf{X}}$ to $G_{\mathbf{X}}$. In certain settings, it may be plausible that $F_{Y|\mathbf{X}}$ is policy-invariant, but usually it is not, due to the usual sources of endogeneity like selection. For example, if individuals sort into low and high education (X) based on unobserved ability that affects wages (Y), then a policy increasing education would move low-ability individuals into the high-education group, affecting the distribution of wages for that group, i.e., affecting $F_{Y|\mathbf{X}}$.

Firpo, Fortin, and Lemieux (2009) and Chernozhukov, Fernández-Val, and Melly (2013) both mention that (2) is useful for policy analysis given policy-invariant $F_{Y|\mathbf{X}}$.[1] Chernozhukov, Fernández-Val, and Melly (2013) write, "changing the covariate distribution... has a causal interpretation as the policy effect... under the assumption that the policy does not affect the conditional distribution" (p. 2215, §2.3). Similarly, Firpo, Fortin, and Lemieux (2009) say this interpretation holds "under the assumption that the conditional distribution $F_{Y|\mathbf{X}}(\cdot)$ is unaffected by this small manipulation of the distribution of $\mathbf{X}$" (p. 955, §2.1). That is, these approaches take $F_{Y|\mathbf{X}}$ to be "structural" in the sense of "invariant to a class of modifications" (Heckman and Vytlacil, 2007, p. 4848), even though it is not explicitly estimated by Firpo, Fortin, and Lemieux (2009).

My contribution is to characterize which policies indeed do not affect the conditional

[1] Additionally, in their Section 2.3 ("When Counterfactual Effects Have a Causal Interpretation"), Chernozhukov, Fernández-Val, and Melly (2013) consider implications of the same conditional independence assumption I consider below (their (2.8)), but only for "type 1 counterfactual effects" where the conditional distribution changes but the covariate distribution remains fixed (Lemma 2.1), not for unconditional quantile regression.

distribution, given a conditional independence assumption.[2] This is true if the changes in policy variable(s) also satisfy conditional independence. I formulate this first with a binary policy variable, where the variable and its changes are both conditionally independent of the potential outcomes given a vector of controls. Then I formulate the result in terms of a nonparametric, nonseparable structural model with a vector of continuous and/or discrete policy variables, which (along with their changes) are conditionally independent of the unobservables given a vector of controls. For such conditionally independent policy changes, the methods of Firpo, Fortin, and Lemieux (2009) and Chernozhukov, Fernández-Val, and Melly (2013) can estimate policy effects, whereas for other policies, the methods generally do not estimate policy effects.

My results complement the identification results of Rothe (2012). He considers the effect on the unconditional Y distribution of a policy that changes the unconditional distribution of one particular variable of interest. Assuming conditional independence of the policy variable and unobservables (given a vector of controls), and assuming the policy changes the unconditional distribution of the policy variable in a deterministic way that maintains rank invariance, the distributional effect is point identified for a continuous policy variable and set identified for a discrete variable. Instead, in Section 3, I allow the policy to affect a vector of variables, replacing the deterministic rank-invariant policy change with a conditionally independent policy change, which achieves point identification even for discrete policy variables. I also allow stochastic policy changes. I further connect our results by showing that the policy change in Rothe (2012) satisfies conditional independence when the policy variable is continuous but not discrete. With a discrete policy variable, my results show point identification is recovered by strengthening conditional independence of the observed policy variable to conditional independence of Rothe's (2012) underlying rank variable.

Section 2 has results in a potential outcomes framework. Section 3 has results for a general structural model with multiple discrete and/or continuous policy variables.

2 Potential outcomes model

I use the following notation and definitions. Let vector $\mathbf{X}$ be partitioned into $\mathbf{X} = (X_1, \mathbf{X}_2)$, where $X_1 \in \{0, 1\}$ is the binary treatment variable of policy interest and vector $\mathbf{X}_2 \in \mathcal{X}_2$ contains control variables. Potential untreated and treated outcomes are Y_0 and Y_1,

[2] The need to define a particular class of policies is not specific to unconditional quantile regression; e.g., Heckman and Vytlacil (2007, p. 4848) write generally, "A system structural for one class of policy modifications may not be structural for another."

respectively. The observed outcome is

$$Y = Y_0(1 - X_1) + Y_1 X_1. \tag{3}$$

Here, "conditional independence" means

$$(Y_0, Y_1) \perp\!\!\!\perp X_1 \mid \mathbf{X}_2, \tag{4}$$

i.e., conditional on the control variables in $\mathbf{X}_2$, there is independence between the treatment X_1 and the potential outcome pair (Y_0, Y_1). This condition has many other names (conditional exogeneity, unconfoundedness, strong ignorability, etc.).

