

# Interpreting Unconditional Quantile Regression with Conditional Independence

David M. Kaplan\*

October 21, 2019

## 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: the policy change itself must also satisfy 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 et al. (2009), the actual and counterfactual distributions are, respectively,

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

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

---

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

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 et al. (2009) consider an infinitesimal change in the direction of  $G_{\mathbf{X}}$ , starting at the actual  $F_{\mathbf{X}}$ . Chernozhukov et al. (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 et al. (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).

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 et al. (2009) and Chernozhukov et al. (2013) both mention that (2) is useful for policy analysis given policy-invariant  $F_{Y|\mathbf{X}}$ .<sup>1</sup> Chernozhukov et al. (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 et al. (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 et al. (2009).

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

---

<sup>1</sup>Additionally, in their Section 2.3 (“When Counterfactual Effects Have a Causal Interpretation”), Chernozhukov et al. (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.<sup>2</sup> For such policies, the methods of Firpo et al. (2009) and Chernozhukov et al. (2013) can estimate policy effects, whereas for other policies, the methods generally do not estimate policy effects.

## 2 Toy example

For intuition, I explore a simple textbook example from Hansen (2019, §2.30). The “treatment” variable takes value  $X_1 = 1$  if an individual attends college and  $X_1 = 0$  if not (i.e., only high school). The control variable takes value  $X_2 = 1$  for a high score on the college entrance exam, and  $X_2 = 0$  for a low score. Let  $Y$  be observed wage (dollars per hour). The observed variables are  $(Y, \mathbf{X})$  where  $\mathbf{X} \equiv (X_1, X_2)$ .

There are two types of individuals defined by potential outcomes  $(Y_0, Y_1)$ . The “untreated” potential outcome  $Y_0$  refers to the outcome when  $X_1 = 0$ , and the “treated” potential outcome  $Y_1$  is the outcome when  $X_1 = 1$ . Each type comprises half the population. Jennifer (higher ability) types have  $(Y_0, Y_1) = (10, 20)$ , whereas George types have  $(Y_0, Y_1) = (8, 12)$ . The observed wage is

$$Y = Y_0(1 - X_1) + Y_1X_1. \tag{3}$$

Although I continue the potential outcomes framework below, the example could be rewritten in terms of a structural model  $Y = h(X_1, X_2, U)$ , with  $U$  representing type. Specifically, let  $U = 1$  represent the Jennifer type and  $U = 0$  the George type. Then,  $h(1, X_2, 1) = 20$  (i.e., all Jennifers with a college degree make 20 dollars per hour),  $h(0, X_2, 1) = 10$ ,  $h(1, X_2, 0) = 12$ , and  $h(0, X_2, 0) = 8$ . Further, the potential outcomes can be written as  $Y_0 = h(0, X_2, U)$  and  $Y_1 = h(1, X_2, U)$ , and the intuition remains the same as below. The intuition also remains the same with discrete or continuous  $X_1$ . Thus, for clarity, I focus on the potential outcomes framework with a binary treatment.

---

<sup>2</sup>The need to define a particular class of policies is not specific to unconditional quantile regression; e.g., Heckman and Vytlačil (2007, p. 4848) write generally, “A system structural for one class of policy modifications may not be structural for another.”

In this example, an individual's type influences  $X_2$  but is conditionally independent (given  $X_2$ ) of  $X_1$ . Naturally, Jennifers are more likely to score high on the exam: they have 3/4 probability of high score ( $X_2 = 1$ ), whereas the George type's high-score probability is 1/4. However, conditional on the score  $X_2$ , college attendance ( $X_1$ ) is independent of type:  $(Y_0, Y_1) \perp\!\!\!\perp X_1 \mid X_2$ . Specifically, for both types,

$$P(X_1 = 1 \mid X_2 = 1, Y_0, Y_1) = 3/4, \quad P(X_1 = 1 \mid X_2 = 0, Y_0, Y_1) = 1/4. \quad (4)$$