Assumption A1. Outcome Y is generated depending on potential outcomes as in (3), and conditional independence holds in the sense of (4).

Assumption A2. The policy changes $(X_1, \mathbf{X}_2)$ to $(X_1 + \Delta_1, \mathbf{X}_2)$, where Δ_1 is a random variable satisfying the conditional independence assumption

$$(Y_0, Y_1) \perp\!\!\!\perp \Delta_1 \mid \mathbf{X}_2. \tag{5}$$

The policy change Δ_1 may depend on X_1, $\mathbf{X}_2$, unobservables, and/or randomization, as long as (5) is satisfied. For example, $\Delta_1 = 1 - X_1$ switches all $X_1 = 0$ to $X_1 = 1$ and vice-versa, and it satisfies (5) since Δ_1 only depends on X_1 and X_1 satisfies conditional independence. Similarly, (5) is satisfied by setting all $X_1 = 0$ with $\Delta_1 = -X_1$, or setting all $X_1 = 1$ with $\Delta_1 = 1 - X_1$. Randomization like $\mathrm{P}(\Delta_1 = 0) = \mathrm{P}(\Delta_1 = 1 - X_1) = 1/2$ also satisfies (5) as long as the randomization mechanism is conditionally independent of the potential outcomes. The policy could even depend on $\mathbf{X}_2$, like setting $\Delta_1 = 0$ for certain ranges of $\mathbf{X}_2$, as long as (5) holds. The policy could affect X_1 indirectly, like by changing incentives to participate as in Heckman and Vytlacil (2001), unless the incentives affect individuals differently depending on their potential outcomes (as is often true), thus violating (5). Other than the restriction of changing only X_1 (and not $\mathbf{X}_2$), the main restriction is that the policy may not (explicitly or implicitly) target individuals based on their potential outcomes.

Theorem 1. *Given Assumption A1, a policy satisfying Assumption A2 does not change the conditional distribution $F_{Y|\mathbf{X}}$.*

Proof. Given A1, the actual conditional distribution can be simplified using (3) and (4).

Evaluating the conditional CDF at value y conditional on $X_1 = x_1$ and $\mathbf{X}_2 = \mathbf{x}_2$,

$$\begin{aligned} F_{Y|\mathbf{X}}(y \mid x_1, \mathbf{x}_2) &\equiv \mathrm{P}(Y \le y \mid X_1 = x_1, \mathbf{X}_2 = \mathbf{x}_2) \\ &= \mathrm{P}(\overbrace{Y_{x_1}}^{\text{by (3)}} \le y \mid X_1 = x_1, \mathbf{X}_2 = \mathbf{x}_2) \\ &= \mathrm{P}(Y_{x_1} \le y \mid \overbrace{\mathbf{X}_2 = \mathbf{x}_2}^{\text{by (4)}}). \end{aligned} \tag{6}$$

Under the new policy, the first element of $\mathbf{X}$ becomes $X_1 + \Delta_1$, so

$$\begin{aligned} F_{Y|\mathbf{X}}(y \mid x_1, \mathbf{x}_2) &\equiv \mathrm{P}(Y \le y \mid X_1 + \Delta_1 = x_1, \mathbf{X}_2 = \mathbf{x}_2) \\ &= \mathrm{P}(\overbrace{Y_{x_1}}^{\text{by (3)}} \le y \mid X_1 + \Delta_1 = x_1, \mathbf{X}_2 = \mathbf{x}_2) \\ &= \mathrm{P}(Y_{x_1} \le y \mid \overbrace{\mathbf{X}_2 = \mathbf{x}_2}^{\text{by (4) and (5)}}). \end{aligned} \tag{7}$$

That is, after conditioning on $\mathbf{X}_2 = \mathbf{x}_2$, further conditioning on $X_1 + \Delta_1$ has no effect on the distribution of Y_{x_1}, since both X_1 and Δ_1 are conditionally independent of potential outcomes. Thus, $F_{Y|\mathbf{X}}$ remains unchanged. □

Theorem 1 readily extends to multi-valued treatment, i.e., when the support of X_1 is $\{0, 1, \ldots, J\}$ instead of just $\{0, 1\}$.