Given conditional independence, it can be checked that the conditional distribution of the observed  $Y$  given  $(X_1, X_2) = (x_1, x_2)$  is equivalent to the conditional distribution of potential outcome  $Y_{x_1}$  given (only)  $X_2 = x_2$ . That is,

$$F_{Y \mid \mathbf{X}}(y \mid x_1, x_2) \equiv P(Y \leq y \mid X_1 = x_1, X_2 = x_2) = P(Y_{x_1} \leq y \mid X_2 = x_2). \quad (5)$$

Given the setup, 1/2 the population is Jennifer type, who has 3/4 probability of a high score, so 3/8 the population consists of high-scoring Jennifers. Similarly, (1/2)(1/4) = 1/8 the population is high-scoring Georges. Thus, conditional on high score ( $X_2 = 1$ ), (3/8)/(3/8 + 1/8) = 3/4 are Jennifers and 1/4 are Georges. Further knowing  $X_1$  does not change the proportion of types because  $X_1$  is independent of type after conditioning on  $X_2 = 1$ . Thus, the distribution of  $Y$  (observed wage) given  $X_1 = 1$  (college) and  $X_2 = 1$  (high score) equals the distribution of  $Y_1$  (since  $X_1 = 1$ , by (3)) given  $X_2 = 1$ . Given  $X_2 = 1$ , which implied 3/4 Jennifer types who have  $Y_1 = 20$  and 1/4 Georges with  $Y_1 = 12$ , then  $P(Y_1 = 20 \mid X_2 = 1) = 3/4$  and  $P(Y_1 = 12 \mid X_2 = 1) = 1/4$ . Altogether,

$$\begin{aligned} P(Y = 20 \mid X_1 = 1, X_2 = 1) &= P(Y_1 = 20 \mid X_2 = 1) = 3/4, \\ P(Y = 12 \mid X_1 = 1, X_2 = 1) &= P(Y_1 = 12 \mid X_2 = 1) = 1/4. \end{aligned} \quad (6)$$

That is, for  $x_1 = 1$  and  $x_2 = 1$ ,  $P(Y = y \mid X_1 = x_1, X_2 = x_2) = P(Y_{x_1} = y \mid X_2 = x_2)$ , as in (5). The conditional CDF for other  $(x_1, x_2)$  can be found similarly.

I now consider three possible policies. The first two policies change  $F_{Y \mid \mathbf{X}}$ , whereas the

third does not.

**Policy changing  $\mathbf{X}_2$**  The counterfactual conditional distribution generally does *not* match (11) if the distribution of  $\mathbf{X}_2$  changes. The actual distribution of  $Y$  given  $(X_1, X_2) = (1, 1)$  is in (6) above; it combines the treated  $Y_1$  potential outcomes for a mix of 3/4 Jennifer types and 1/4 George types. Imagine an extreme policy that somehow made everyone get a high score,  $X_2 = 1$ . This particularly benefits the George types. If the probability of college admission still depends only on score, then the mix of types among the college-educated subpopulation ( $X_1 = 1$ ) shifts to have 1/2 Jennifers and 1/2 Georges, instead of 3/4 and 1/4. Thus, the conditional distribution changes to have

$$P(Y = 20 \mid X_1 = 1, X_2 = 1) = P(Y = 12 \mid X_1 = 1, X_2 = 1) = 1/2, \quad (7)$$

different than (6). Even if college admissions criteria shifted in response to the policy, in order to match (6), they would have to shift precisely to keep the same proportion of Jennifers and Georges in college. Although this may happen to be true in some cases, it is not true generally.

**Type-dependent policy** Maintaining the distribution of  $\mathbf{X}_2$  is not sufficient for the counterfactual conditional distribution to match (5); policies that depend on the individual's type also fail to maintain the same  $F_{Y|\mathbf{X}}$ . Consider the following policy that affects college admissions without affecting the distribution of exam score ( $X_2$ ). Specifically, low-scoring individuals who are initially rejected are allowed to appeal. The appeal is similar to retaking the exam, so the probability of a successful appeal is 3/4 for Jennifers and 1/4 for Georges. Previously, the population proportion of low-scoring college Jennifers was

$$P(\text{Jennifer}) P(X_2 = 0 \mid \text{Jennifer}) P(X_1 = 1 \mid X_2 = 0, \text{Jennifer}) = (1/2)(1/4)(1/4) = 1/32,$$

whereas the population proportion of low-scoring college Georges was

$$P(\text{George}) P(X_2 = 0 \mid \text{George}) P(X_1 = 1 \mid X_2 = 0, \text{George}) = (1/2)(3/4)(1/4) = 3/32.$$

Similarly, the previous population proportions of low-scoring non-college Jennifers was

$$(1/2)(1/4)(3/4) = 3/32,$$

and the previous population proportion of low-scoring non-college Georges was

$$(1/2)(3/4)(3/4) = 9/32.$$

Thus, conditional on  $(X_1, X_2) = (1, 0)$ , the wage distribution has 1/4 probability of the Jennifer type's treated  $Y_1 = 20$  potential outcome and 3/4 probability of the George type's  $Y_1 = 12$ . Similarly, conditional on  $(X_1, X_2) = (0, 0)$ , the wage distribution has 1/4 probability of the Jennifer  $Y_0 = 10$  and 3/4 probability of the George  $Y_0 = 8$ . However, under the new policy, 3/4 of the low-scoring rejected Jennifers (who are 3/32 of the population) successfully appeal, whereas only 1/4 of the low-scoring rejected Georges (who are 9/32 of the population) successfully appeal. Thus, only  $(1/4)(3/32) = 3/128$  of the population remains low-scoring non-college Jennifer, while  $(3/4)(9/32) = 27/128$  of the population remains low-scoring non-college George. Previously, the low-scoring non-college subpopulation had 1/4 Jennifer types and 3/4 George types, but now it has 1/10 Jennifer types and 9/10 George types. This changes the conditional distribution of wage given  $(X_1, X_2) = (0, 0)$ : there is now only 1/10 probability of Jennifer's  $Y_0 = 10$  and 9/10 probability of George's  $Y_0 = 8$ , whereas before the probabilities were 1/4 and 3/4, respectively. Similarly, the conditional distribution of wage given  $(X_1, X_2) = (1, 0)$  also changes. Thus, even though the marginal distribution of  $X_2$  is unchanged by the policy, the conditional distribution still changes, so the methods of Firpo et al. (2009) and Chernozhukov et al. (2013) do not estimate the effect of the policy when only consider the policy's effect on the distribution of  $(X_1, X_2)$ . The problem is essentially that the policy violates conditional independence since the appeal's success probability implicitly

depends on the individual's type.

**Conditionally independent policy** There is a type of policy that does maintain the conditional distribution from (5), which is stated more generally and formally in Theorem 1. Consider a policy that changes (only) the college admissions probabilities given score, which still (crucially) do not depend on type. Perhaps there is concern that the traditional exam does not correctly identify which individuals would most benefit from college (the Jennifers), but rather there is a sub-group of Jennifers who for various reasons score low. Lacking a better exam that can actually distinguish type, but motivated by equity concerns, a new policy simply admits more low-scoring students. Specifically, now  $P(X_1 = 1 \mid X_2 = 0) = 1/2$ , but still  $P(X_1 = 1 \mid X_2 = 1) = 3/4$ . Under the new policy, the high-scoring subpopulation remains 3/4 Jennifers and 1/4 Georges since the exam does not change; similarly, the low-scoring subpopulation remains 1/4 Jennifers and 3/4 Georges. The only difference is that relatively more (1/2 instead of 1/4) of the low-scoring subpopulation is admitted to college. The conditional wage distribution given  $(X_1, X_2) = (1, 0)$  remains probability 1/4 of Jennifer's  $Y_1 = 20$  and probability 3/4 of George's  $Y_1 = 12$ ; similarly, the conditional wage distribution given  $(X_1, X_2) = (0, 0)$  remains probability 1/4 of Jennifer's  $Y_0 = 10$  and probability 3/4 of George's  $Y_0 = 8$ . The only change is in the marginal probabilities of  $(X_1, X_2) = (1, 0)$  and of  $(X_1, X_2) = (0, 0)$ . As a related but crazier (in this case) example, college could be abolished altogether; i.e., the policy with  $P(X_1 = 1 \mid X_2 = 0) = P(X_1 = 1 \mid X_2 = 1) = 0$  also maintains the same conditional distribution.