Theorem 1 could be applied to the empirical example from Firpo, Fortin, and Lemieux (2009, §4). There, Y is log wage (for U.S. males); X_1 is a dummy for union membership; and $\mathbf{X}_2$ includes dummies for non-white, married, education categories, and ranges of experience. Footnote 18 (p. 962) says, "For simplicity, we maintain the assumption that union coverage status is exogenous. Studies that have used selection models or longitudinal methods [to treat endogeneity] suggest that the exogeneity assumption only introduces small biases." That is, they suggest Assumption A1 is at least a good approximation. However, Assumption A2 must also hold to interpret the UQR estimates as policy effects.

In their setting, Assumption A2 would arguably hold for some policies but not others. It holds for the extreme case of outlawing unionization, but possibly not for marginal policy changes that operate by changing incentives or information sets. For example, if a policy popularizes empirical results that union membership benefits lower-skilled workers more than higher-skilled workers, consequent changes in union membership would likely depend on potential outcomes (via unobserved skill) even conditional on $\mathbf{X}_2$. It is difficult to guess whether a right-to-work law would (approximately) satisfy Assumption A2. If the law deters workers from union membership independently of their potential outcomes conditional on their $\mathbf{X}_2$, then Assumption A2 would hold. However, if low-wage workers are deterred more than high-wage workers by union membership becoming relatively more costly than non-membership, then Assumption A2 may not hold.

3 Structural model with general policy variable

The intuition of Theorem 1 applies to a general structural model with a general vector $\mathbf{X}_1$ affected by the policy. This $\mathbf{X}_1$ can have continuous, discrete, and/or categorical components.

The structural model is

$$Y = h(\mathbf{X}_1, \mathbf{X}_2, \mathbf{U}), \tag{8}$$

where Y is the scalar outcome, $\mathbf{X}_1$ is now allowed to be a vector affected by the policy, $\mathbf{X}_2$ is a vector of control variables, and $\mathbf{U}$ is a vector of unobserved determinants of Y. The structural function $h(\cdot)$ is unknown and unrestricted (i.e., nonparametric and nonseparable) but assumed invariant to the policy considered.

The conditional independence assumption is

$$\mathbf{U} \perp\!\!\!\perp \mathbf{X}_1 \mid \mathbf{X}_2, \tag{9}$$

i.e., conditional on the control variables in $\mathbf{X}_2$, there is independence between $\mathbf{X}_1$ and $\mathbf{U}$.

Assumption A3. Given the policy-invariant structural model in (8), conditional independence holds in the sense of (9).

Assumption A4. The policy changes $(\mathbf{X}_1, \mathbf{X}_2)$ to $(\mathbf{X}_1 + \mathbf{\Delta}_1, \mathbf{X}_2)$, where $\mathbf{\Delta}_1$ is a random vector satisfying the conditional independence assumption

$$\mathbf{U} \perp\!\!\!\perp \mathbf{\Delta}_1 \mid \mathbf{X}_2. \tag{10}$$

Theorem 2. *Given Assumption A3, a policy satisfying Assumption A4 does not change the conditional distribution $F_{Y|\mathbf{X}}$.*

Proof. Given A3, the actual (initial) conditional distribution $F_{Y|\mathbf{X}}$ simplifies to

$$\begin{aligned} F_{Y|\mathbf{X}}(y \mid \mathbf{x}_1, \mathbf{x}_2) &\equiv \mathrm{P}(Y \le y \mid \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2) \\ &= \mathrm{P}(\overbrace{h(\mathbf{X}_1, \mathbf{X}_2, \mathbf{U})}^{\text{by (8)}} \le y \mid \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2) \\ &= \mathrm{P}(h(\mathbf{x}_1, \mathbf{x}_2, \mathbf{U}) \le y \mid \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2) \\ &= \mathrm{P}(h(\mathbf{x}_1, \mathbf{x}_2, \mathbf{U}) \le y \mid \overbrace{\mathbf{X}_2 = \mathbf{x}_2}^{\text{by (9)}}). \end{aligned} \tag{11}$$

Under the new policy, the first elements of $\mathbf{X}$ change from $\mathbf{X}_1$ to $\mathbf{X}_1+\boldsymbol{\Delta}_1$, so

$$
\begin{aligned}
F_{Y|\mathbf{X}}(y \mid \mathbf{x}_1,\mathbf{x}_2) &\equiv \mathrm{P}(Y \le y \mid \mathbf{X}_1+\boldsymbol{\Delta}_1=\mathbf{x}_1, \mathbf{X}_2=\mathbf{x}_2) \\
&= \mathrm{P}(\overbrace{h(\mathbf{X}_1+\boldsymbol{\Delta}_1,\mathbf{X}_2,\mathbf{U})}^{\text{by (8) and A4}} \le y \mid \mathbf{X}_1+\boldsymbol{\Delta}_1=\mathbf{x}_1, \mathbf{X}_2=\mathbf{x}_2) \\
&= \mathrm{P}(h(\mathbf{x}_1,\mathbf{x}_2,\mathbf{U}) \le y \mid \mathbf{X}_1+\boldsymbol{\Delta}_1=\mathbf{x}_1, \mathbf{X}_2=\mathbf{x}_2) \\
&= \mathrm{P}(h(\mathbf{x}_1,\mathbf{x}_2,\mathbf{U}) \le y \mid \overbrace{\mathbf{X}_2=\mathbf{x}_2}^{\text{by (9) and (10)}}).
\end{aligned} \tag{12}
$$

That is, after conditioning on $\mathbf{X}_2=\mathbf{x}_2$, further conditioning on $\mathbf{X}_1+\boldsymbol{\Delta}_1$ has no effect on the distribution of $\mathbf{U}$ and thus no effect on the distribution of $h(\mathbf{x}_1,\mathbf{x}_2,\mathbf{U})$ (given values $\mathbf{x}_1$ and $\mathbf{x}_2$), since both $\mathbf{X}_1$ and $\boldsymbol{\Delta}_1$ are conditionally independent of the unobserved $\mathbf{U}$. Thus, $F_{Y|\mathbf{X}}$ remains unchanged. □

As noted in the introduction, Rothe (2012) considers a related setting; details follow. Assumption A3 is the same, but Assumption A4 differs. Rothe (2012) has scalar X_1 that is (without loss of generality) written in terms of a latent rank variable $W \sim \mathrm{Unif}(0,1)$ as $X_1 = F_1^{-1}(W)$, where $F_1^{-1}(\cdot)$ is the generalized inverse of CDF $F_1(\cdot)$. The policy then changes X_1 from $F_1^{-1}(W)$ to $F_p^{-1}(W)$, where $F_p^{-1}(\cdot)$ is the generalized inverse of CDF $F_p(\cdot)$, the counterfactual CDF of X_1. This imposes rank invariance because $w \ge w' \iff F_1^{-1}(w) \ge F_1^{-1}(w') \iff F_p^{-1}(w) \ge F_p^{-1}(w')$, and it is deterministic in the sense that $F_p(\cdot)$ is nonrandom. If X_1 is continuous, then $W = F_1(X_1)$ and

$$
\Delta_1 = F_p^{-1}(W) - X_1 = F_p^{-1}(F_1(X_1)) - X_1. \tag{13}
$$

This satisfies Assumption A4, suggesting the policy effect is point identified, and indeed Rothe (2012) provides point identification (by a different argument).

However, with discrete X_1, (13) fails. With discrete X_1, $W \ne F_1(X_1)$ because $W \sim \mathrm{Unif}(0,1)$ but $F_1(X_1)$ has a discrete distribution. Consequently, Δ_1 cannot be written solely in terms of X_1 because $F_p^{-1}(W) \ne F_p^{-1}(F_1(X_1))$; instead, $\Delta_1 = F_p^{-1}(W) - X_1$ also depends on W. With the additional assumption of conditional independence of W, point identification would result. However, W may violate conditional independence even if X_1 satisfies it. Thus, generally, policy effects are only partially identified with discrete X_1, and Rothe (2012) derives the identified sets.

References

Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly. 2013. "Inference on Counterfactual Distributions." *Econometrica* 81 (6):2205–2268. URL https://www.jstor.org/stable/23524318.

Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux. 2009. "Unconditional Quantile Regression." *Econometrica* 77 (3):953–973. URL https://www.jstor.org/stable/40263848.

Heckman, James J. and Edward Vytlacil. 2001. "Policy-Relevant Treatment Effects." *American Economic Review* 91 (2):107–111. URL https://www.jstor.org/stable/2677742.

Heckman, James J. and Edward J. Vytlacil. 2007. "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation." In *Handbook of Econometrics*, vol. 6B, edited by James J. Heckman and Edward E. Leamer, chap. 70. Elsevier, 4779–4874. URL https://doi.org/10.1016/S1573-4412(07)06070-9.

Rothe, Christoph. 2012. "Partial Distributional Policy Effects." *Econometrica* 80 (5):2269–2301. URL https://doi.org/10.3982/ECTA9671.