### 3 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  $Y_1$ ,

respectively. The observed outcome is

$$Y = Y_0(1 - X_1) + Y_1X_1. \quad (8)$$

Here, “conditional independence” means

$$(Y_0, Y_1) \perp\!\!\!\perp X_1 \mid \mathbf{X}_2, \quad (9)$$

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 (8), and conditional independence holds in the sense of (9).

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

$$(Y_0, Y_1) \perp\!\!\!\perp \Delta_1 \mid \mathbf{X}_2. \quad (10)$$

The policy change  $\Delta_1$  may depend on  $X_1$ ,  $\mathbf{X}_2$ , unobservables, and/or randomization, as long as (10) is satisfied. For example,  $\Delta_1 = 1 - X_1$  switches all  $X_1 = 0$  to  $X_1 = 1$  and vice-versa, and it satisfies (10) since  $\Delta_1$  only depends on  $X_1$  and  $X_1$  satisfies conditional independence. Similarly, (10) is satisfied by setting all  $X_1 = 0$  with  $\Delta_1 = -X_1$ , or setting all  $X_1 = 1$  with  $\Delta_1 = 1 - X_1$ . Completely randomized  $\Delta_1$  would satisfy (10) but lead to cases with  $X_1 + \Delta_1 = 1 + 1 = 2$ ; more feasibly,  $\Delta_1$  could be randomized between 0 and  $1 - 2X_1$ . The policy could depend on  $\mathbf{X}_2$ , like setting  $\Delta_1 = 0$  for certain ranges of  $\mathbf{X}_2$ , again as long as (10) 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 (10). 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 (8) and (9). Evaluating the conditional CDF at value  $y$  conditional on  $X_1 = x_1$  and  $\mathbf{X}_2 = \mathbf{x}_2$ ,

$$\begin{aligned}
 F_{Y|\mathbf{X}}(y | x_1, \mathbf{x}_2) &\equiv \text{P}(Y \leq y | X_1 = x_1, \mathbf{X}_2 = \mathbf{x}_2) \\
 &\stackrel{\text{by (8)}}{=} \text{P}(\overbrace{Y_{x_1}} \leq y | X_1 = x_1, \mathbf{X}_2 = \mathbf{x}_2) \\
 &= \text{P}(Y_{x_1} \leq y | \overbrace{\mathbf{X}_2 = \mathbf{x}_2}^{\text{by (9)}}). \tag{11}
 \end{aligned}$$

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

$$\begin{aligned}
 F_{Y|\mathbf{X}}(y | x_1, \mathbf{x}_2) &\equiv \text{P}(Y \leq y | X_1 + \Delta_1 = x_1, \mathbf{X}_2 = \mathbf{x}_2) \\
 &\stackrel{\text{by (8)}}{=} \text{P}(\overbrace{Y_{x_1}} \leq y | X_1 + \Delta_1 = x_1, \mathbf{X}_2 = \mathbf{x}_2) \\
 &= \text{P}(Y_{x_1} \leq y | \overbrace{\mathbf{X}_2 = \mathbf{x}_2}^{\text{by (9) and (10)}}). \tag{12}
 \end{aligned}$$

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  $\Delta_1$  are conditionally independent of potential outcomes. Thus,  $F_{Y|\mathbf{X}}$  remains unchanged.  $\square$

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

Theorem 1 could be applied to the empirical example from Firpo et al. (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. This would be true 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 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 A2 may not hold.

## 4 Structural model

A version of Theorem 1 can also be derived for the structural model

$$Y = h(X_1, \mathbf{X}_2, \mathbf{U}), \quad (13)$$

where  $Y$  is the scalar outcome,  $X_1$  is the scalar variable of interest,  $\mathbf{X}_2$  is a vector of control variables,  $\mathbf{U}$  is a vector of unobserved determinants of  $Y$ . The structural function  $h(\cdot)$  is unknown and unrestricted (i.e., nonparametric, nonseparable) but assumed invariant to the policy considered. The conditional independence assumption is

$$\mathbf{U} \perp\!\!\!\perp X_1 \mid \mathbf{X}_2, \quad (14)$$

i.e., conditional on the control variables in  $\mathbf{X}_2$ , there is independence between  $X_1$  and  $\mathbf{U}$ .

**Assumption A3.** Given the policy-invariant structural model in (13), conditional independence holds in the sense of (14).

**Assumption A4.** The policy changes  $(X_1, \mathbf{X}_2)$  to  $(X_1 + \Delta_1, \mathbf{X}_2)$ , where  $\Delta_1$  is a random variable satisfying the conditional independence assumption

$$\mathbf{U} \perp\!\!\!\perp \Delta_1 \mid \mathbf{X}_2. \quad (15)$$

**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

$$\begin{aligned} F_{Y|\mathbf{X}}(y \mid x_1, \mathbf{x}_2) &\equiv \text{P}(Y \leq y \mid X_1 = x_1, \mathbf{X}_2 = \mathbf{x}_2) \\ &= \text{P}(\overbrace{h(X_1, \mathbf{X}_2, \mathbf{U})}^{\text{by (13)}} \leq y \mid X_1 = x_1, \mathbf{X}_2 = \mathbf{x}_2) \\ &= \text{P}(h(x_1, \mathbf{x}_2, \mathbf{U}) \leq y \mid X_1 = x_1, \mathbf{X}_2 = \mathbf{x}_2) \\ &= \text{P}(h(x_1, \mathbf{x}_2, \mathbf{U}) \leq y \mid \overbrace{\mathbf{X}_2 = \mathbf{x}_2}^{\text{by (14)}}). \end{aligned} \quad (16)$$

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

$$\begin{aligned} F_{Y|\mathbf{X}}(y \mid x_1, \mathbf{x}_2) &\equiv \text{P}(Y \leq y \mid X_1 + \Delta_1 = x_1, \mathbf{X}_2 = \mathbf{x}_2) \\ &= \text{P}(\overbrace{h(X_1 + \Delta_1, \mathbf{X}_2, \mathbf{U})}^{\text{by (13)}} \leq y \mid X_1 + \Delta_1 = x_1, \mathbf{X}_2 = \mathbf{x}_2) \\ &= \text{P}(h(x_1, \mathbf{x}_2, \mathbf{U}) \leq y \mid X_1 + \Delta_1 = x_1, \mathbf{X}_2 = \mathbf{x}_2) \\ &= \text{P}(h(x_1, \mathbf{x}_2, \mathbf{U}) \leq y \mid \overbrace{\mathbf{X}_2 = \mathbf{x}_2}^{\text{by (14) and (15)}}). \end{aligned} \quad (17)$$

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  $\mathbf{U}$  and thus no effect on the distribution of  $h(x_1, \mathbf{x}_2, \mathbf{U})$  (given values  $x_1$  and  $\mathbf{x}_2$ ), since both  $X_1$  and  $\Delta_1$  are conditionally independent of the unobserved  $\mathbf{U}$ . Thus,  $F_{Y|\mathbf{X}}$  remains unchanged.  $\square$

## References

- Chernozhukov, V., Fernández-Val, I., Melly, B., 2013. Inference on counterfactual distributions. *Econometrica* 81 (6), 2205–2268.  
URL <https://www.jstor.org/stable/23524318>
- Firpo, S., Fortin, N. M., Lemieux, T., 2009. Unconditional quantile regression. *Econometrica* 77 (3), 953–973.  
URL <https://www.jstor.org/stable/40263848>
- Hansen, B. E., 2019. *Econometrics*, Department of Economics, University of Wisconsin.  
URL <https://www.ssc.wisc.edu/~bhansen/econometrics>
- Heckman, J. J., Vytlačil, E., 2001. Policy-relevant treatment effects. *American Economic Review* 91 (2), 107–111.  
URL <https://www.jstor.org/stable/2677742>
- Heckman, J. J., Vytlačil, E. J., 2007. Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In: Heckman, J. J., Leamer, E. E. (Eds.), *Handbook of Econometrics*. Vol. 6B. Elsevier, Ch. 70, pp. 4779–4874.  
URL [https://doi.org/10.1016/S1573-4412\(07\)06070-9](https://doi.org/10.1016/S1573-4412(07)06070-9)